Università Campus Bio-Medico di Roma
Corso di dottorato di ricerca in Scienze e Ingegneria per l'Uomo e l'Ambiente/Science and Engineering for Humans and the Environment
XXXV ciclo a.a. 2019-2020

Artificial Intelligence in Breast Cancer MRI: Applications for TNM staging

Domiziana Santucci

Coordinatore
Prof Giulio Iannello

Tutore
Prof Paolo Soda

Ai miei Fratelli David e Livia

Sostegno ed esempio in questo mio percorso

"

SUMMARY

Figures Index

Tables Index

1 INTRODUCTION

Breast cancer (BC) is the most common non-cutaneous malignancy in women: a woman living up to the age of 90 has a one in eight chance of developing breast cancer[1]. It is ironic and tragic thinking that a neoplasm of an organ as exposed as the breast, easily accessible for clinical and diagnostic examination, continues to require such a heavy toll in terms of human lives.

In the past breast tumors were considered as a single group of diseases. Currently, we know that BC is a very heterogeneous pathology, with a great variety of presentation and clinical, anamnestic and histological scenarios that determine not only an important phenotypic variability, but they affect the prognosis, the probability of recurrence and metastasis, and the therapeutic choice.

It is therefore essential to identify the type, the biological behavior, the histochemical characteristics of the tumor lesions through invasive investigations, such as histological or cytological examination, and non-invasive ones, such as imaging methods, in order to recognize the details of the tumor and make a real "identikit", in order to guide the therapeutic path and accurately establish the prognosis.

Through invasive procedures we obtained tumor histological prognostic factors, which are divided in two groups: the morphological ones, including the histotype, the grading, the nuclear grade, the degree of invasion of the lymphatic vessels and tumor neooangiogenesis, and the biofunctional ones, including the prognostic factors that derive from cellular and molecular biology, such as receptors for steroid hormones, estradiol (ER) and progesterone (PgR), the proliferative activity expressed by antigens such as Ki67, ploidy, the presence of adhesion proteins, such as E- cadherin, of oncogenes / tumor suppressors, of epithelial receptors such as HER2[2–4].

On the other hand, non-invasive procedures, as medical images, allow us to investigate tumor characteristics but also to determine the correct TNM (Tumor, Nodes, Metastases) staging[5]. Nowadays, basing the aforementioned assumptions, non-invasive methods are increasingly evaluated and studied in order to guarantee greater precise and refined tumor characterization

that allows an increasingly personalized therapy and staging that falls within the concept of *precision medicine*[6].

The breast cancer (BC) loco-regional staging is performed with pre- and post-contrast Magnetic Resonance Imaging (MRI) sequences. The former has the task of characterizing the mammary gland, identifying lesions and mainly defining their morphology. Then, the characterization of the lesions, which begins in the pre-contrast sequences, is specified and established with the post-contrast ones, evaluating the contrast agent kinetics within the lesion itself.

In recent years, pilot studies have attempted to correlate the tumor characteristics extracted from MRI with the molecular subtypes of BC through artificial intelligence systems, with the aim of predicting their biological behavior[7,8].

Compared to radiologist inspection, AI-based automated image analysis avoids time consuming processes, extracting valuable and relevant information as quantitative data. There are two kinds of AI algorithms, based on the data interpretation and extraction modality employed:

- the conventional Machine Learning (ML) processes, known as Radiomics algorithms, based on handcrafted features, which learn from experience without being explicitly programmed[9]. (paragraph 3.3.1).
- the Deep Learning (DL)/Deep Neural Network (DNN) algorithms, which process images/raw data and extract information based on mathematical optimization of classification and automatically learn representations from data, minimizing or completely removing human intervention[10]. (paragraph 3.3.2).

In conventional ML field, the extracted features may further be processed through features selection algorithms in order to find the most relevant ones. In a second stage these features are merged as a single value identifying one single tumor signature.

There are four types of conventional ML algorithms: supervised (output: classification, regression), unsupervised (output: clustering), semi-supervised and reinforced. They differ according to the type of training and the output generated by the algorithms themselves.

Supervised algorithms are trained using labelled data, indeed the algorithm learns from prior knowledge. Unsupervised algorithms do not have labelled data and learn patterns from the data, associating them with new connection proposed by the algorithm itself through mathematical methods. Semi-supervised models are between the supervised and unsupervised learning algorithms and use both labeled and unlabeled data for training. Reinforced algorithms use the estimated errors as rewards or penalties.

The most common handcrafted features extraction algorithms are: Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), Graph Run Length Matrix (GRLM), and Histogram of Gradient (HOG). The main limitation of these algorithms is the focus on just one type of feature, such as texture or edge, and then they could not be proposed for other kind of images.

Because of such limitations of extracting inadequate information, systems based on handcrafted features are non-used for unknown data.

To address these limitations DL approaches are used. DL algorithms directly process raw data to produce an output without any other manual and explicit features extraction. DL approaches can process any kind of features from input data (image, text, audio, etc.) providing standardized and robust results[11–14].

DL algorithms originate from Artificial Neural Network (ANN). ANN is a network which mimics the human neurons functioning, where each artificial neuron is the basic building block of the network[15].

In this type of ANN the neurons (called also nodes) are organized in hierarchical layers and can send and receive signals based on the weights that are assigned to them. The response is coded as rejection or acceptance and, in case of acceptance, it is interpreted as a

nonlinear activation function. The weight associated with the input signal affects the transmission to the output layer. The output layer is made up of activation units, each corresponding to label/classes present in the dataset, carrying the weighted output of the hidden layer and predicts the corresponding class[16]. Even more, the functionality of back-propagation and the weight dependence are used during the training phase to reduce the error[17].

Convolutional Neural Network (CNN) is the most used and powerful architecture for medical image analysis. As a DNN, it consists of hierarchical levels of trainable filters (convolution operation) and pooling operations, which manage incoming inputs to improve output performance. These convolutional filters are known as kernels. The number of levels and convolutional filters are defined as hierarchical since the level of extraction of features goes from low-level to high-level and gradually becomes more refined.

There are dozens of studies which employ conventional ML methods in BC MRI field. But, only few studies are present about DL application in BC field and in particular only two works used MRI data[18,19].

The purpose of my PhD program was to follow the path of definition of BC aggressiveness through MRI investigation in relation to the TNM staging in an innovative and up to date way, not only through a detailed analysis of the current literature and its contextual review, but also through the proposal of new conventional ML and DL algorithms, mainly focusing on new or little investigated topics as input for research.

As regards the pre-contrast sequences, we wanted to highlight their role in the characterization of tumor aggression. The first step was therefore to study the T2 weighted (T2w) and diffusion weighted (DW) sequences. For the T2w we focused on the role of tumor associated edema in the correlation with tumor histology using a Random Forest type classifier (chapter 3.1). For the DWI we wanted to validate the role of the Apparent Diffusion Coefficient (ADC) as a predictor of tumor cellularity (chapter 3.2). Both works have shown significant results.

Based on these results, indicating that the tumor and the edema associated with it contains information describing the degree of aggression, we wondered if the lesion itself could also contain information to predict the spread to the axillary lymph nodes.

We first conducted a review of the literature on the prediction of lymph node metastases by AI applied to breast MRI, which highlighted a lack of research in this area (chapter 4.1).

Then we developed a pipeline and applied a Random Forest classifier able to predict the state of the lymph nodes on 99 patients (27 with positive LNS and 72 with negative LNS), using only the post-contrast sequences, considered as the most performing ones, obtaining an AUC higher than 80% (chapter 4.2).

We subsequently investigated whether even DL-type algorithms could give us such satisfactory results in terms of prediction of the lymph node status. We first evaluated the post-contrast images (chapter 4.3) and then assessed the added contribution of the pre-contrast ones proposing a multimodal approach (4.4).

If in the radiomics approach we used a convex hull algorithm to reduce the variability of segmentation with the aim of reach a greater information content, in the DL approach we studied different Bounding box options (basing on the results of the first works about the pre-contrast sequences that showed how the parenchyma around the lesions could influence tumor staging).

With such encouraging results, we went beyond the lymph node prediction, to understand if tumor and peritumor information could also predict distant spread. After a review on the role of AI on the prediction of bone metastases (chapter 5.1), we applied the same DL approach previously used for lymph nodes study, on a sample of 157 patients (39 with metastases and 118 without metastases) to predict the presence or absence of metastases at distance. However, in this case, the results were not significant (chapter 5.2).

The following chapters report the details of each step of my research referring to the relative paper published in medical and non-medical journals. Each chapter is furthermore the work

resulting from the collaboration with radiologists, physicians and engineers coming from different research institutes.

In summary, through of this thesis, concerning breast cancer patients:

- we demonstrated that the BC edema radiomic features extracted from the T2 MRI sequences significantly correlate with the main histological prognostic factors (chapter 3.1);
- we demonstrated that ADC values predict main macroscopic cell growth and replication factors in a highly accurate manner, offering an additional quantitative tool for the evaluation of the entire tumor, assisting the information obtained from the biopsy, which represents only tumor portions (chapter 3.2);
- we conduct of two systematic reviews about the role of artificial intelligence in prediction of lymph node status using breast cancer MRI and in prediction of breast cancer bone metastases using CT and MRI (chapters 4.1 and 5.1);
- we demonstrated that a Radiomics based approach associated to histological and clinical information could predict with high accuracy the lymph node status offering to clinicians and to patients the possibility of avoiding invasive procedures such as lymphadenectomy or lymph node biopsy in unnecessary cases (chapter 4.2);
- we propose three different CNNs and implement different tumor bounding options to systematically assess the contribution gathered by the healthy tissue in the prediction of the axillary lymph node status, evaluating the impact of the 3D and 2D features extracted from input volumes and the influence of healthy tissue in cases of multifocal and multicentric tumors (chapter 4.3);
- we introduce an MDL-based methodology for ALNS evaluation, exploiting three MRI sequences (DCE, T2, and DWI), and clinical and histological characteristics of the primary tumor, implementing a Multi-Input Single-Output 3D CNN where the TF makes the different modality-specific DNNs influence each other in the extraction of features, thus improving the integration of different image modalities (chapter 4.4);

- we demonstrated a currently insufficient ability of the Deep Learning approach in predicting distant metastasis status using CNNs (chapter 5.2).

2 BACKGROUND

2.1 Breast Cancer TNM staging

AJCC 8th edition staging was defined on January 1, 2018. All carcinomas of the breast are covered by this staging.

T describes the size of the tumor and it is correctly evaluated by ultrasound, mammography and definitely by magnetic resonance imaging (MRI). Even more, results published so far regarding MRI-provided tumor details, such as tumor size, morphology, shape, and vascularity characteristics, have been shown to be significant in differentiating breast cancer subtypes.[20–23] Several studies have highlighted the value of MRI after contrast medium administration (DCE-MRI) in reporting the anatomical and functional properties of tumors, guiding the therapeutic choice. However, recognition of the aforementioned tumor characteristics is generally limited to a few qualitative descriptors, and it is still dependent on the radiologist.

N defines Axillary lymph node status (ALNS). It is one of the most important prognostic factors and it has an independent role in determining prognosis and guiding adjuvant therapy decisions[24]. Due to the numerous side effects associated with complete axillary dissection (AD), sentinel node biopsy (SNB) has been proposed in recent years as an alternative to routine staging for patients with early stage BC with clinically negative axillary lymph nodes[5]. However, SNB is also an invasive procedure with significant complications, including shoulder dysfunction, neurogenic damage, upper arm numbness and lymphedema.

M describes distant spread of tumor. About 30% of early-stage breast cancer patients will experience distant metastases, and 90% of patient deaths derive from complications from metastases.[25] Most common sites of metastases are bone (58%), lung and pleura (26%) and contralateral lymph nodes (16%). Actually, the metastasis spread depends on various factors, such as histological and intrinsic behavior of the tumor, timeliness of diagnosis, genetic and individual variability and so on. Metastasis status is determined through a first step total body

CT and more deeply investigated with dedicated MRI. Sometimes the metastatic lesion requires a biopsy confirmation.

2.2 The role of MRI in Breast Cancer

Magnetic Resonance Imaging (MRI) is considered the most accurate imaging technique in the determination of breast cancer lesions, in particular the importance of its diagnostic role has been widely defined through the execution of dynamic sequences after administration of contrast agent based on Gadolinium.[26]

The indications for the use of MRI in breast cancer imaging were defined in detail by the European Society of Breast Cancer Specialists (EUSOMA) in a workshop in October 2008 and are as follows: screening in high-risk women, breast implants, pre-surgical staging, response to neo-adjuvant therapy evaluation, post-surgical follow-up, CUP-syndrome (Cancer of Unknown Primary - occult neoplasm syndrome).[27]

Breast MRI has evolved from a primarily contrast-enhanced technique to a multiparametric technique, in which T2-weighted and diffusion-weighted imaging (DWI) are routinely performed. Still, the basis for any MRI protocol is a dynamic T1-weighted contrast-enhanced sequence.

Pre-contrast sequence: T2 and DW images

T2-weighted imaging is nowadays included in the standard MRI protocol. T2-weighted imaging are usually performed both with fat suppression, for cysts visualization, and without fat suppression, for lesion morphology characterization. Benign lesions are depicted with high T2-signal intensity, while most cancers do not show high signal intensity compared to the normal breast parenchyma for their high cellularity and low water content, with some exceptions (mucinous carcinoma, necrotic cancer, and metaplastic carcinoma). T2-weighted images are also used to describe the presence and the localization of associated tumor edema, improving the lesion classification. Several studies have demonstrated the role of T2-weighted sequences in increasing the specificity for differentiation of benign and malignant lesions[28,29].

DWI quantifies the random movement of water molecules in the analyzed tissue called Brownian motion. It depends on tissue microstructure and cellular density. Tumor lesions

present an increased cell density and an alteration of microstructure with subsequent decreasing of water diffusion and higher signal intensity at DWI sequences. DWI does not require a contrast agent administration and it is performed in a short acquisition time, optimizing the b values and the fat suppression and minimizing the artifacts improving the SNR[30]. The DWI is also performed before the contrast medium administration because it could vary the DWI intrinsic resolution[31]. While the DWI sequence gives the radiologist a qualitative information about tissue, the apparent diffusion coefficient (ADC), obtained by ADC maps derived from DWI images, is a quantitative measure of the cell diffusivity: benign lesions are characterized by high ADCs (range, 1.2–2.0 3 1023 mm2/sec), while malignant ones are associated to lower ADCs. Some studies have reported that diagnostic performance in distinguishing benign and malignant lesions is affected neither by the choice of the field strength (1.5-T or 3.0-T MRI unit)[32] nor by the combination of multiple b values.[33,34] The ADC measurements vary depending on b values[34].

Post-contrast sequences

Among the MRI sequences, from the very beginning, DCE-MRI, or MRI after administration of contrast medium (contrast medium), stood out for its high sensitivity and its high positive predictive value. This examination is still considered the most appropriate in the field of breast cancer, for the solution of diagnostic-clinical problems not resolved by conventional study techniques (Mammography and Ultrasound).

The contrast medium used is usually based on Gadolinium molecules. Gadolinium contrast temporarily alters the molecular properties within the magnetic field of the structures where it is located, changing the signal intensity of the tissue itself dependently on their location[35].

The acquisition of images includes:

- Detection of a T1-weighted pre-contrast image (T10map);
- Acquisition of T1-weighted images during and after administration of the contrast medium.

The images obtained describe the distribution pattern of the contrast medium in the tissue which is better evident and easier to interpret in the “subtracted” sequences.

The signal documented by the passage of the contrast medium is used to generate the so-called intensity/time curves (I/TC) which reflect the absorption of the tissue upon arrival of the contrast medium, through the different signal enhancement values ("enhancement").

There are mainly three types of I/TC:

- Type I: characterized by a slow and progressive wash-in and wash-out. It describes benign lesions that has poor vascularization and therefore poor and slow proliferation;
- Type II: defined by a slow wash-in and a terminal plateau, which indicates a slow, gradual and lasting impregnation of the contrast medium. It characterizes borderline lesions and requires further investigation with further sequences and morphological evaluation of the lesion itself;
- Type III: identified by a rapid wash-in, with surge, and an early and steep wash-out. This is typical for highly vascularized lesions and high replication speed, therefore characteristic of malignant tumor lesions.

The intralesional distribution of the contrast medium allows the distinction between different vascularization patterns: homogeneous, characteristic of benign lesions; heterogeneous, typical of borderline or malignant lesions; or with rim-enhancement, which identifies malignant lesions with high mitotic activity, where an expansion corresponds to a rapid eccentric growth and consequent central necrosis due to lack of angiogenesis.

2.3 Application of Artificial Intelligence in Breast Cancer

The primary goal of AI application in medical imaging is to achieve greater efficacy and efficiency in routine clinical practices getting support in the decision-making process.

As previously described, Images contain a huge amount of data, that radiologists cannot analyze manually. The use of AI for medical imaging analysis reduces the inter-readers variability, decreases the uncertainty, and provides quantitative information which may support physician evaluation.

In MRI breast cancer field, conventional AI (ML) algorithms are largely used to detect breast lesions, define their characteristics, discriminating benign vs malignant lesions and predict their histological behaviour and prognosis.

On the other hand, DL approaches have only been investigated recently and few works are present in literature nowadays.

2.3.1 Conventional ML algorithms: the Radiomics approaches

The radiomics process involves many challenging steps, as listed by Gillies et al.[8]

Images acquisition

Although the breast lesion can already be characterized on MRI examination, the evident features are generally described subjectively and qualitatively through medical reports with terms such as: "modestly heterogeneous enhancement", "strongly spiculated margins", "large central necrotic core" and so on. for this reason, these evaluations suffer from a great intra- and inter-observer variability. The Radiomics aim is to standardize image acquisition and analysis, through quantitative, more objective and precise image descriptors potentially usable as non-invasive prognostic biomarkers as well as to objectively quantify the imaging features that could reveal tumor biology.[36]

The first step towards an objective analysis is to perform imaging exams, following the same protocols and standardized methods during image acquisition and reconstruction, with the intent of reducing distortion and variability between images acquired in different periods or from different sources.

Volume of interest and segmentation

The lesions in question should be treated as volumes of interest. Normally the radiological reports do not report a detailed analysis of the so-called subvolumes within the lesions, which can have a prognostic value; instead Radiomics captures and analyzes these sub-volumes of interest using quantitative measures.

Segmentation is the crucial part of the entire analysis process: during segmentation only the Region of Interest (ROI) must be isolated from the rest of the organ, in our case the breast, paying attention to the edges, which are often not well defined. This process can also be achieved with the help of algorithms capable of describing the edges, but, due to the high difficulty in detecting the true boundaries of the lesion and the heterogeneity of the volume of interest itself, the resulting ROI requires an additional check by a radiologist.

Extraction and selection

The core of radiomics is the extraction of data from the features of the ROI, which can quantify the intensity, shape, texture and morphology of the tumor. These features include form-based, first-, second- and higher-order statistical determinants, and model-based characteristics. Shape-based features are external representations of a region and characterize the shape, size, and surface information of ROIs. Typical metrics include sphericity and firmness.[37] First-order characteristics (e.g. mean, median) describe the overall intensity and variation of ROIs, ignoring spatial relationships[36,38]. Contrasting second order (texture) characteristics can provide interrelationships between voxels. Material characteristics can be extracted from different matrices, for example, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), etc. [39] Semantic characteristics describe the qualitative characteristics of the image typically used in the radiological workflow.

Features can be extracted through many commercial software and algorithms and their quantity is crucial: if too many features are extracted, it is possible to run into overfitting caused by the small sample generally taken into consideration; if on the other hand the features are limited in number, they may not be able to acquire the real distribution of the data. It is possible to avoid the large size of the features vector by eliminating those features that can be redundant (highly correlated with each other) or irrelevant: this type of function manipulation is called feature selection.

To get the best predictive model, the selection of the appropriate set of features and their use to train machine learning models is crucial. Having irrelevant features in the data can reduce the accuracy of models, in particular linear algorithms, because it can increase the training time and can lead to overfitting.

Sometimes, the feature selection and model construction can be implemented together, (embedded method), such as the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm. On the other hand, wrapper methods select the features based on the models' performance for different subsets of features and the model needs to be rebuilt such as the Recursive Feature Elimination Support Vector Machines (SVM-RFE) algorithm.

2.3.2 DL algorithms: the DNN characteristics

For breast cancer analysis, two types of ANN have been used i.e. Shallow Neural Network (SNN)[40], and DNN[41]. In particular the last one has been used for breast cancer lesions due to its capability of automatically processing the input signals and computing features from irregular and abnormal ROI without any human intervention; those are named automatic features. Among the DNN, CNN is the most powerful and used for breast lesions analysis.

All CNN-based models employed in oncology have been shown to improve the accuracy of diagnosis and reduce false negative detection[42]. There are several CNN architectures applied in the literature in the field of BC (LeNet[43], AlexNet[44], GoogLeNet[45], Inception V4[46], ResNet[47]

and others): all have reported robust results when applied to different datasets. CNNs have been used for the study of the geometry of breast tissues, for the classification of tumor lesions and for their detection. The performances of the works reported in this recent review about the application of AI in breast cancer have shown excellent results and in particular a dependence on the complexity depth of the architecture[48].

In medical imaging, DCE, DWI, or T2 sequences are examples of image modalities. The use of multiple and different sources provides complementary information that improves the detection and analysis of diseases, resulting in the need for physicians of techniques that aim to combine and simultaneously analyze different image modalities. Multimodal Deep Learning (MDL) exploits DL approaches to implement methods allowing the fusion of complementary information coming from heterogeneous sources with the aim of providing a richer data representation than the unimodal approach.

In recent years, several data fusion techniques have been investigated in the research community[49,50] resulting in three main levels: early fusion or data-level, late fusion or decision-level, intermediate or joint fusion. Early fusion consists of the integration of different and heterogeneous sources of data in a single structure that is then used as input for a DL model. In the case of medical image analysis using CNNs, the simplest strategy involves concatenating the acquisitions in a single volume[51]. However, the inherent characteristics of each imaging modality, such as different resolutions or sampling times, may make the creation of a single, very complex structure. As a consequence, it is possible to extract a high-level representation from each input modality before performing the fusion. In this case, handcrafted features and those extracted exploiting a DNN are merged in a single layer. Since the features are first extracted from each imaging modality and then merged in a single structure, the early fusion may not fully take into account the complementary nature of the images, creating vectors with redundancy.

This characteristic requires the implementation of techniques for features reduction like the principal component, independent analysis, and correlation analysis.
Late fusion integrates the decision coming from different models, each trained on a specific

image modality. In other words, this technique combines the decision of independent "experts", exploiting the fact that errors from multiple models tend to be uncorrelated. There are different combining strategies such as majority voting, averaged fusion, Bayes' rule, or those exploiting the use of a meta-model. Indeed, late fusion started to be popular with the growth of solutions involving ensemble learning. It is worth noting that, among all multimodal fusion techniques, late fusion represents the simplest technique to implement.

Intermediate fusion (IF) exploits the characteristic of DNNs to be able to transform raw inputs into higher-level representations, with the aim of creating a shared representation. In the literature, the shared representation layer can be implemented according to the two main methodologies, proposed in [52] and [53] respectively. The first simple and effective approach implements a concatenation of the extracted vectors from a given layer of each unimodal model[52]. The resulting representation is processed by the following hidden layers. On the other hand, the approach described in[53] is a multiplicative method that computes the Kronecker product between all the modality-specific features vectors, extracted from a given hidden layer, considering also the unimodal representation. In the intermediate approach, it is not easy to understand when the modality-specific representation should be merged into a shared representation. In the literature, while different solutions proposed a single fusion layer, several approaches implement a gradual fusion strategy[54,55]. The choice of which modality to fuse at which depth of representation can be very challenging, especially in cases where more than two sources are present.

DL-based multimodal learning offers several advantages when compared to classical ML techniques, where data fusion usually consists of a handcrafted rigid architecture. ML approaches typically implement early or late fusion, explicitly performing features selection and dimensionality reduction. Since features are manually designed by domain experts requiring prior knowledge about the problem to be solved, the fusion techniques may be sensitive to data pre-processing. On t+he other hand, in MDL, the DNNs support the development of solutions based on early, late, and intermediate fusion and autonomously learn both the best features representation, making the features extraction step useless.

3 TUMOR

3.1 T2w-sequences and breast tumor associated edema

Breast cancer main prognostic factors include tumor size, lymph node involvement, tumor histology and nuclear grade, lymphatic and vascular invasion, and molecular status (which include estrogen-receptor (ER) and progesterone-receptor (PgR) status, epithelial growth factor receptor HER2 status and expression of proliferation-related genes, such as Ki-67) [56,57].

Breast edema is not considered a major prognostic factor in current national and international guidelines (ACR and AIOM) [5,58]. However, several studies have shown a correlation between breast edema and tumor aggressiveness[59,60]. Focal edema is mostly correlated with malignancy, representing an indirect sign of peritumoral vascular invasion and inflammation.

As a highly sensitive and non-invasive method, magnetic resonance imaging (MRI) has a prominent role in BC diagnosis, pre-operative tumor staging and patient management, providing both morphologic and functional information[26]. However, MRI tumor information is qualitative, and results depend on the radiologist experience. While the post-contrast sequences are used for the characterization of lesion behavior, the pre-contrast images give mainly morphologic information[26,61]. Focal edema appears as high signal intensity on T2-weighted (T2- WI) MRI sequences and, based on its location, can be classified in three different types: peritumoral, prepectoral, and subcutaneous edema[59].

To our knowledge, there are no articles associating tumor edema Radiomics with primary tumor imaging information in order to predict breast cancer biological aggressiveness.

Our aim is, therefore, to evaluate the contribution of edema associated with clinical and MRI features for breast cancer histological prognostic factors prediction, using T2- weighted MRI radiomics.

3.1.1 Population

In this study all breast cancer MRI examinations performed at the Department of Radiological Sciences, Sapienza University in Rome, for local staging, from January 2015 to January 2019, were retrospectively reviewed. A total of 160 patients, with 164 histologically proven invasive BC lesions, were enrolled.

The following were considered as inclusion criteria: staging 3T MRI examination; presence of Dynamic Contrast-Enhanced MRI (DCE) T2-WI and DWI sequences; presence of focal or diffuse edema in the T2-WI sequences, histopathologic diagnosis of invasive BC; complete histologic analysis, including molecular receptor evaluation (estrogen receptor ER, progesterone receptor PgR; epidermal growth factor receptor HER2) and Ki-67 index calculation.

Exclusion criteria were: presence of breast implants, post-chemotherapy follow-up patients, neo-adjuvant chemotherapy, and images that were not of good diagnostic quality. Patients clinical data (age, menopausal state, family history, hormone therapy), tumor MRI-features (location, margins, dimensions, morphology, kinetic curves, edema type), and histological features (histological type, grading, ER, PR, HER2, and Ki-67 index) were collected.

Lymph node tumoral involvement for each patient was also recorded, using definitive surgical characterization, as a dichotomous result: positive, if there was at least sentinel LN involved by BC metastasis, or negative, if there was no lymph node metastasis.

Patients' clinical data and tumor MRI-features were used as "semantic features" while the histological factors were used as labels to be predicted for the algorithm.

3.1.1.1 MRI examination

All MRI exams were performed on a 3T magnet (Discovery 750; GE Healthcare, Milwaukee, WI, USA). Patients were positioned prone, and a dedicated eight-channel breast coil (8US

TORSOPA) was used. Three orthogonal localizer sequences were employed, then images were acquired following this protocol:

- T2-weighted axial single-shot fast spin echo sequence with a modified Dixon technique (IDEAL) for intravoxel fat-water separation (TR/TE 3500–5200/120–135 ms, matrix 352 × 224, FoV 370 × 370, NEX 1, slice thickness 3.5 mm).
- Diffusion weighted axial single-shot echo-planar with fat suppression sequence (TR/TE 2700/58 ms, matrix 100 × 120, FOV 360 × 360, NEX 6, slice thickness 5 mm) with diffusion-sensitizing gradient applied along the three orthogonal axes and with a b-value of 0, 500, and 1000 s/mm^2.
- T1-weighted axial 3D dynamic gradient echo sequence with fat suppression (VI-BRANT) (TR/TE 6.6/4.3 ms, flip angle 10°, matrix 512 × 256, NEX 1, slice thickness 2.4 mm), before and five times after intravenous contrast medium injection.
- Current guidelines suggest at least three time points to measure during the post-contrast-phase: one before the administration of contrast medium, one approximately 2 min later to capture the peak, one in the late phase. This allows us to evaluate whether a lesion continues to enhance or is characterized by contrast agent wash-out. At least two measurements after contrast medium administration are recommended, even if the optimal number of repetitions is unknown. In our center, we usually perform five acquisitions after contrast medium administration ensuring to obtain a specific signal intensity curve time without penalizing the duration of the examination.

Gadobenate-dimeglumine (Multihance®; Bracco Imaging, Milan, Italy) was administered at a concentration of 0.2 mmol/kg and injected intravenously (20 G cannula) at a rate of 2 mL/s via an automatic injector; this was followed by infusion of 15 mL of saline at the same rate. In post-processing, subtracted images were automatically produced from the images after contrast medium administration for a more accurate tumor analysis.

For each lesion the following MRI characteristics were collected using DCE-sequences:

- Location on the breast quadrant;

- Margins: regular, irregular, lobulated, spiculated or non-mass;
- Dimensions (mm);
- Morphology: round, oval, or irregular;
- Contrast enhancement, quantified using the signal intensity/time curve: type I, characterized by a slow wash-in and without wash-out, type II, defined by a plateau curve after a rapid/slow wash-in, and type III, with rapid wash-in and rapid wash-out;
- Associated-tumor edema type

3.1.1.2 Edema evaluation at MRI imaging

The 164 lesions were divided according to the presence or absence of edema (0). When present, edema was classified as peritumoral (1), pre-pectoral (2), subcutaneous (3), or diffuse (4). Two example of edema types are reported in Figure 1.

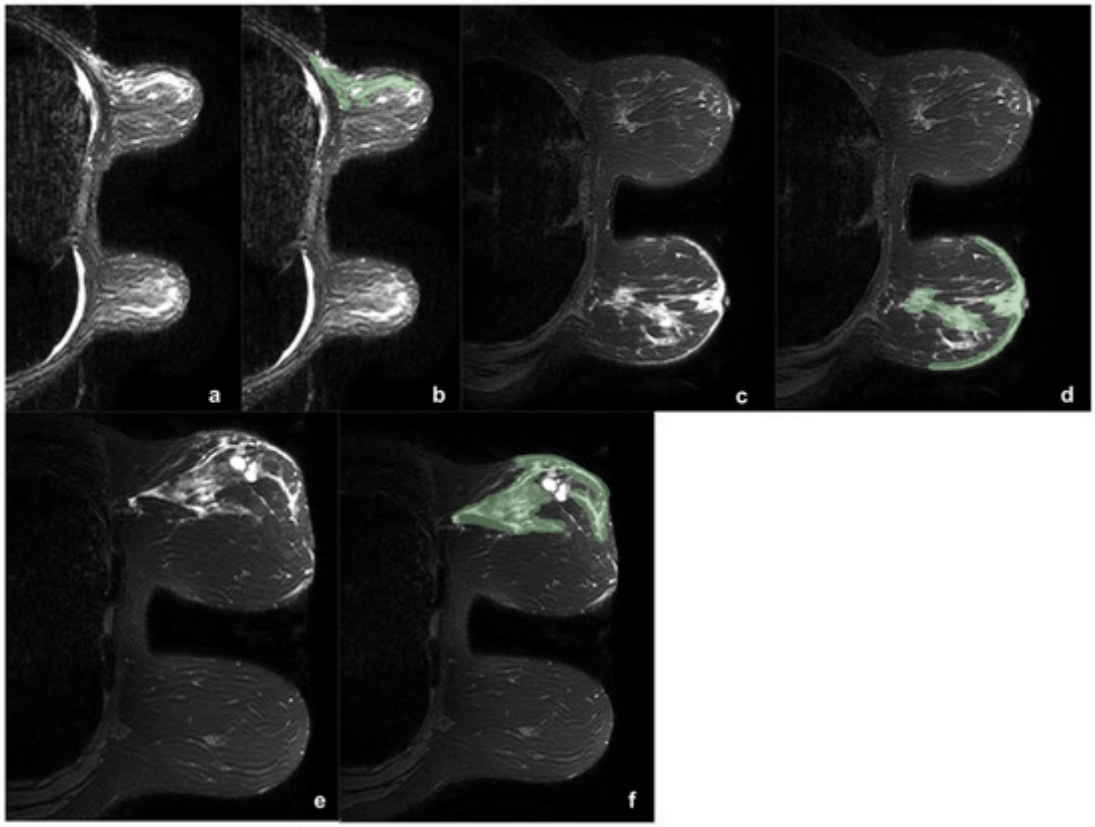

Figure 1. Three example of breast tumor associated edema before (a,c,e) and after (b,d,f) the segmentation process (green).
In the first case (a,b) was reported a case of upper-external quadrant CDI, LUMINAL B, G2 tumor, with pre-pectoral and peritumoral edema. In the second case (c,d) was shown a case of central CDI, LUMINAL B, G3 tumor, with peritumoral and sub-cutaneous edema. It is interesting to note as in the second case the comparison with the contralateral breast helps in the edema definition. The third case (e,f) shows a CDI, LUMINAL B, G3 tumor, with diffuse edema. During the segmentation the two hyperintense areas with regular margins were identified as cysts and accurately avoided during the ROI definition.

Only lesions associated with edema were eligible for this study. Edema was identified as high signal intensity on T2-weighted sequences. However, edema may have relatively similar signal to tumor lesions and breast tissue in the conventional T1-weighted images and the T2-fat-suppression sequences. Therefore, the DCE and DWI sequences were also used, allowing the best definition between tumor and edema. In addition, a comparison between the two breasts

was exploited in order to distinguish breast edema, in the breast side with tumor, from breast glandular tissue, in the breast side without tumor.

3.1.1.3 Histological characteristics

Breast tissue specimens obtained after biopsy or surgery were analyzed by a pathologist with more than 20 years of experience. Histological diagnosis was performed following the WHO classification: the histopathological grade was evaluated according to NGS (Nottingham Grading Score), through a scoring system which evaluate tubule formation, pleomorphism, and mitotic count. The total numerical score ranges from 3 to 9. A score of 3–5 corresponds to grade 1 (G1), a score of 6 or 7 was interpreted as grade 2 (G2), and a total score of 8 or 9 leads to diagnosis of grade 3 (G3).

Immunohistochemical (IHC) analysis was performed to evaluate molecular receptors status (ER, PgR, and HER2) and to calculate Ki-67 index. Only nuclear reactivity was considered for ER. The monoclonal antibody Mib-1 (1:200 dilution; Dako, Glostrup, Denmark) was used to assess the Ki-67 index, which was reported as the percentage of immune-reactive cells out of 2000 tumor cells in randomly selected high-power fields surrounding the tumor core. HER2 status was re-evaluated using the Hercep test (Dako, Glostrup, Denmark). Samples that gave an equivocal IHC result were subjected to fluorescence in situ hybridization (FISH) analysis. A ratio of HER2 gene signals to chromosome 17 signals greater than 2.2 was used as a cut-off value to define HER2 gene amplification. ER and PgR status was considered positive if expression was ≥1% and negative if expression was <1%. HER2 expression was graded as follows: 0, 1+, 2+, or 3+; tumors with a score greater or equal than 2+ were considered HER2 positive, whereas scores lesser than 2+ were considered HER2 negative. Ki-67 expression ≥14% was considered positive and <14% was considered negative.

3.1.1.4 Statistical analysis

Descriptive statistics were accomplished using the statistical software program SPSS© version 25.0. Statistical significance was set at $p < 0.05$.

Spearman's Rank-Order Correlation was evaluated to assess whether there was a correlation between the edema types and categorical variables, both clinical (menopausal status, hormone therapy, family history), MRI (location, margins, dimensions, morphology, kinetic curves, edema type), histological characteristics (histologic type, grading, expression of ER, PgR, and HER2b and Ki-67 index) and lymph node status (positive or negative). We also performed Spearman's Rank-Order Correlation to assess a possible correlation between edema types and positive lymph node status, excluding lesions with peritumoral edema.

The Kolmogorov-Smirnov test was performed to determine whether age and tumor size followed a normal distribution.

Statistical comparison of edema type and age and tumor size was performed using the Kruskal-Wallis H test.

3.1.2 Methodology

3.1.2.1 Radiomics Approach

Edema segmentation

Each case has been anonymized and identified with a progressive identification number (ID). For the analysis of bilateral tumors, lesions were considered one at a time with different IDs. For each breast sample the T2-weighted sequence was selected and loaded into a workstation. To get the region of interests (ROIs), a semi-automatic segmentation of the edema was performed by an expert radiologist, and then proofread by another one, with 3D Slicer (http://www.slicer.org, accessed date: 1 January 2015). All the following analysis were run using home-made algorithms on MATLAB. Segmentation was performed at first in the axial

projections, avoiding tumor lesions and necrosis when present. Each segmentation was improved in the sagittal and coronal projections and finally optimized using three-dimensional version of the well-known convex-hull algorithm.

Indeed, although all tumor edemas usually present a spherical and irregular geometry, when considering an MR acquisition it is easy to incur in some tissues that are partially difficult to segment, either because of tissue stretching due to the imaging phase, or because of hard-to-detect shapes within the images when edemas are close to other tissues with similar density, such as cysts or pure dense glandular tissue, thus often resulting in jagged edges and sometimes concavities in the resulting ROI. On this ground, the calculation of the three-dimensional convex hull serves a twofold purpose: firstly, it removes all possible segmentation deformities otherwise hindering further calculation of textural features and secondly, it can slightly expand the ROIs thus also containing a perimeter portion of healthy tissue at the edge of the segmentation that might carry information about the tumor status.

Feature Extraction and Selection

A total of 253 features were extracted and filtered, selecting the most informative through a Best Firsts wrapper algorithm, and using a Random Forest and the labels we would like to predict. We performed five experiments, by using the different histological labels set with dichotomic values (0–1): histology, grading, ER, PgR, and Ki-67. All the experiments will be detailed in the next section.

Furthermore, it is worth noting that the wrapper-based feature selection was applied avoiding any bias between the training and test set and we used all default Matlab's parameters for the classification.

The classification stage straightforwardly employed the same learning paradigm used in the feature selection phase; note also that the use of a Random Forest alleviates the curse of dimensionality and permits us to benefit by the aggregation of several decision trees, mimicking the consensus mechanism among different experts. This configuration was applied

using only semantic features (i.e., patient clinical data and features derived by visual inspection of the MR images), and semantic features associated to edema radiomics descriptors.

We preferred to manually implement the code in order to better control specific computational aspects, usually bulked into external software, considered when generalizing features generally extracted on planar images to a three-dimensional volume. All the steps to build our code were tested against well-established packages.

A total of 253 features were extracted, of which 11 were semantic (age, menopausal status, family history, hormone therapy, MRI location, stadiation, margins, dimensions, morphology, kinetic curves, edema type) and 241 were radiomics. In order to fully take advantage of gathered acquisitions, no data nor gray level reduction was performed to the images and all analysis were intrinsically computed in the three-dimensional voxel space. It worth noting that although the relatively large slice thickness of the MRs could alter the isotropy of the acquired stacks, all the features we extracted which are sensitive to the voxels texture orientation were computed for all the existing three-dimensional orientations and then filtered out, in the further feature selection step, all the less informative ones. In detail, there were 11 first-order features (mean, standard deviation, skewness, kurtosis, energy, entropy, maximum absolute value, position of the maximum absolute value, energy around the maximum absolute value, range, number of maximum relative values, energy around the maximum relative values), 48 s-order features extracted from the histograms of four variations of the three Orthogonal Planes-Local Binary Patterns (TOP-LBP) (the same feature extracted from the first-order group, per each variation), and 182 features extracted from the 3D version of the Gray Level Co-occurrence Matrix (GLCM) (autocorrelation, covariance, inertia, absolute of the inertia, inverse of the inertia, energy and entropy, per each of the 26 possible three-dimensional directions). A summary of such features is reported in Table 1.

	Features *n*	Parameters	Description
Semantic features	11	-	Clinical (age, HT, family history, menopausal state) and imaging (location, stadiation, margins, dimensions, morphology, kinetic curves, edema type)
First-order	12	# bins = 2^{16}	All these features base on the count of the voxels in a ROI and therefore on the associated histogram computed on such count
Second-order TOP-LBP	48	Radius = 1 # neighbors = 8	These features attempt to extract the shape's pattern of tumour inside a ROI analysing the neighborhood of each voxel
Second-order GLCM	182	Interpixel distance = 1	These are the multi-dimensional generalization of the histogram and aim to determine the tissue's orientation inside a ROI

Table 1: Description of the extracted features.
#: "number of"

Classification

As mentioned before, the labels have always been assumed dichotomic values of 0 and 1: for histology (ductal or lobular invasive carcinoma), for ER and PgR status (≥ or <10%), for Ki67 proliferation index (≥ or <1%), and also for grading (G1 + G2 vs. G3), grouping the class with the lowest number of elements (G1 in the specific case). The dichotomic labeling approach was chose in order to avoid a class unbalance and to ease and standardize the classifier approach.

Each of the five experiments was performed in ten-fold cross-validation and repeated for each histological label to predict, as detailed in the Figure 2.

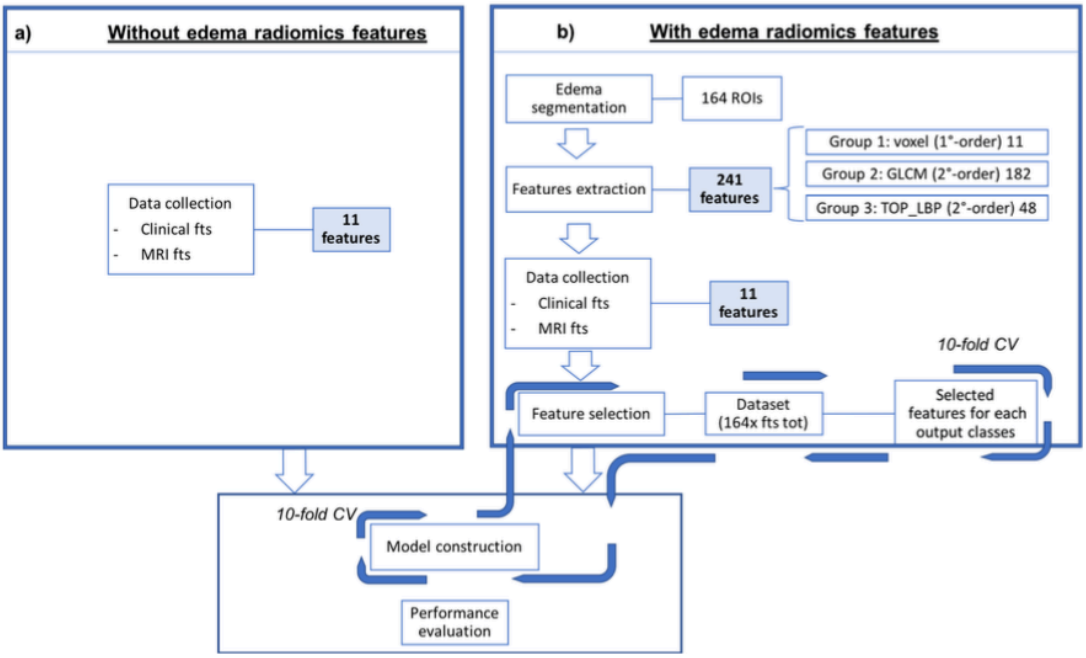

Figure 2: Schematic representation of the proposed approach to predict the five histological labels. In the left box (a) is reported the classic medical approach, using only the semantic features, while in the right box (b) the classification was performed associating the radiomics features to the semantic ones. The model is then built using the best performing features.

3.1.3 Results

In this study, 164 breast cancer lesions with 127 presenting edema tumor were used. Four patients had bilateral breast cancer, but only one patient with bilateral BC presented edema. Edema types were represented as follows: 62 (49%) peritumoral, 26 (20%) pre-pectoral, 16 (13%) subcutaneous, 23 (18%) diffuse. The history-clinical data of the study population were collected. The mean age of the patients presenting edema was 54.86 years (range 30–84). Seventy-two (n = 72, 45.5%) patients were pre-menopausal and 88 (55.5%) were postmenopausal; 120 (73%) patients did not have any family member affected by breast cancer, 32 (20%) patients had one family member, and 8 (5%) patients had at least 2 female or male family members affected by breast cancer at any age.; 9 (6%) patients assumed hormone

therapy for at least 3 continuous months, whether for contraceptive, substitution, or medical therapy reasons, and 151 (94%) patients did not assume any hormone therapy.

The median diameter of the lesions measured was 19 mm (range 9–60 mm).

Analyzing tumor location, peritumoral edema was more present when the tumor was located at the upper-outer quadrant (30.6%), pre-pectoral was more frequent when the tumor was in the internal quadrant (upper 12% and lower 46%), subcutaneous edema was predominant for lesions located at the lower-inner quadrant (21%), while the diffuse class was more frequent in case of central tumors (25%).

The lesions were more frequently round or oval in peritumoral edema (56%), irregular in case of subcutaneous and pre-pectoral edema (36.4% and 41.8% respectively) and non-mass like in case of diffuse edema (30.6%).

The other main clinical, imaging, and histological characteristics of the 126 patients presenting edema, divided basing on edema type, are shown in Table 2.

			Edema Type				Total	*p*-Value
			Peritumoral	Pre-Pectoral	Subcutaneous	Diffuse		
Family History	None	*n*	37	17	14	18	86	0.0250
		%	29.1%	13.4%	11.0%	14.2%	67.7%	
	1	*n*	18	8	2	4	32	
		%	14.2%	6.3%	1.6%	3.1%	25.2%	
	>1	*n*	7	1	0	1	9	
		%	5.5%	0.8%	0.0%	0.8%	7.1%	
Hormone Therapy	None	*n*	58	26	15	19	118	0.267
		%	45.7%	20.5%	11.8%	15.0%	92.9%	
	Positive	*n*	4	0	1	4	9	
		%	3.1%	0.0%	0.8%	3.1%	7.1%	
Menopause	Pre-m	*n*	24	14	5	6	49	0.444
		%	18.9%	11.0%	3.9%	4.7%	38.6%	
	Post-	*n*	38	12	11	17	78	
		%	29.9%	9.4%	8.7%	13.4%	61.4%	
Kinetic Curve	I	*n*	9	2	2	6	19	0.375
		%	7.1%	1.6%	1.6%	4.7%	15.0%	
	II	*n*	32	13	4	7	56	
		%	25.2%	10.2%	3.1%	5.5%	44.1%	
	III	*n*	21	11	10	10	52	
		%	16.5%	8.7%	7.9%	7.9%	40.9%	
Margins	Regular	*n*	3	0	1	0	4	0.746
		%	2.4%	0.0%	0.8%	0.0%	3.1%	
	Irregular	*n*	28	14	11	12	65	
		%	22.0%	11.0%	8.7%	9.4%	51.2%	
	Lobulated	*n*	7	5	0	2	14	
		%	5.5%	3.9%	0.0%	1.6%	11.0%	
	Spiculated	*n*	21	6	4	5	36	
		%	16.5%	4.7%	3.1%	3.9%	28.3%	
	Non-mass	*n*	3	1	0	4	8	
		%	2.4%	0.8%	0.0%	3.1%	6.3%	
Histology	IDC	*n*	54	23	16	17	110	0.513
		%	42.5%	18.1%	12.6%	13.4%	86.6%	
	ILC	*n*	8	3	0	6	17	
		%	6.3%	2.4%	0.0%	4.7%	13.4%	
Grade	1	*n*	11	1	0	2	14	<0.001 *
		%	8.7%	0.8%	0.0%	1.6%	11.0%	
	2	*n*	33	8	6	6	53	
		%	26.0%	6.3%	4.7%	4.7%	41.7%	
	3	*n*	18	17	10	15	60	
		%	14.2%	13.4%	7.9%	11.8%	47.2%	
LNS	Negative	*n*	53	23	14	14	104	0.064
		%	41.7%	18.1%	11.0%	11.0%	81.9%	
	Positive	*n*	9	3	2	9	23	
		%	7.1%	2.4%	1.6%	7.1%	18.1%	
ER Status	Negative	*n*	7	8	5	6	26	0.029 *
		%	5.5%	6.3%	3.9%	4.7%	20.5%	
	Positive	*n*	55	18	11	17	101	
		%	43.3%	14.2%	8.7%	13.4%	79.5%	

PR Status	Positive	*n*	16	11	8	11	46	0.018 *
		%	12.6%	8.7%	6.3%	8.7%	36.2%	
	Negative	*n*	46	15	8	12	81	
		%	36.2%	11.8%	6.3%	9.4%	63.8%	
HER2 Status	Negative	*n*	58	24	14	15		0.003 *
		%	45.7%	18.9%	11.0%	11.8%		
	Positive	*n*	4	2	2	8		
		%	3.1%	1.6%	1.6%	6.3%		
Ki-67	<20%	*n*	30	6	1	6	43	0.004 *
		%	23.6%	4.7%	0.8%	4.7%	33.9%	
	>20%	*n*	32	20	15	17	84	
		%	25.2%	15.7%	11.8%	13.4%	66.1%	

Table 2 Clinical, histologic, and MRI characteristics of the patients presenting with breast edema, classified according to the type of edema
(* means the correlation is statistically significant).
HT: hormone therapy, IDC: Invasive Ductal Carcinoma, ILC: Invasive Lobular Carcinoma, LNS: Lymph Node Status.

Using Spearman's Rank test no correlation we did not found between edema type and menopausal status, hormone therapy, lesion margins, kinetic curve, histological type, and positive lymph node status (p value > 0.5). A significant correlation was found between tumor histologic class, grading, ER, PgR and HERb2 expression, and Ki-67 index. A significant correlation was found between edema type and positive lymph node status, when peritumoral edema was excluded (p = 0.022). The Kruskal-Wallis H test demonstrated a significant correlation between edema class and lesion size (p value < 0.001). No correlation was found between patient age and edema type.

Radiomics Approach

After the feature extraction a signature, composed by the most significative features that lead to the final results, was individuated for each histological label (histological class, ER and PgR status, Ki-67 index and grading). The selected features and their predictive impact for each histological label were reported in the following charts (Figure 3).

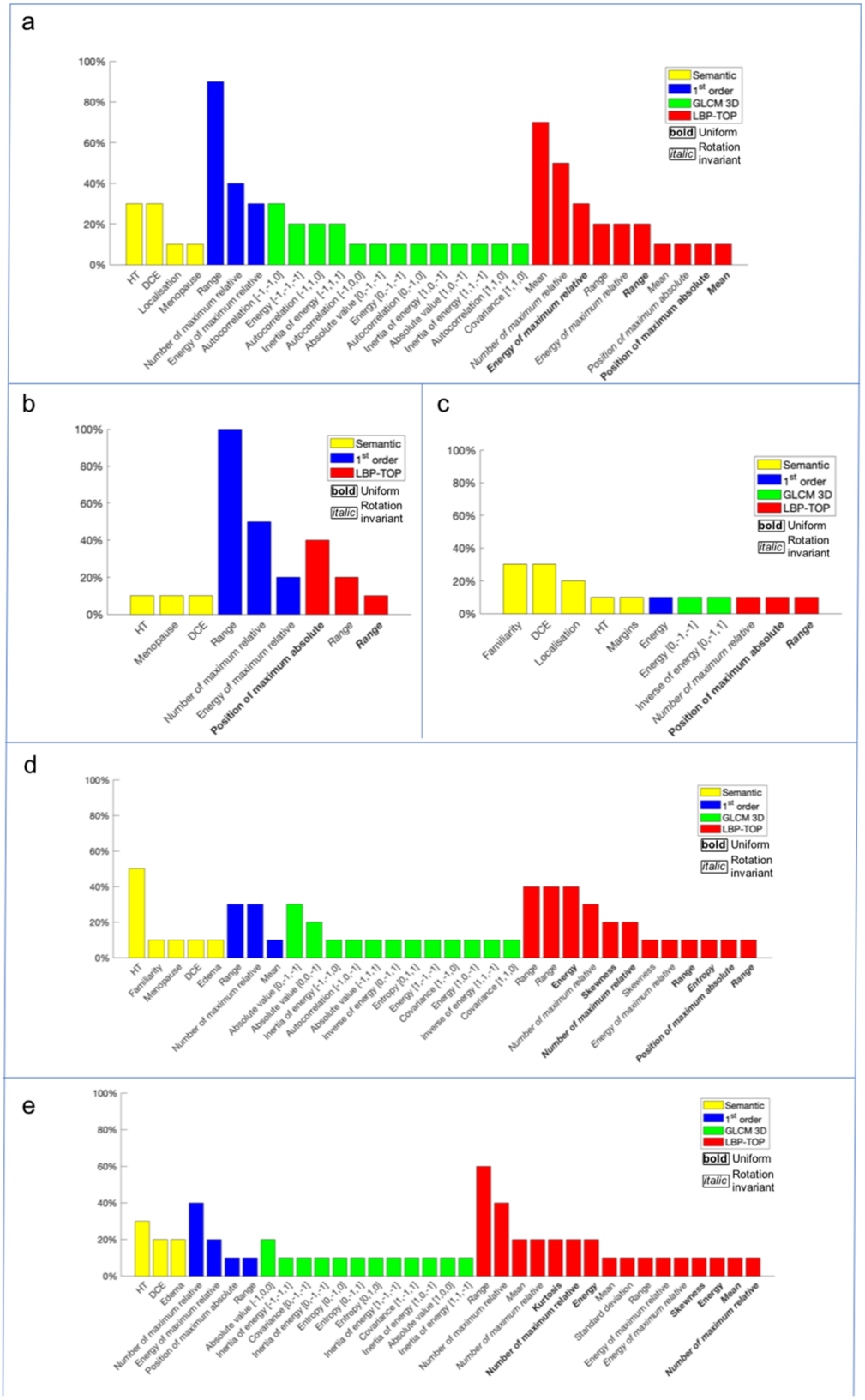
a
100%
80%
60%
40%
20%
0%
Semantic
1st order
GLCM 3D
LBP-TOP
bold Uniform
italic Rotation invariant
HT
DCE
Localisation
Menopause
Range
Number of maximum relative
Energy of maximum relative
Autocorrelation [-1,-1,0]
Energy [-1,-1,-1]
Autocorrelation [-1,1,0]
Inertia of energy [-1,1,1]
Autocorrelation [-1,0,0]
Absolute value [0,-1,-1]
Energy [0,-1,-1]
Autocorrelation [0,-1,0]
Inertia of energy [1,0,-1]
Absolute value [1,0,-1]
Inertia of energy [1,1,-1]
Autocorrelation [1,1,0]
Covariance [1,1,0]
Mean
Number of maximum relative
Energy of maximum relative
Range
Energy of maximum relative
Range
Mean
Position of maximum absolute
Position of maximum absolute
Mean
b
100%
80%
60%
40%
20%
0%
Semantic
1st order
LBP-TOP
bold Uniform
italic Rotation invariant
HT
Menopause
DCE
Range
Number of maximum relative
Energy of maximum relative
Position of maximum absolute
Range
Range
c
100%
80%
60%
40%
20%
0%
Semantic
1st order
GLCM 3D
LBP-TOP
bold Uniform
italic Rotation invariant
Familiarity
DCE
Localisation
HT
Margins
Energy
Energy [0,-1,-1]
Inverse of energy [0,-1,1]
Number of maximum relative
Position of maximum absolute
Range
d
100%
80%
60%
40%
20%
0%
Semantic
1st order
GLCM 3D
LBP-TOP
bold Uniform
italic Rotation invariant
HT
Familiarity
Menopause
DCE
Edema
Range
Number of maximum relative
Mean
Absolute value [0,-1,-1]
Absolute value [0,0,-1]
Inertia of energy [-1,-1,0]
Autocorrelation [-1,0,-1]
Absolute value [-1,1,1]
Inverse of energy [0,-1,1]
Entropy [0,1,1]
Energy [1,-1,-1]
Covariance [1,-1,0]
Energy [1,0,-1]
Inverse of energy [1,1,-1]
Covariance [1,1,0]
Range
Range
Energy
Number of maximum relative
Skewness
Number of maximum relative
Skewness
Energy of maximum relative
Range
Entropy
Position of maximum absolute
Range
e
100%
80%
60%
40%
20%
0%
Semantic
1st order
GLCM 3D
LBP-TOP
bold Uniform
italic Rotation invariant
HT
DCE
Edema
Number of maximum relative
Energy of maximum relative
Position of maximum absolute
Range
Absolute value [-1,0,0]
Inertia of energy [-1,1,-1]
Covariance [0,-1,-1]
Inertia of energy [0,-1,-1]
Entropy [0,-1,0]
Entropy [0,0,1]
Entropy [0,1,1]
Inertia of energy [1,-1,-1]
Covariance [1,-1,-1]
Inertia of energy [1,0,-1]
Absolute value [1,0,0]
Inertia of energy [1,1,-1]
Range
Number of maximum relative
Mean
Number of maximum relative
Kurtosis
Number of maximum relative
Energy
Mean
Standard deviation
Range
Energy of maximum relative
Energy of maximum relative
Skewness
Energy
Mean
Number of maximum relative

Figure 3: Features selection
The charts show the most significant features, selected by a wrapper filter, for each histological label: (a) grading, (b) ER, (c) histological class, (d) Ki-67 and (e) PgR. The features were individuated by a color based on the specific feature class (yellow for semantic features, i.e., the 11 features collected without considering the edema information, blue for radiomics first-order features, green for radiomics GLCM 3D second-order features and red for radiomics LBP 3D second-order features).

For the sake of completeness, we would like to stress the fact that in order not to incur in the issue known as Curse of Dimensionality, i.e., when during a classification task the number of features is empirically higher than 10% of the samples present in the dataset, thus biasing the predictions, after the feature selection phase we should have concluded with at least 10 edemas in the minority class per feature, and instead the number of selected features exceeds the number of analyzed samples in all the classes considered. In fact, after the feature selection we accepted a certain abundance of extra features without proceeding with a deeper filtering, due to the fact that the algorithm considered in the next step, the Random Forest, is quite capable of dealing with a larger number of attributes than the relevant ones, performing an additional intrinsic selection of only the most significant features and avoiding bias.

The difference between the AUC and the accuracy obtained to predict each of the five histological prognostic labels, with and without edema radiomics features contribution, is reported in Table 3.

	Without Edema	With Edema	Difference
Histology	AUC: 0.520 Accuracy: 85.8% Sensibility: 100% Specificity: 5.6% PPV: 85.7% NPV: 100%	AUC: 0.645 Accuracy: 64.17% Sensibility: 64.7% Specificity: 61.1% PPV: 90.4% NPV: 23.4%	AUC: +0.125 * Accuracy: −21% Sensibility: −35.3% Specificity: +55.5% * PPV: +4.7% * NPV: −76.6%
Grading	AUC: 0.590 Accuracy: 90% Sensibility: 0% Specificity: 100% PPV: 0% NPV: 90.8%	AUC: 0.789 Accuracy: 90.8% Sensibility: 36.4% Specificity: 96.3% PPV: 50% NPV: 93.8%	AUC: +0.199 * Accuracy: +0.8% * Sensibility: +36.4% * Specificity: −3.7% PPV: +50% * NPV: +3% *
ER	AUC: 0.466 Accuracy: 72.5% Sensibility: 0% Specificity: 92.6% PPV: 0% NPV: 77%	AUC: 0.487 Accuracy: 81.7% Sensibility: 23.1% Specificity: 97.9% PPV: 75% NPV: 82.1%	AUC: +0.021 * Accuracy: +9.2% * Sensibility: +23.1% * Specificity: +5.3% * PPV: +75% * NPV: +5.1% *
PR	AUC: 0.546 Accuracy: 55% Sensibility: 35.4% Specificity: 68.1% PPV: 42.5% NPV: 61.3%	AUC: 0.659 Accuracy: 61.7% Sensibility: 54.2% Specificity: 66.7% PPV: 52% NPV: 68.6%	AUC: +0.113 * Accuracy: +6.7% * Sensibility: +18.8% * Specificity: −1.4% PPV: +9.5% * NPV: +7.3% *
Ki-67	AUC: 0.573 Accuracy: 59.2% Sensibility: 17.1% Specificity: 81% PPV: 31.8% NPV: 65.3%	AUC: 0.621 Accuracy: 64.2% Sensibility: 34.1% Specificity: 79.7% PPV: 46.7% NPV: 70%	AUC: +0.048 * Accuracy: +5% * Sensibility: +17% * Specificity: −1.3% PPV: +14.9% * NPV: +4.7% *

Table 3: Experimental results
The AUC and the accuracy for each histological label are obtained, at first, using only the semantic features (second column) and, then, adding the edema radiomics features (third column). In the last column is reported the difference between the two AUC/accuracies, indicating with * the cases in which edema radiomics features increased the result.

3.1.4 Discussion and Conclusions

Among the different imaging methods, MRI has demonstrated an increasingly important role as a breast cancer management guide, from BC diagnosis and loco-regional stadiation, to aiding in therapy decision-making[5,27]. High-field MRI contrast resolution allows a very accurate definition of the main morphological and functional tumor characteristics.

In recent years, in the literature there have been preliminary studies that have attempted to seek a correlation between MRI tumor characteristics with BC molecular subtypes and prognostic factors using artificial intelligence (AI). Radiomics, which represents the main AI medical application, consists of the analysis of medical images and aims to convert the relative

information and characteristics, called features, that the observer is not able to see by themselves, into quantitative and measurable high-data functions.

MRI post-contrast sequences have proved to be the main sequences for the phenotypic description of the lesion, characterization of the true tumor extension, margins, morphology and, almost importantly, definition of tumor neoangiogenesis[39]. On the other hand, the pre-contrast sequences, T2-weighted and DWI, are the source of morphological information and are used to improve tumor aggression characterization, contributing with data about the degree of peri-tumor inflammation.

Edema, identified as a high-intensity signal in T2-weighted sequences, can therefore be useful, although non-specific, data, correlating with peritumoral malignant lesion spread and lymphangioinvasiveness[59,60,62]. Edema can be classified into peritumoral, pre-pectoral, subcutaneous, and diffuse, presenting a different degree of correlation with tumor aggressiveness. Both malignant and benign lesions, such as mastitis, previous radiotherapy, post-surgical inflammation, nephrotic syndrome, lymphoma, venous congestion, and chronic heart disease can be a cause of diffuse breast edema[63]. Focal edema, studied in T2WI sequences, however, is associated with malignant lesions in most cases[59,64–67].

Peritumoral edema is caused by the tumoral process of neoangiogenesis, resulting in increased vascular permeability in the newly formed vessels and the release of peritumoral cytokines. This is specific of invasive breast cancer, even if it is less frequent in lobular cancer than in ductal cancer [68]. This could be due to the low density of the lesion and the growth pattern of lobular carcinoma. Peritumoral edema is also significantly related to rim enhancement[69].

Breast tissue is drained primarily by lymphatic vessels leading to the axillary lymph nodes[70,71]. Invasive BC causes lymphatic vessels to increase in number and size, consequently increasing the risk of lymph node metastasis. When the main axillary lymphatic drainage is blocked due to carcinoma, collateral intramammary and pre-pectoral lymphatic drains take over[59,67,70–72]. Pre-pectoral edema is related to tumor cells in the retro-mammary areas[27,67,70]. When lymphatic drainage in the subcutis is blocked due to tumor emboli, subcutaneous edema will occur[67,72]. In theory, pre-pectoral edema usually precedes subcutaneous edema as it is considered by

Uematsu et al. as “the final stage of breast edema associated with malignancy”. During imaging studying, it is therefore useful to clarify the type and the quantity of edema observed, as this finding can help in predicting the malignancy of the lesion.

On this basis, we wanted to explore edema over-reader information, with the aim to analyze features which can be added as imaging breast cancer descriptors to the traditional radiologist evaluation, by combining information from radiomics, based on T2-weighted 3T MRI sequences, with the patient clinical and traditional imaging data of the lesion to predict the histological characteristics of the tumor.

To our knowledge there are no articles that associate tumor edema radiomics to tumor imaging in order to improve breast cancer biological aggressiveness.

In our study, the class of edema correlates significantly with tumor size, histological class, grading and expression of ER, PgR and HER2b, and Ki-67 index, reflecting how the presence of edema is associated with particularly aggressive tumors. As already hypothesized by Baltzer et al. invasive tumor growth and progression are associated with phenomena of tumor proteolysis and neoangiogenesis; these factors contribute to an increased vascular permeability, because of basal membranes that perform less than in the physiological vasculature [**35**]. Tumor growth and progression are nothing but the expression of tumor aggressiveness, which is in turn expressed at histological level by the nuclear grade, mitotic growth index, and receptor expression of the tumor lesion.

We did not find a significant correlation between the class of edema and the presence of lymphadenopathy, when considering all the edema types. When considering the lesions with pre-pectoral, subcutaneous, and diffuse edema, a significant correlation was found, confirming Uematsu’s studies. Uematsu describes the classes of pre-pectoral and subcutaneous edema as signs of lymphovascular invasion (LIV), which is responsible for disruption of lymphatic drainage in the dermal and subdermal area due to tumor emboli: LIV is significantly associated with pre-pectoral edema, and subcutaneous edema which is its final stage, and both are associated with inflammatory breast cancer and occult inflammatory breast cancer [**10**].

Regarding the radiomics impact, our results show consistently higher AUC using breast edema radiomics features. The best improvement in AUC was obtained regarding grading (0.789 vs. 0.590) while estrogen- and progesterone-receptors showed a lower increase of AUC (ER: 0.487 vs. 0.466; ER: 0.659 vs. 0.546). The improvements can be confirmed by looking at the values of sensitivity, specificity, PPV, and NPV of each experiment. In fact, the percentages almost always show higher values when considering predictions obtained using edemas when calculating features rather than not, only in a few comparisons we have that specificity shows a lower result as well as a single NPV, suggesting a more conservative trend for the classifier using edema features. These results can probably be explained as edema is an imaging manifestation of advanced or inflammatory BC, while hormone-receptors are histological features. The only case in this study where the accuracy decreases is in case of histological type. This can be explained by the imbalance of our sample (108 CDI vs. 17 CLI).

It is also interesting to note that only some semantic features survived to the features selection process, such as the DCE-Kinetic Curve, which represents the tumor intrinsic vascularization, closely connected with the lesion aggressiveness and edema, as strongly reported in literature [7,8,26,59–62,64,73–76]. However, for all the histological labels, the majority of selected features were radiomics, and in particular second-order ones. This result suggests that, for this experiment, most of the information resides within the tissue pattern and the micro-structure of the tumor contours, which are the properties analyzed by these feature families. The spread of tumor-associated edema follows specific patterns related to tumor aggressiveness, and this can explain this class features selection.

The use of a 3Tesla magnet allows the generation of high-quality images with high information content. Our study collected images with high data content to be exploited to obtain features that are not only numerous, but also highly detailed in order to achieve the maximum correlation with the histological microstructure and facilitate the radiomic path. This is possible based on the greater detail provided by high-resolution images, as demonstrated by various studies performed on high-field MRI[77,78].

High resolution image production guarantees an excellent definition of the details allowing a selection of most appropriate features, but also optimizes the work of the classifier, both from a computational point of view, streamlining the amount of information to be evaluated, and the result, ensuring greater appropriateness of the data.

It has emerged over the years that the application of Artificial Intelligence, even in other fields, requires very advanced computational sciences and statistics. Moreover, radiomics needs an incredibly collaborative inter-disciplinary, inter-institutional team[79].

The main limitations of this study include the small sample size and the single center nature of the study, the unbalanced number of the group component (smaller number of the minority class), the manual segmentation, leading to time-consumption and user error and variability, the large number of features utilized which may lead to overparameterization, the lack of an independent test set, and the lack of an assessment of inter-reader variation. A larger and preferably multi-center cohort is needed for a more rigorous analysis. Moreover, an automatic, standardized, and validated segmentation method would be ideal even if not yet available at present. It is certainly necessary to keep in mind the inherent limits of radiomics: this technique is currently still tied to the radiologist's ability to select ROIs and the engineer's ability to build the algorithm capable of feature selection and their experience.

However, the results obtained in this preliminary study, as well as those published so far in the literature, are highly promising. Considering the significant clinical impact that the appropriate use of these tools could offer the physician, the growing research in this area is justified. There is the need to produce common datasets in the public domain, in order to make the collected data usable and to standardize the protocols.

This study demonstrates that the BC edema radiomic features of preoperative 3 Tesla MRI extracted from the T2 sequences significantly correlate with the main histological prognostic factors. Despite its currently marginal role in guidelines as a predictor, the breast associated edema may be useful in the preliminary phase to provide additional information on tumor bio-histological aggressiveness. This concept is more evident by applying Artificial Intelligence algorithms to images that can provide objective quantitative data to the limited vision of the

radiologist. Radiomics is a complex tool in its implementation phase but simple in its use phase. This will inevitably lead to its greater development in the medical field and this study represents only a primordial evaluation of what AI can offer as a decision supporting tool.

3.2 DWI-sequences and ADC prognostic value

The highly heterogeneous nature of breast cancer is a consequence of the different histological behaviors of the tumor and the different molecular receptor structures. Based on this heterogeneity, an improvement in disease-specific survival rates was observed through increasingly personalized treatment.

The cellular microenvironment is crucial to define the aggressiveness of breast cancer and can be quantified by cell proliferation and intrinsic cellular changes, such as mitotic count and degree of pleomorphism.

Tumor cellularity is defined by the proportion of tumor cells in the tumor bed: this assessment is performed by a pathologist who estimates the local cellularity of the specimen by comparing the area containing the tumor with the reference standard. Being a semi- quantitative assessment, it is, therefore, subject to inter-rater variability[80].

At the time of this study, cancer cellularity is not employed as a method of further sub-differentiation of breast cancer and its role is reserved in assessing tumor response to neoadjuvant therapy, according to the Sinn, Sataloff, and Miller-Payne Pinder methods of evaluating response to treatment[81–84].

The current limit is due to the different cellularity score obtained from biopsy and definitive surgical specimens. The former, in fact, could be representative of a single portion and not of the whole tumor, distorting the final evaluation.

Diffusion weighted imaging (DWI) is traditionally included in MRI routine protocol improving the specificity of post-contrast-MRI, which is considered the golden standard for tumor detection. DWI restriction is proportional to the degree of movement of water molecules and can be quantified by the apparent diffusion coefficient (ADC) value. In ADC maps, tissues with high cellularity showed lower ADC values[85,86]. Recent literature has demonstrated an inverse correlation between ADC values and histological grade of the tumor, proving its usefulness in identifying high-grade invasive breast cancer prior to the surgery[87,88].

The aim of this study is to define whether ADC values, obtained by 3T staging-MRI, vary with Ki-67 expression, tumor grade and cellularity, the latter obtained both after biopsy and after surgery, to identify whether ADC may represent a non-invasive predictor of aggressive cellular definitive pathophysiology.

3.2.1 Population

In this study, all breast cancer MRI examinations performed at our Department of Radiological Sciences for local staging from January-2010 to September-2019 were retrospectively reviewed. A total of 100 patients with histologically proven invasive BC lesions were enrolled.

The following inclusion criteria were considered: staging 3T-MRI examination, performed after biopsy and before surgery; presence of DCE-MRI, T2-WI and DWI sequences; ADC evaluation of the main lesion for each exam; histopathological diagnosis confirming invasive BC; complete histological analysis including molecular receptor assessment (estrogen receptor ER, progesterone receptor PgR; epidermal growth factor receptor HER2), Ki-67 index, and calculation of cellularity, both at biopsy and on the operative specimen.

Exclusion criteria were: presence of breast implants, post-chemotherapy follow-up examinations, neo-adjuvant chemotherapy and images that were not of good diagnostic quality.

Patients clinical data (age, menopausal state, familiarity, hormone therapy), tumor MRI features (stadiation, localization, margins, kinetic curves, size) and histological features (histological type, grading, ER, PgR, HER2, Ki-67 index, cellularity rate) were collected.

Institutional Review Board approval was not required because this is a retrospective observational study, and only existing information collected from human participants was used and there are no identifiers linking individuals to the data/samples.

All methods and procedures were in accordance with institutional and research committee ethical standards and the 1964 Declaration of Helsinki and its subsequent amendments or comparable ethical standards.

3.2.1.1 MRI examination

All MRI exams were performed on a 3T magnet (Discovery 750; GE Healthcare, Milwaukee, WI, USA). Patients were positioned prone and a dedicated eight-channel breast coil (8US TORSOPA) was employed. Three orthogonal localizer sequences were performed, then the following protocol was acquired:

- T2-weighted axial single-shot fast spin echo sequence with fat suppression (DIXON) (TR/TE 3500–5200/120–135 ms, matrix 352 × 224, FoV 370 × 370, NEX 1, slice thickness 3.5 mm).
- Diffusion weighted axial single-shot echo-planar sequence with fat suppression (TR/TE 2700/58 ms, matrix 100 × 120, FOV 360 × 360, NEX 6, slice thickness 5 mm) with b values of 0, 500 and 1000 s/mm^2.
- T1-weighted axial 3D dynamic gradient echo sequence with fat suppression (VIBRANT) (TR/TE 6.6/4.3 ms, flip angle 10°, matrix 512 × 256, NEX 1, slice thickness 2.4 mm), before and five times after intravenous contrast medium injection.

Current guidelines suggest at least three time points to measure during the post-contrast-phase: one before the administration of contrast medium, one approximately 2 min later to capture the peak, one in the late phase. This allows us to evaluate whether a lesion continues to enhance or is characterized by contrast agent wash-out. At least two measurements after contrast medium administration are recommended, even if the optimal number of repetitions is unknown. In our center, we usually perform five acquisitions after contrast medium administration ensuring obtaining a specific signal intensity curve time without penalizing the duration of the examination.

Gadobenate-dimeglumine (Multihance®; Bracco Imaging, Milan, Italy) was administered as contrast agent (concentration of 0.2 mmol/kg; rate of 2 mL/s) followed by injection of 15 mL of saline. In post-processing, subtracted images were automatically produced from the images after contrast medium administration for a more accurate tumor analysis.

The entire exam for each patient was transferred to a workstation (Advantage Windows Workstation 4.4; GE Medical System, Milwaukee, WI, USA) for post-processing analysis. For a quantitative analysis, ADC values were calculated according to the following equation:

$$ADC = -(1/b)\ln (S0/S1)$$

where b is the diffusion factor, S1 is the attenuated signal (b-value of 1000 s/mm^2) and S0 is the full spin echo signal without diffusion gradient (b-value of 0 s/mm^2), as reported in the literature.

DCE sequences were considered to be reference images for tumor detection and lesion characterization. The largest lesion was considered to be the index lesion and included in the statistical analysis. The greatest axial diameter was submitted to statistical analysis. On the basis of their morphology, lesions were classified into mass tumors with regular, lobulated, irregular and spiculated margins, or non-mass tumors.

For each index lesion, a signal intensity-to-time curve (SI/T) was automatically generated by placing a region of interest (ROI) within the lesion on a subjectively recognized area of maximal contrast enhancement and evaluating all five of the acquired DCE series. The kinetics curves were classified as I (progressive wash-in), II (plateau) or III (rapid wash-out), in accordance to the BIRADS guidelines.

For qualitative analysis, DWI sequences were subsequently evaluated and the lesion was simply considered to be visible (characterized by diffusion restriction) or non-visible (without any diffusion restriction). For quantitative analysis, the ADC value of the index lesion was calculated by superimposing the subtracted images on the ADC map. The ROI was circular, measuring 3–6 mm, and was manually drawn on the slice where the lesion reached its greatest

diameter. Then, the ADC value was generated and stored automatically. ADC measurements were performed only on the enhanced solid portion to avoid areas of T2 shine-through, i.e., the necrotic core of the tumor. All ADC values were retrospectively measured by a radiologist with more than 10-years of experience with breast MRI, as they were not originally included in the reports.

3.2.1.2 Histologic Characteristics

All breast lesions were characterized on the histological specimen obtained by core biopsy and on the histological definitive sample after surgery, by two pathologists. Histological diagnosis was performed according to WHO classification. In the cases with lobular histotype where assessing cellularity was particularly troublesome, the count was performed on sections stained with immunohistochemistry for cytokeratin (CKAE1AE3).

The histopathological grade was evaluated according to NGS considering the tubule formation, the pleomorphism and the mitotic count through a scoring system. The total score ranges from 3 to 9: 3–5 corresponds to grade 1 (G1), 6 or 7 to grade 2 (G2) and 8 or 9 to grade 3 (G3).

Immunohistochemical (IHC) analysis was performed to evaluate molecular receptors status (ER, PgR, and HER2) and to calculate Ki-67 index. Evaluation of ER and PgR status was performed by IHC using Dako monoclonal antibody, 1:100 dilution. The monoclonal antibody Mib-1 (1:200 dilution; Dako, Glostrup, Denmark) was used to assess the Ki-67 index, which was reported as the percentage of immune-reactive cells out of 2000 tumor cells in randomly selected high-power fields surrounding the tumor core. HER2 status was re-evaluated using the Hercep test (Dako, Glostrup, Denmark), following published guidelines [**16**]. Samples that gave an equivocal IHC result were subjected to fluorescence in situ hybridization (FISH) analysis. A ratio of HER2 gene signals to chromosome 17 signals greater than 2.2 was used as a cut-off value to define HER2 gene amplification.

ER and PgR status were considered to be positive if the expression was ≥1% and negative if the expression was <1%. HER2 expression was classified as 0, 1+, 2+ or 3+; only tumors reaching a score of 3+ were considered to be HER2-positive. The lesions were divided into four groups based on the rate of Ki-67 index: <10%, between 10% and 14%, between 14% and 20%, and >20%, respectively. Patients were further grouped according to Ki-67 expression into Ki-67 index <20% and >20%, considering the first group as unequivocally negative and the second as unequivocally positive. Cancer cellularity was assessed semi-quantitatively on the biopsy specimen and the surgical specimen by estimating the percentage of the tumor area covered by neoplastic cells. When more than one tumor bed was identified, cellularity was calculated as an average of each area's cellularity, weighted for its approximate size. Foci of necrosis and in situ carcinoma were excluded from the assessment. Subdivision was made into groups as follows: <50%, between 50% and 70%, >70%, respectively.

3.2.2 Methodology

3.2.2.1 2.4. Statistical Analysis

The ADC was treated as a continuous dependent variable, whereas Ki-67 expression, tumor grade and the cellularity rate, were considered to be independent variables. All the analyzed variables did not follow a normal distribution, and non-parametric tests were used for statistical computations.

Spearman's Rho correlation test was used to correlate ADC values with histological features (histotype, class, ER, PgR, HER2, grading, Ki-67, biopsy cellularity and surgical cellularity) and with other MRI features (tumor size, kinetic curves, and margins).

To detect significant differences in ADC values among the grading groups, Ki-67 groups and biopsy and cellularity groups, both the Wilcoxon–Mann–Whitney U test (two groups comparison: G1 vs. G2, G2 vs. G3 and G1 vs. G3; Ki-67 < 20% vs. Ki-67 > 20%; cellularity < 50 vs. cellularity 50–70%, cellularity 50–70% vs. cellularity > 70%, cellularity < 50 vs.

cellularity > 70%) and Kruskal–Wallis H test (multiple-groups comparisons: G1 vs. G2 vs. G3; cellularity < 50 vs. cellularity 50–70% vs. cellularity > 70%) were carried out.

A ROC curve, plotted with TPR against the FPR, was performed to demonstrate the accuracy of ADC values in predicting the most aggressive patterns: G3 class alone, G2 + G3 classes together, the Ki-67 > 20% group and the surgical cellularity >70% group. An Area Under the Curve (AUC) > 9 indicated an excellent test, between 8 and 9 a good test, between 7 and 8 a fair test, between 6 and 7 a poor test and <6 a worthless test.

Statistical significance was set at $p < 0.05$. All data analyses were processed using SPSS (IBM Statistical Software Program, IBM, Armonk, NY, USA), version 25.0.

3.2.3 Results

A total of 100 histologically proven invasive breast carcinomas were included in the study, and their MRI examinations were retrospectively reviewed. The mean age was 54.72 years (range 38–83 years); 54 patients were postmenopausal, 46 premenopausal; 10 had undergone hormone therapy during their lifetime; 24 patients had a relative with a history of breast cancer, and 12 had two or more affected relatives. Descriptive statistics for tumor MRI characteristics, which are reported for each classification group along with relative frequencies, and Spearman's Rho test results are summarized in Table 4.

Variation			Grade				Surgical Cellularity Rate				Ki-67 Index			p Value
			1	2	3	Total	<50	50–70%	>70%	Total	<20%	>20%	Total	
Kinetic Curve	I	*n*	4	18	2	24	8	9	9	26	17	9	26	0.300
		%	4.0%	18.0%	2.0%	24.0%	8.0%	9.0%	9.0%	26.0%	17.0%	9.0%	26.0%	
	II	*n*	5	24	10	39	11	25	4	40	23	17	40	
		%	5.0%	24.0%	10.0%	39.0%	11.0%	25.0%	4.0%	40.0%	23.0%	17.0%	40.0%	
	III	*n*	3	19	14	36	9	13	12	34	20	14	34	
		%	3.0%	19.0%	14.0%	36.0%	9.0%	13.0%	12.0%	34.0%	20.0%	14.0%	34.0%	
Margins	Regular	*n*	0	7	2	9	2	2	4	8	5	3	8	0.032 *
		%	0.0%	7.0%	2.0%	9.0%	2.0%	2.0%	4.0%	8.0%	5.0%	3.0%	8.0%	
	Irregular	*n*	7	28	14	49	12	28	12	52	30	22	52	
		%	7.0%	28.0%	14.0%	49.0%	12.0%	28.0%	12.0%	52.0%	30.0%	22.0%	52.0%	
	Lobulated	*n*	4	12	3	19	7	7	4	18	12	6	18	
		%	4.0%	12.0%	3.0%	19.0%	7.0%	7.0%	4.0%	18.0%	12.0%	6.0%	18.0%	
	Spiculated	*n*	1	7	6	14	4	8	2	14	8	6	14	
		%	1.0%	7.0%	6.0%	14.0%	4.0%	8.0%	2.0%	14.0%	8.0%	6.0%	14.0%	
	Non-mass	*n*	0	8	1	9	3	2	3	8	5	3	8	
		%	0.0%	8.0%	1.0%	9.0%	3.0%	2.0%	3.0%	8.0%	5.0%	3.0%	8.0%	
Size	Mean 19.46 mm		19.67	19.00	20.07		17.32	19.87	21.08		20.38	18.08		0.560

Table 4 Description of the extracted MRI characteristics.
*** indicates statistical significance ($p < 0.05$).**

The histopathological and immunohistochemical results for all lesions and classification groups, with the results of Spearman's Rho test, are shown in Table 5.

Variation			Grade				Surgical Cellularity Rate				Ki-67 Index			p Value
			1	2	3	Total	<50	50–70%	>70%	Total	<20%	>20%	Total	
Histology	IDC	*n*	12	34	24	70	17	35	18	70	40	30	70	0.182
		%	12.0%	34.0%	24.0%	70.0%	17.0%	35.0%	18.0%	70.0%	40.0%	30.0%	70.0%	
	ILC	*n*	0	28	2	30	11	12	7	30	20	10	30	
		%	0.0%	28.0%	2.0%	30.0%	11.0%	12.0%	7.0%	30.0%	20.0%	10.0%	30.0%	
ER Status	Negative	*n*	1	4	6	11	1	2	8	11	4	7	11	0.02 *
		%	1.0%	4.0%	6.0%	11.0%	1.0%	2.0%	8.0%	11.0%	4.0%	7.0%	11.0%	
	Positive	*n*	11	58	20	89	27	45	17	89	56	33	89	
		%	11.0%	58.0%	20.0%	89.0%	27.0%	45.0%	17.0%	89.0%	56.0%	33.0%	89.0%	
PR Status	Negative	*n*	2	15	13	30	5	11	14	30	15	15	30	0.413
		%	2.0%	15.0%	13.0%	30.0%	5.0%	11.0%	14.0%	30.0%	15.0%	15.0%	30.0%	
	Positive	*n*	10	47	13	70	23	36	11	70	45	25	70	
		%	10.0%	47.0%	13.0%	70.0%	23.0%	36.0%	11.0%	70.0%	45.0%	25.0%	70.0%	
HER2 Status	Negative	*n*	12	58	23	93	26	45	22	93	58	35	93	0.373
		%	12.0%	58.0%	23.0%	93.0%	26.0%	45.0%	22.0%	93.0%	58.0%	35.0%	93.0%	
	Positive	*n*	0	4	3	7	2	2	3	7	2	5	7	
		%	0.0%	4.0%	3.0%	7.0%	2.0%	2.0%	3.0%	7.0%	2.0%	5.0%	7.0%	
Ki-67	Mean 20.06%		4.83	16.74	33.81		10.89	18.87	32.56		11.32	33.18		0.01 *

Table 5 Description of the extracted histologic characteristics.

* indicates statistical significance ($p < 0.05$).

The mean ADC value of the lesions detected at DWI was 1.09×10^{-3} mm^2/s (range: $0.7–1.5 \times 10^{-3}$ mm^2/s). No statistical significance was found by correlating ADC values with kinetic curves and tumor size; a significant correlation was found between ADC values and tumor margins ($p = 0.032$). No statistical significance was found between ADC values and histological type, PgR and HER2 receptor status ($p > 0.05$). However, ADC correlates significantly with grading, ER receptor status and Ki-67 index ($p < 0.05$). A high correlation was found between ADC values and cellularity rate, both biopsy ($p < 0.01$) and surgical ($p << 0.001$) (Figure 5, 6 and 7).

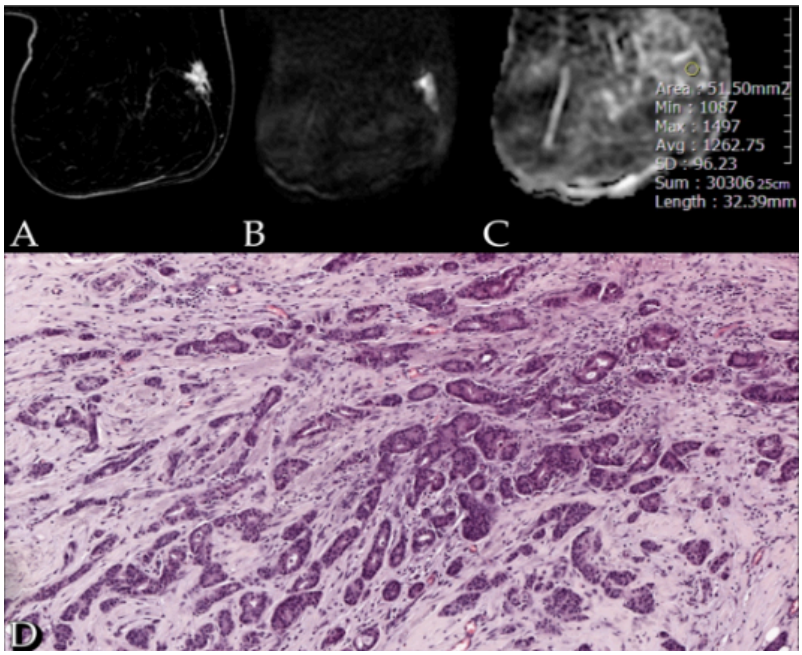

Figure 4: Case 1

Case of a 51-year-old woman with a G1 Luminal A invasive ductal carcinoma, Ki67 of 12%. (A) The post-contrast image shows an irregular spiculated enhancing lesion in the outer quadrants of the right breast. (B) The diffusion weighted image (b-value of 1000 s/mm2) shows high restriction signal in the tumor region. (C) ADC map with an ADC value of 1.26 mm2/s. (D) Section from surgical specimen shows 40% cellularity (10× HE).

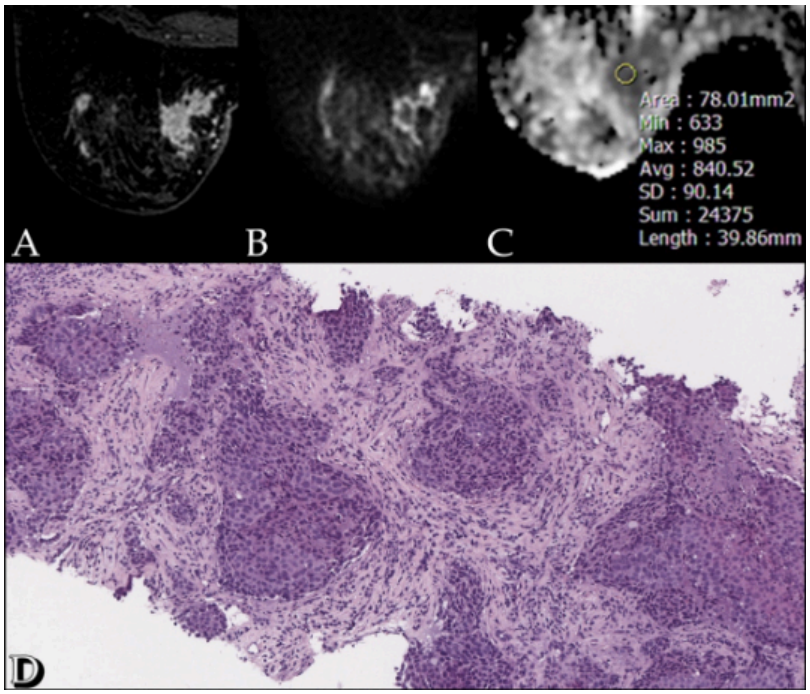

Figure 5: Case 2

Case of a 51-year-old woman with a G2 Luminal A invasive ductal carcinoma, Ki67 of 60%. (A) The post-contrast image shows an irregular enhancing lesion in the inner inferior quadrant of the left breast. (B) The diffusion weighted image (b-value of 1000 s/mm2) shows high restriction signal in the tumor region. (C) ADC map with an ADC value of 0.84 mm2/s. (D) Section from surgical specimen shows 60% cellularity (10× HE).

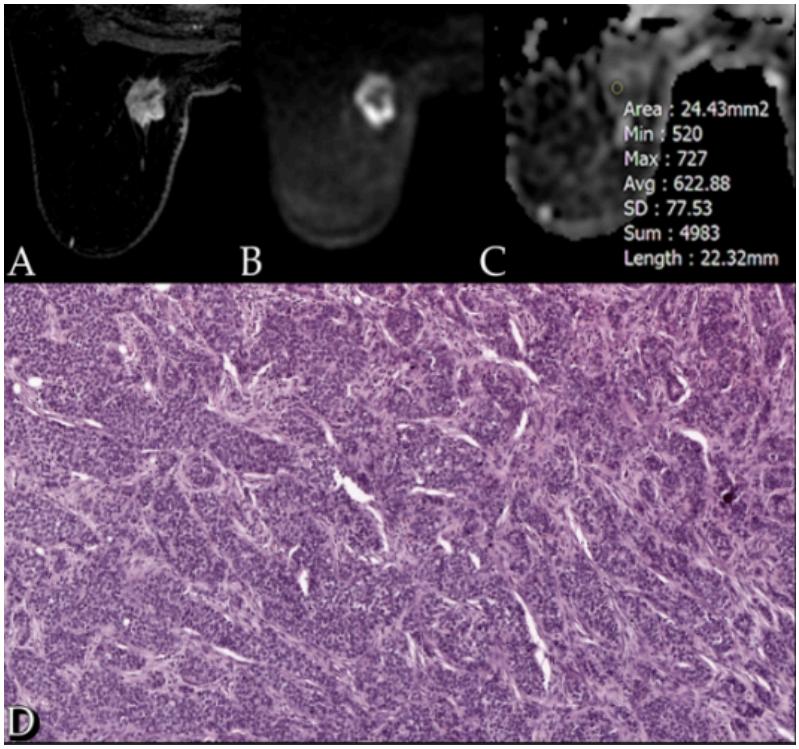

Figure 6: Case 3
Case of a 72-year-old woman with a G3 Luminal B invasive lobular carcinoma. (A) The post-contrast image shows an irregular heterogeneously enhancing lesion in the inner upper quadrant of the left breast. (B) The diffusion weighted image (b-value of 1000 s/mm2) shows high restriction signal in the tumor region. (C) ADC map with an ADC value of 0.62 mm2/s. (D) Section from surgical specimen shows 90% cellularity (10× HE).

The Wilcoxon-Mann–Whitney test showed significant different ADC values when the following grading groups were compared individually, with a p value of 0.009 for G1 vs. G2 and a p value of 0.001 for G1 vs. G3, but no significant difference was found for G2 vs. G3. ADC values were significantly higher for Ki-67 < 20 % than Ki-67 > 20%, with a *p* value < 0.03. There was a significant difference in ADC values when the biopsy cellularity groups were compared individually, with a p value of 0.014 for <50% vs. >70%. No significant difference was found for the other classes (<50% vs. 50–70% and 50–70% vs. >70%). ADC values were also statistically different for the comparison of the individual surgical cellularity groups, with a *p* value << 0.001 for <50% vs. >50–70%, whereas *p* was 0.07 for 50–70% vs. >70% and *p* << 0.001 for <50% vs. >70%.

The Kruskal-Wallis test showed that ADC values were significantly different among the three grading groups ($p = 0.009$), the three biopsy cellularity groups ($p = 0.21$) and, most strongly, the three surgical cellularity groups ($p << 0.001$).

Using ADC values, the prediction for G3 corresponded to an AUC of 0.720, whereas it corresponded to a score of 0.835 if G2 and G3 were grouped (Figure 8 and 9 respectively). The AUC for Ki-67 < 20% corresponded to 0.679 (Figure 10). The AUC for surgical cellularity > 70% was 0.805 (Figure 11).

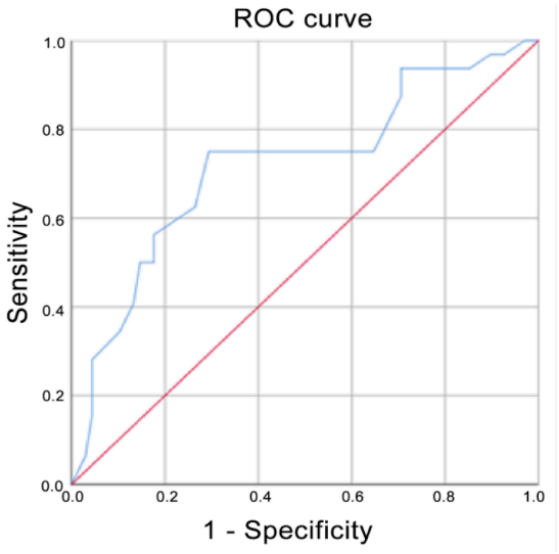

Figure 7: Receiver operating characteristic (ROC) curve when G3 was compared with G1 and G2. Area under the curve (AUC) = 0.720.

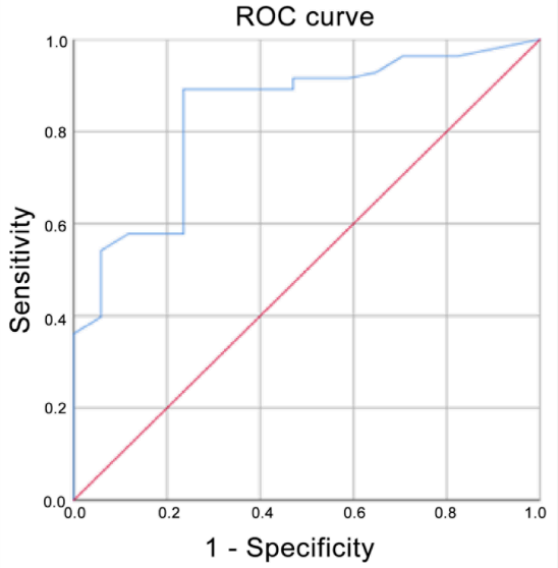

Figure 8: Receiver operating characteristic (ROC) curve when G2 and G3 were compared with G1.
Area under the curve (AUC) = 0.835

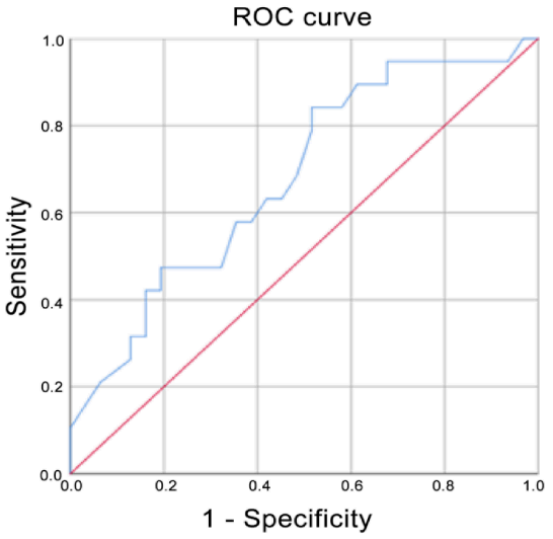

Figure 9: Receiver operating characteristic (ROC) curve when Ki-67 Index > 20% was compared with Ki-67 Index < 20%.
Area under the curve (AUC) = 0.679

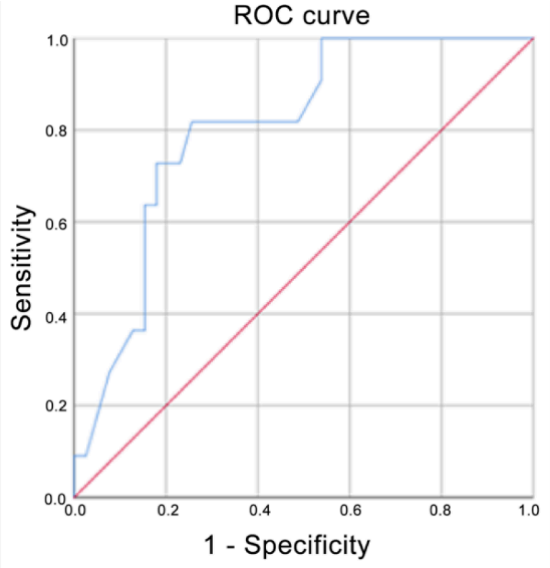

Figure 10: Receiver operating characteristic (ROC) curve when surgical cellularity rate > 70% was compared with surgical cellularity rate > 50 and surgical cellularity rate 50–70%.
Area under the curve (AUC) = 0.805.

3.2.4 Discussion and Conclusions

ADC values and their correlation with malignant lesions have been extensively studied in several papers, which have shown that ADC values of infiltrating lesions are statistically lower than those of in situ tumors, and that similarly ADC values of in situ tumors are lower than those of B3 lesions and these in turn of benign lesions[89]. Since ADC values are an expression of tumor cellularity, perfusion and angiogenesis, in the context of malignant tumor lesions they can be used to assess the degree of malignancy and aggressiveness, being inversely related to tumor grade and Ki-67 index, as already demonstrated by literature[61,90–92].

Therefore, while DWI cannot be strictly considered to be an expression of the presence or absence of receptors on cell membranes, it can be representative of the microscopic assessment of morphological and cytological features of tumor cells: those include the degree of tubule formation, nuclear pleomorphism, and mitotic count, all of which are included in the grading assessment.

The histological tumor grade is one of the most effective and practical prognostic factors in breast cancer[93]. The morphological characteristics of the tumor mass correlate with its biological and clinical behavior and thus can accurately predict the response to therapy and the patient prognosis. The prognostic significance of the Nottingham Grade System was first demonstrated in 1991[2], providing important tumor information with reproducible results. NGS has been shown to be a simple, easily reproducible, and relatively cost-effective method for predicting tumor evolution and prognosis. For this reason, it has not been completely superseded by newer molecular tests, which have cost limitations, are not as widespread as grade analysis[94], and may lead to misclassification in the presence of rare histologic subtypes (e.g., neuroendocrine tumors)[95].

Our aim is to test in our sample whether ADC, obtained from 3T MRI examination, can be used as a predictable non-invasive index of tumor aggressiveness and whether it can summarize the macroscopic biological behavior of the tumor, considering not only grading and Ki-67, but also exploring the role of ADC values in cellularity prediction.

The results present in the literature concerning the correlation between ADC and grading are highly variable between studies. In particular, some papers did not find a significant association between ADC values and histological grade of breast cancer[96,97] and there is currently no consensus on the b values to determine ADC values, and no cut-off has yet been proved to predict pre-biopsy tumor grade.

In our study, a statistically significant inverse correlation between ADC value and tumor grading was demonstrated. G1 tumors presented significantly higher ADC values compared with G2 tumors, and also G2 compared with G3 tumors. Our results are in agreement with Yuan et al., who found significantly lower ADC values in the high pathological grade class compared to the low pathological grade class ($p = 0.001$), with a median value of 0.864×10^{-3} mm^2/s for G2 + G3 lesions vs. 0.946×10^{-3} mm^2/s for G1 lesions[61].

The relationship between ADC values and histological grade was also evaluated by ROC analysis. The ROC curve showed that the evaluation of ADC values was a fair test when considering the accuracy in predicting tumors in the G3 group (AUC = 0.720); interestingly,

an AUC = 0.835 was obtained if G2 and G3 were combined. When compared to the study by Kızıldağ Yırgın et al.[92], our AUC was smaller considering the G3 class alone (0.720 vs. 0.875), but very similar when G2 and G3 were combined together (0.835 vs. 0.840). These results, expressed by the ROC curves, reflect the difficulty of distinguishing G2 tumors from the other two classes of lesions and in particular from G3 tumors, as is the case in anatomo-pathological evaluations. ADC is nothing more than an ex-vivo expression of the histological behavior of the lesion: to obtain a better predictive value of ADC, a better histo-pathological stratification between G2 and G3 classes is needed. While it is easier to identify the G1 class, the G2 definition is still controversial.

Ki-67 is a protein that is not expressed in the G0 phase of the cell cycle, and is therefore closely linked to tumor proliferation and consequently to tumor aggressiveness and cellularity. As cellularity increases, there is a reduction in the free diffusion of water molecules, which in turn corresponds to a reduction in ADC values. Ki-67 has also already been shown to be a good indicator of response to neoadjuvant chemotherapy and is a useful predictor of pCR[98]. In the field of MRI, changes in ADC values can be a non-invasive alternative to biopsy to estimate the effect of Ki-67-positive BC chemotherapy[99].

Our results regarding the correlation between ADC values and Ki-67 are in line with what has been basically demonstrated in the literature, where an inverse correlation between ADC and Ki-67 index has already been demonstrated[20,100–102]. Ki-67 positive tumors have significantly lower ADC values than Ki-67 negative tumors. In our study, we obtained an AUC of 0.679 in differentiating between Ki-67 positive BCs, which is similar to the AUC estimated by Shen et al. (0.683) even if they considered Ki-67 status as positive when Ki-67 index was greater than 14%[101]. Our AUC was significantly lower than Mori et al., who obtained an AUC of 0.81: this phenomenon is probably explained by the lower heterogeneity of their study sample, in which the included lesions were composed solely of luminal-type invasive breast cancers NOS[102].

Cellularity is presently assessed in the breast cancer specimen after surgery, reserving the assessment of cellularity biopsy only for patients destined for neo-adjuvant therapy. Some studies have shown that cellularity rates are already evident on MRI exams, and ADC values

may help in distinguishing between benign and malignant lesions and predict in the early stages the patients who may respond to the eventual neoadjuvant treatment. Cytotoxic effects are, in fact, responsible for a reduction in tumor cellularity, resulting in reduced signal in DWI-weighted sequences and increased ADC values[103,104]. However, the biopsy cellularity cannot be representative of the real cellularity of the whole tumor, coming from a tumor portion sample. Ahn, S et al. and Reisenbichler, E et al. found no correlation between cellularity and survival, while other studies demonstrated a correlation between stroma-rich tumors and an increased risk of relapse and decreased survival, especially in triple-negative BC[81,105–107]. In the present work, we aimed to demonstrate an inverse correlation between ADC and tumor cellularity, and in particular the ability to more accurately predict the cellularity of the definitive sample compared with the biopsy sample, providing an additional decision support tool in the pre-surgical phase.

To our knowledge, there are no articles aiming to identify a non-invasive prognostic factor that can support the assessment of the cellularity of the biopsy specimen and, also, in specific cases, replace this assessment, giving indirect information on the histological aggressiveness of the tumor. We successfully demonstrated an inverse correlation between the pathophysiological phenomenon of cancer cellularity and the change in cellularity as shown in DWI sequences. The relative ROC curve showed that the evaluation of ADC values was a fair test when considering the accuracy in predicting tumors in the surgical cellularity >70% group (AUC = 0.805). These findings are in agreement with the results of Hatakenaka et al.[100] and Matsubayashi et al.[20], which, in contrast to us, studied cellularity in the biopsy specimen and in the neo-adjuvant setting.

The major contribution of our work lies in demonstrating that ADC values predict main macroscopic cell growth and replication factors in a highly accurate manner. In particular, our results reinforce the current literature regarding grading and Ki-67, and introduce a new chapter regarding cellularity. There are no works that compare cellularity in the biopsy and in the surgical phases, therefore this work needs further validation. The ability of the ADC to predict surgical cellularity before surgery, during the MRI staging phase, certainly appears promising,

offering an additional quantitative tool for the evaluation of the entire tumor, assisting the information obtained from the biopsy, which represents only tumor portions.

The main limitation of this preliminary study is that we did not perform a comparison of ADC values with biopsy/surgical cellularity values before and after treatment to define whether there is a correlation even during treatment. Additionally, another limitation is the presence of both IDCs and ILCs in the study population: the significantly lower number of ILCs may have contributed to bias in the statistics.

Our results indicate a significant correlation between ADC values detected on 3T-MRI in invasive breast cancers, biopsy and surgical cellularity, tumor histologic grade, and Ki-67 index. A higher ADC value corresponds to a lower cellularity rate, which in subsequent follow ups may be associated with a better response to treatment. Tumors with a lower histologic grade, corresponding to fewer mitoses, are characterized by a larger extracellular volume, corresponding to higher ADC values; vice versa for tumors with a high histologic grade. ADC values may predict histologic grade, but further studies are needed to evaluate the value of ADC as a prognostic factor to predict tumor behavior. In the future, ADC values could be taken together with DCE curves to achieve a better understanding of invasive breast cancers in an attempt to predict their evolution, prognosis, and response to therapy.

4 LYMPH NODES: Artificial Intelligence in LNS prediction, from Radiomics to Deep Learning Approaches

Until the mid-1990s, the axillary cavity involvement was established through radical Axillary Dissection (AD), with the removal of all the lymph nodes of I, II order and, in many centers, of the III, with the aim of plan eventual systemic adjuvant therapies. Sentinel lymph node biopsy (SLNB) has demonstrated to be a valid alternative to AD. Its high diagnostic accuracy has allowed it to be elected as the Gold Standard in the vast majority of early breast cancers with clinically negative axillary cable. The most important advantage is the significant reduction in morbidity compared to AD.

The AD was actually reserved only for those patients who pre-operatively presented a clinically and/or ultrasonographically positive axillary cable or in whom the SLN is metastatic on histological examination[108].

However, about half of patients with macrometastatic SLN (metastases> 2mm) and many more with micrometastases (<2mm), undergoing AD, do not have other positive lymph nodes besides SLN. Various studies such as the ZOO11, the IBCSG study and the AMAROS, did not demonstrate differences in disease free survival and overall survival, at more than 6 years of follow-up, in patients randomized which received AD or just SLN removal, with clinically negative axillary cavity undergoing conservative surgery, and with 1 or 2 macro- or micrometastatic SLNs. Controversies still concern the setting of patients with locally advanced breast cancer or T2>3cm who, today more and more, are subjected to neoadjuvant therapies (especially chemotherapy).

In the light of these data, it is easy to understand how the topic "the state of the axilla" is of great interest and topicality[5].

Nowadays, it is well known that histopathological data of primary tumor, such as lymphovascular invasion, Ki-67 proliferation index, histological grade, estrogen receptor (ER) status and progesterone receptor (PgR) status, are predictors of SLN metastasis. However, they

are available postoperatively and cannot be used to guide decisions on performing SLN biopsy[4,38].

In this scenario, MRI has played an evolving role in providing anatomical and functional properties of breast tissues, in a non-invasive way. Findings on MRI such as tumor size, morphology, shape and enhancement have been shown as significant in differentiating breast cancer subtypes[20–22,61]. In particular, several studies have highlighted the value of dynamic contrast-enhanced MRI (DCE-MRI), identifying it as the main sequence for the detection and characterization of breast lesions, facilitating treatment choice[109]. However, manual annotation of tumor characteristics is generally limited to a few qualitative descriptors and is dependent on the operator.

Over the past few years, pilot studies have attempted to correlate the tumor features extracted from MRI with the molecular subtypes of BC. This filed is relatively new and is referred to as radiomics, i.e., the conversion of the information contained in medical images into high-dimensional, mineable and quantitative imaging characteristics, usually referred to as features or descriptors, via high-throughput extraction of data-characterization algorithms[7,8].

A growing interest is expressed in radiomics research in many fields. However, almost all previous works focused on breast cancer, correlated histological and radiomics features of primary tumors and only few works correlate the main breast tumor characteristics with lymph-node status[37,109–112].

4.1 Literature review about the role of AI applied in BC MRI for LNS prediction

The impact of analysis of radiomics-derived data on BC lesion diagnosis, prediction of NAC response, risk of recurrence and disease-free survival have already been investigated[76,113–116]. However, there are only few studies about the prediction ability of radiomics regarding lymph node status in breast cancer and no specific reviews. Only Dong-Man et al. explored the state of the art of the application of radiomics in breast MRI including the prediction of axillary lymph node metastasis in a less lymph node-specific review[7].

The purpose of this review is to investigate the potential role of radiomics as a decision support tool in predicting lymph node status in breast cancer patients.

4.1.1 Methodology

Two reviewers conducted the search, selected the studies, and extracted data from each study independently. From a total of 30 papers, 10 research articles on predicting lymph node metastasis in breast cancer using a radiomic approach in breast MRI were considered eligible. MEDLINE databases, such as PubMed and Web of Science, were searched using the following keywords: “breast cancer” AND “lymph node” AND (“radiomics” OR “radiomics”). No limitations were applied to the search strategy. Case studies, abstracts, reviews, letters to editors, editorials, and comments were excluded. Diagnostic modalities other than MRI were excluded. Only publications in English were included. We completed the search by manually reviewing the bibliography of all selected articles. The quality of the methodology was evaluated according to the RQS (Radiomics Quality Score) as described by Lambin et al. (2017). This score is useful to evaluate the quality of articles concerning radiomics through the evaluation of 16 criteria; each criterion was assigned a different maximum score in relation to its importance. The two reviewers assigned, in accordance, the RQS to the selected studies in absolute values and percentage (maximum value of 36, representing 100%). The following data were extracted from each study: title, authors, publication year and journal, study design

(retrospective or prospective), number of patients, MRI technical information (magnetic field strength and sequences), software used to perform segmentation and feature selection, number and type of radiomics features considered, algorithms used for the classification and resulting accuracy.

4.1.2 Results

Our search found 30 publications on predicting axillary lymph node status in BC patients. All these studies were published from 2017 to 2020. Ten studies were retrospective in design; only Liu et al. was prospective[117]. All the ten study characteristics, as recorded by the reviewers, are shown in Table 6. Chai et al. compared the different discriminative abilities of both pre-contrast and post-contrast MR imaging sequences, showing that the combination of CE2 features and kinetic features had the highest performance and that preoperative radiomic signatures of primary BC were associated with ALNM[118].

Liu et al. demonstrated that ALNM could be predicted using DCE-MRI-based radiomics by applying three different classifiers (SVM, logistic regression, and XGboost), finding that SVM results based on the strongest enhanced DCE-MRI images gave the best classification efficacy (accuracy of 0.85 and AUC of 0.83)[119]. The SVM showed the best accuracy also in the paper by Cui et al. (89.54) in comparison to KNN and LDA[118]. In their work, the combination of morphological and texture features had the highest performance. In addition, a nomogram that scored morphological and texture features to calculate the probability of ALNM was established. Dong et al. were the first to predict ALNM status in BC patients using radiomics based on T2-FS and DWI sequences, demonstrating that the predictive performance of features derived from T2-FS and DWI combined (AUC: 0.805) outperformed those of T2-FS and DWI when taken independently (0.770 and 0.787, respectively)[111]. A nomogram incorporating the radiomic signature, MRI-reported LN status and LN palpation was developed by Han et al.[112]. The nomogram showed better results than the radiomic signature alone and a radiomic signature to distinguish the number of metastatic LNs was also investigated (less than two positive nodes/more than two positive nodes). Liu et al. evaluated the first attempt to combine

DCE-MRI radiomic features with clinic-pathological features to improve the predictive performance of ALNM, citing that the predictive performance was comparable to Dong et al[110]. They reported that DWI is not available in all hospitals and T2-FS sequences alone are not as high performing as DCE-MRI. Tan et al. predicted ALNM in BC based on T2-FS images alone and established a nomogram incorporating radiomic signatures with clinic-pathological features using a linear regression model (AUC: 0.805), also plotting calibration curves to assess the consistency between the probability of ALNM predicted by the nomogram and the actual results[120]. Shan et al. validated a nomogram model to detect ALNM in patients with invasive BC, which incorporated the kinetic curve model and only five radiomic features extracted from DCE-MRI with a high AUC of 0.86[121]. Patients with early-stage BC were investigated by Yu et al. which developed and validated a clinical-radiomic nomogram that successfully stratified these patients according to their risk of ALNM[122]. In addition, they defined a nomogram to provide individualized prediction of ALNM and risk of disease recurrence in patients with early-stage BC. There is only one work that combines radiomic and hemodynamic features to improve preoperative prediction model for ALNM status by including quantitative parameters (such as Ktrans and Kep) to allow estimation of angiogenesis and tumor proliferation, by Liu et al. (2020). The mean RQS score calculated was 11.1 (maximum possible value=36). The RQSs and the criteria by which the studies scored lowest are shown in Table 7. In five studies, multiple segmentation was performed [110,111,119,122]. In all studies, feature selection was performed to avoid the curse of dimensionality. Six studies calculated a multivariable analysis with non-radiomic features and discussed biological and radiological correlations[110,112,120–122]. Han et al. and Yu et al. evaluated the clinical importance of applying radiomics by determining the decision curve analysis[112,122]. Only Tan et al. plotted the calibration curves to evaluate the consistency between the nomogram-predicted probability of ALNM and the actual surgical outcomes[120]. In Shan et al. and Liu et al., validation was based on datasets from different institutions, two and four, respectively. The studies had the lowest RQ scores in the following criteria: study type (only Liu et al. was prospective in design)[110,121], validation, comparison with a gold standard, potential clinical utility, economic analysis, and open science data (none of the studies made their datasets open source, although Liu et al., Dong et al. and Shan et al. allow access to the datasets upon explicit request). Some bias were described by the authors in

the discussion paragraphs. In most cases, the small samples size and the monocentricity were considered the major limitations. The population investigated was composed by less than 200 patients for seven studies[37,110,111,118,119,121]. Yu et al. used the largest sample size (over 1000 patients) to develop and validate DCE-MRI radiomic signatures for preoperative identification of ALNM. Eight studies lacked external validation for their models[121,122].

Authors	Publication year	Journal	Study design	Number of patients	Magnete	Sequence	Segmentation	Software used	Technique used for feature selection	Classification	Features	Results	Conclusion
Chai et al.	2019	J. Magn. Reson. Imaging	Retrospective	120	3 T	T1, T2, DWI, CE2, kinetics	Manual	Mazda	LASSO	LDA + 10CV	Morphological, histogram, texture (475) + kinetics features (109)	T1: Acc: 78.67; AUC: 0.87 CE2: Acc: 77.32; AUC: 0.85 T2: Acc: 74.16; AUC: 0.79 DWI: Acc: 79.15; AUC: 0.85 CE2 + k: Acc: 86.37; AUC: 0.91 union: Acc: 84.43; AUC: 0.92	Assessed the effect of ALNM status using radiomics features, and compared the discriminating abilities of different MR sequences
Liu et al.	2019	Frontiers in Oncology	Retrospective	62 total: training set (49 patients), validation set (13 patients)	3 T	CE	Manual	Radcloud	Select K best, 10CV + LASSO	SVM, LR, XGboost	first order, shape- and size- based, texture, higher order statistics features (1409)	SVM train: Acc: 76, AUC: 0.82; valid: Acc: 85, AUC: 0.83 XGb train: Acc: 84, AUC: 0.92; valid: Acc: 85, AUC: 0.83 LR train: Acc: 71, AUC: 0.82; valid: Acc: 77, AUC: 0.88	Demonstrated that combining artificial intelligence and radiomics from DCE-MRI of primary tumors to predict ALNM status
Cui et al.	2018	Scientific Reports	Retrospective	115	3 T	CE2	Regional growth (interactive and semiautomatic)	MATLAB 2017a	LASSO	SVM, KNN, LDA	Morphological, texture (58)	SVM: Acc: 89.54, AUC: 0.8615 KNN: Acc: 86.82, AUC: 0.8436 LDA: Acc: 74.27, AUC: 0.6367	Evaluated the diagnostic ability of traditional machine learning method combining with the radiomics features for preoperative prediction of ALNM
Dong et al.	2017	European Radiology	Retrospective	146	1.5 T	T2, DWI	Manual	ITK-SNAP, MATLAB 2014a	0.623 + bootstrap AUC	logistic regression models	Non-texture, texture (10,962)	T2-FS: AUC train: 0.847, AUC valid: 0.770 DWI: AUC train: 0.847, AUC valid: 0.787 T2FS + DWI: AUC train: 0.863, AUC valid: 0.805	Assessed the effect of ALN metastasis status using radiomics features, and compared the discriminating abilities of different MR sequences

Han et al.	2018	European Radiology	Retrospective	411 total: training set (279 patients), validation set (132 patients)	1.5 T	CE	Manual	ITK-SNAP, Pyadiomics	LASSO + LOOCV	SVM + 6CV	Shape, first-order statistical, texture (808)	Radiomc signature: Train, AUC: 0.78, Valid, AUC: 0.76 Nomogram: Train, AUC: 0.84, Valid, AUC: 0.87	Developed a nomogram based on a radiomic signature and clinical characteristics to predict LN metastasis
Liu et al.	2018	J. Magn. Reson. Imaging	Retrospective	163 total: training set (109 patients), validation set (54 patients)	1.5 T	CE	Manual	MRIcron, MATLAB 2017b, LIFEx v. 3.4233	LASSO + 3CV	logistic regression models	Shape, histogram, texture, Laws (590)	Radiomic features plus clinicopathologic characteristics: train, AUC: 0.91; valid, AUC: 0.87 Radiomics feat alone: train, AUC: 0.84; valid, AUC: 0.81	Combined DCE-MRI radiomic features with clinicopathologic characteristics to improve prediction performance of ALNM
Tan et al.	2020	Academic radiology	Retrospective	329 total: training set (269 patients), validation set (60 patients)	3 T	T2	Manual	ITK-SNAP, MATLAB 2019b	Mann–Whitney U-test, SVM with recursive feature elimination with linear kernel function and 10CV	SVM	Shape, first-order statistical, texture, wavelet (647)	Radiomics feat alone: Train: Acc: 76.58; AUC: 0.880, Valid: Acc: 75; AUC: 0.819 Radiomic features plus clinicopathologic characteristics: Train: Acc: 81.78; AUC: 0.894; Valid: Acc: 78.33; AUC: 0.810	Combined T2FS radiomic features with clinicopathologic characteristics to improve prediction performance of ALNM
Shan et al	2020	Frontiers in Oncology	Retrospective	145 total: training set (101 patients), validation set (44 patients)	3 T	CE, kinetics	Manual	ITK-SNAP, AK software	ANOVA and MW, Spearman correlation test, LASSO + 10CV	Logistics regression models	morphology, histogram, texture (396)	Nomogram: train AUC: 0.91, Acc: 89; valid AUC: 0.86, Acc: 82	Established and validated a nomogram model to detect ALNM which incorporated the kinetic curve pattern and radiomic features

Yu et al.	2020	JAMA Network Open	Retro-spective	1214 total: training set (849 patients), validation set (365 patients)	1.5 T/3 T	T1, T2, DWI, CE	Semiau-tomatic	3D Slicer, PyRadi-omics	LASSO and random forest algorithm	SVM, logistics regression models	Shape, first-order, gray-level cooccurrence matrix, gray-level size zone matrix, gray-level dependence matrix, and neighboring gray tone difference matrix (863)	LASSO–logistic regression model: train AUC: 0.88; valid AUC: 85 clinical-radiomic nomogram train AUC: 0.92; valid AUC: 0.90	Developed and validated nomo-grams that incorpo-rated radi-omic and clinical signatures to predict ALNM and risk of disease recur-rence
Liu et al.	2020	Cancer Imaging	Prospec-tive	164 total: training set (124 patients), validation set (40 patients)	3 T	CE, kinet-ics	Manual	AK soft-ware	LASSO + 10CV	logistics regression models	Shape, first-order statisti-cal, texture (396)	Radiomics model: train AUC: 0.81; valid AUC: 0.74, Acc: 69 Pharma-cokinetic parameters model: train AUC: 0.77; valid AUC: 0.74, Acc: 67 Combined model: train AUC: 0.80; valid AUC: 0.76, Acc. 76	Showed that the combina-tion of radiom-ics and hemo-dynamic character-istics can obtain an improved preop-erative prediction model for ALNM status

Table 6: Characteristics of the studies on lymph node and tumor MRI radiomics included in the review For each one of the ten papers, authors, journal and year of publication, number of patients and type of study, MRI technical information, radiomic properties, resulting accuracy and of study conclusions are reported

Authors	RQS	Prospec-tive study	External validation	Detection and discussion of biological correlates	Comparison to 'gold standard'	Potential clinical utility	Cost-effec-tive analysis	Open science
Chai et al.	8 (22%)	No	No	No	No	No	No	No
Liu et al.	10 (28%)	No	No	Yes	No	No	No	No
Cui et al.	7 (19%)	No	No	No	No	No	No	No
Dong et al.	9 (25%)	No	No	No	No	No	No	No
Han et al.	11 (30%)	No	No	Yes	No	Yes	No	No
Liu et al.	11 (30%)	No	No	No	No	No	No	No
Tan et al.	12 (33%)	No	No	Yes	Yes	No	No	No
Shan et al.	10 (28%)	No	Yes	No	No	No	No	No
Yu et al.	16 (44%)	No	Yes	Yes	No	Yes	No	No
Liu et al.	17 (47%)	Yes	No	Yes	No	No	No	No

Table 7: Overview of the Radiomic Quality Score (RQS)

4.1.3 Discussion and Conclusions

Radiomics is an emerging field of relatively recent development and application. The consequence, as shown in this review, is the low number of studies in the current literature, with only ten articles available regarding ALNM status prediction in breast MRI, all published after 2017 and mostly between 2018 and 2020. In addition, nine out of ten studies are retrospective in design and three include fewer than 120 patients. Dong-Man et al. (Ye et al. 2020) have already written a rapid review about the relationship between radiomics and breast cancer, including the prediction of axillary lymph node status, however, in our study, we preferred to more deeper investigate this topic, updating the results with recent studies and assessing the quality of the literature. A lymph node is considered clinically suspicious if at least one of the following criteria is present: palpability on physical examination of the axilla, suspicious features on imaging, and/or positivity on cyto-histological examinations after fine needle biopsy, core-needle biopsy or SLNB/ ALND. Clinical examination of the axillary cavity is associated with a high false negative rate, reaching values of 45% (Sacre 1986). Regarding imaging methodologies, the main one for the study of the axilla after the diagnosis of breast cancer, during NAC or follow-up, is the US examination, which allows the evaluation of the I, II and III lymph node level. Considering only the size criteria, US sensitivity and specificity range from 49 to 87% and 55 to 97%, respectively (Alvarez et al. 2006). MRI and PET-CT are the most accurate methods for studying supraclavicular and internal mammary lymph node stations, with MRI performing better than PET-CT (Liang et al. 2017). MRI is also the primary imaging modality to evaluate BC response and axillary involvement after NAC (Scheel et al. 2018). The gold standard for axillary staging is, however, represented by surgical approach (SLNB and ALND): if SLNB is negative, no further surgical evaluation is required (Giuliano et al. 2017). Although complications in both methods are decreasing over the years, there is currently no complication-free method to stage the axilla. For this reason, several non-invasive methods are being investigated. Radiomics, as quantitative information extraction and analysis, is a very promising new approach for the prediction of axillary metastases, but still limited due to its technical and methodological complexity and, therefore, not yet applied to routine clinical practice. This review focuses on the current status of the radiomic approach in predicting lymph node status based on MRI for breast cancer. Among all the available imaging techniques, MRI is non-invasive, generally well-accepted and routinely used in clinical practice for BC staging.

For axillary staging, MRI is unaffected by patient habitus and less operator-dependent than US, demonstrating the best visualization of the entire axilla (Baltzer et al. 2011). Typical MRI suspicious features include lymph node size, morphologic features (cortical thickening, loss of fat hilum, irregular margins, and round shape), presence of peripheral edema and asymmetry with the contralateral axilla (Baltzer et al. 2011). Through the combination of MRI pure information routinely used in breast cancer staging and radiomic features, unnecessary ALND and SLNB could be avoided, with a reduction of surgical complication rate and improvement of patient's quality of life. In addition, radiomic signatures applied to MRI could be a step toward more personalized and precision medicine in treatment strategies for breast cancer patients. Some studies have also developed nomograms, that are graphical representations of the mathematical relationship between prognostic factors, both clinical and diagnostic, including radiomic features, to visually aid the clinical decision process (Balachandran et al. 2015). Among breast MRI sequences, DCE is considered the best for primary tumor identification and characterization, although the diagnosis of metastatic lymph nodes by MRI is still not ideal. However, the current literature indicates promising results in terms of predicting metastatic axillary lymph nodes using MRI radiomic signatures in breast cancer patients, particularly using DCE sequences. Currently, DWI sequences are rarely used alone in radiomics studies because of relatively low image quality and spatial resolution, leading to less quantitative data availability. Among all the studies analyzed by this review, one includes DWI sequences with T2-FS as pre-contrast sequences (Dong et al. 2018), one is based only on T2-FS images (Tan et al. 2020) and two studies combined DWI and T2-FS images with DCE sequences (Chai et al. 2019; Yu et al. 2020).

Considering the different DCE phases evaluated, Chai et al. show that the second phase of post-contrast imaging (CE2) performed the best. It is acquired 60–90 s after contrast administration, achieving the most relevant contrast between tumor and background (Chai et al. 2019). This is also supported by other studies using CE sequences (Liu et al. 2019a, b, 2020; Cui et al. 2019; Han et al. 2019; Shan et al. 2020; Yu et al. 2020). Feature extraction of radiomics usually requires prior segmentation. Although studies in which ROIs are manually drawn by radiologists may be more prone to error and user variability, prediction is still good, despite being a time-consuming method. It would be ideal to develop a reliable and validated automatic

segmentation method, but it is not yet available. A critical issue concerns the quality of study results, which is currently relatively low. According to Lambin et al. (2017), the quality of the results is assessed using the 16 RQS criteria, with a maximum total score of 36. The average RQS score, obtained from the analysis of the papers considered by this review, is 11.1 (30.1%), indicating modest overall quality. The most critical points are the type of study, the detection and discussion of biological correlations, the validation, the comparison with a gold standard, the potential clinical utility, the economic analysis and the open science data (details are given in Table 7). The study design, year of publication and sample analyzed by the various studies are depending on the recent lymph node radiomics application to BC field. Even more, the researchers are characterized by an extreme variability in the software employed for feature selection and extraction as well as the number and type of classifiers applied. The absence of standardization of the applied methodologies reduces or limits the results reproducibility. Interestingly, in this regard, all articles were published in scientific medical imaging journals. Even more, in the analyzed papers, there are some important biases concerning the number of features selected, in some cases too high, as in the work of Dong et al. (>10,000 features) (Cui et al. 2019), risking overfitting and redundancy, and the pre-processing phases, such as manual segmentation, with results reproducibility reduction. This review has some important weaknesses. The first is the low number of articles analyzing lymph node status in breast cancer, with all studies but one being retrospective in design. The population numbers are quite low for most studies, although this is in accordance with their experimental nature. However, we expect that, as the application of radiomics progresses and becomes more widespread, there will be an increase in the number of cases examined, also considering the possibility of dataset sharing. Second, the results are not homogeneous, which leads to low reproducibility of the studies. There is still no standardized approach, for example in terms of the software used (some of which are produced “in house”) or the type and number of features analyzed. This weakness is reflected in the modest quality of the studies, as described by the RQS, especially in the areas of discussion of potential clinical utility and correlation between radiomic information and biological data. In conclusion, new prospective studies with larger sample sizes and focused on the clinical aspect are needed to thoroughly investigate the behavior of BCs in relation to axillary involvement and their radiomic mirroring. It remains clear that

radiomics results on axillary lymph node status predictability obtained so far are definitely encouraging. Dedicated libraries, open-access datasets and interactive comparison between different algorithms are required to test the accuracy of the results and allow for their generalization. The results of the review indicate the direction of future research: new works should be conducted using DCE-MRI, particularly CE2, introducing automation systems for the pre-processing (segmentation) phase and improving feature selection.

4.2 A new proposed radiomics approach to predict LNS using DCE sequences

Basing on the growing interest in radiomics and its application in axillary lymph node prediction in patients with BC, with this work, we aim to:

- Demonstrate that a radiomics-based approach can provide a non-invasive approach for predicting LN metastases useful for clinical practice, which helps identifying those patients who have certain negative lymph node invasion and can avoid unnecessary invasive procedure;
- Show that the combination of texture features extracted from anatomical and functional DCE-MRI images combined with clinical and histological descriptors boost the performance;
- Explore a large set of radiomics features, including also a 3D extension of Local Binary Patterns to enrich the description.

4.2.1 Population

All breast MRI exams performed at the Central Radiology Department of Policlinico Umberto I, from January 2017 until January 2019 for pre-operative evaluation, were retrospectively reviewed. The inclusion and exclusion parameters of the study were defined. Only the patients having the following characteristics were included: MRI-examination executed with 3 Tesla magnetic field, post-contrast sequences, mass-like tumors, diagnostic confirmation of invasive breast cancer by histopathological analysis, complete histological analysis including molecular

receptor structure and proliferation index Ki67, final lymph-node status (ipsilateral axillary cable).

Patients who had breast implants or expanders, patients in post-chemotherapy follow-up, patients in neo-adjuvant treatment and patients whose images were not of excellent diagnostic quality were excluded.

For all patients, a written informed consensus was obtained before contrast-MRI execution.

Following the mentioned criteria, a total of 97 breast cancer patients (age range: 37–82 yo; mean age: 55.48 yo) with 99 breast lesions were included in the study (2 bilateral breast cancer cases, 83 invasive ductal carcinoma (83.8%), 12 invasive lobular carcinoma (12.1%) and 4 medullary carcinoma (4.0%).

4.2.1.1 MRI examination

All MRI investigations were performed with a 3 Tesla magnetic field with Discovery 750 machinery, from GE Healthcare (Milwaukee, WI, USA), using a breast dedicated 8-channel coil with patient in prone position. Written informed consensus was obtained before each procedure by the patient.

After a sequence performed for framing and carried out along the three space orthogonal planes, were acquired a T2-weighted single shot fast spin eco (IDEAL), a diffusion-weighted sequence (DWI) with b values of 0 and 1000 sec/mm^2 and T1-weighted 3D axial sequences, dynamic gradient echo, with fat suppression (VIBRANT), before and five times after administration of the intravenous contrast medium (Gadobenate dimeglumine), administered at a concentration of 0.2 mmoL/kg and at a speed of 2 mL/s for a total of 14 mmoL in an ideal patient of 70 kg.

Therefore, "subtracted" images were obtained automatically, using the post-contrast images to which the mask was removed. All the technical details of the sequences are summarized in Table 8.

Sequences	Values
Axial T2 IDEAL FSE	
TR/TE (ms)	3500–5200/120–135
Matrix	352 × 224
FoV (mm)	370 × 370
NEX	1
Slice thickness (mm)	3.5
Axial EPI DW	
TR/TE (ms)	2700/58
Matrix	100 × 120
FoV (mm)	360 × 360
NEX	6
B-values (s/mm^2)	0–1000
Slice thickness (mm)	5
Axial T1 VIBRANT	
TR/TE (ms)	6.6/4.3
Angle	10°
FoV (mm)	380 × 380
Matrix	512 × 256
NEX	1
Slice thickness (mm)	2.4

Table 8: Study protocol adopted for MRI examination.
(TE: Time of Echo. TR: Time of Repetition. FOV: Field of View. FSE: Fast Spin Echo. EPI DW: Echo Planar Imaging Diffusion-Weighted. VIBRANT: Volume Imaging for BReast AssessmeNT.)

For each lesion, the following features were collected: localization (breast quadrant position, retro-areolar, upper- or lower-external, upper- or, lower-internal quadrant); margins (divided in regular, irregular, lobulated and spiculated); maximum diameter (mm), measured on post-contrast images; contrast enhancement after contrast agent administration, performed using

signal intensity/time curve (type I, with both slow wash-in and wash-out, type II, depicted as a plateau curve and type III, with both rapid wash-in and wash-out).

4.2.1.2 Clinical and histological data

Specific anamnestic-clinical data for each patient were previously collected and, according to them, the population was divided into groups: age, menopausal status (42 patients in the pre- and 57 in the postmenopausal phase), hormone therapy (10 patients who have performed at least 3 continuous months of hormone therapy of any kind contraceptive, replacement or therapeutic therapy and 89 patients who did not assume any hormone therapy), familiarity (62 patients without any familiar, 33 patients with one familiar and 4 patients with at least 2 female or male family members affected by breast cancer at any age).

The histological examination was performed for all the 99 breast lesion included in the study. The samples were obtained by core-biopsy or surgery and analysed by an anatomo-pathologist with more than 15 years of experience. The classification of the tumor histotype was performed in accordance with the WHO classification. The tumor histological grade was assigned following the NGS for which a score from 1 to 3 was given for each of these parameters: tubular formation, nuclear pleomorphism and number of mitoses. Immunohistochemical analysis was carried out to determine the receptor structure, Estrogen Receptor (ER), Progesteron Receptor (PgR), Human Epidermal growth factor Receptor (HER2), and the proliferation index Ki67. ER and PgR expression was considered as positive when >10%; Her2 was considered as positive when >+2 and ki67 when >14%.

Hence, the following histological data were collected for each tumor: histotype (divided in invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC), medullary carcinoma), grading (three groups) and tumor class, basing on hormone receptor expression and proliferation index (Luminal A: ER+, HER2- and low ki67; Luminal B: ER+, HER2 –/+ and high ki67; HER2 overexpressed; Triple Negative (TN): ER–, PgR–, HER2–).

The status of the axilla was assessed after the diagnosis of breast cancer, analysing radiologically (ultrasound of the axillary cable in the diagnostic phase and breast MRI in the staging phase), clinically and then histologically the status of the axillary lymph nodes during definitive surgery. Then, axillary cable definition consisted of sentinel node dissection, sampling dissection or total lymphadenectomy, basing on surgeon decision, but curative in all cases. The patients were simply classified as positive or negative, depending on whether there was, in the first case, at least one lymph node involved, or, in the second case, no positive lymph node. The status of the axillary cable defines the so-called final label, dichotomized into positive LN or negative LN.

In the following sections, we present the method that predicts axillary lymph node metastasis, therefore being a safe and non- invasive prognostic approach. A schema is offered in Figure 12, which consists of 4 blocks, numbered I–IV. According to the figure, next sections first describe the segmentation and the pre-processing approach adopted. Then, we present the proposed method to compute the quantitative descriptors, i.e., feature computation, and next we move to feature selection. Finally, we present the classification approach.

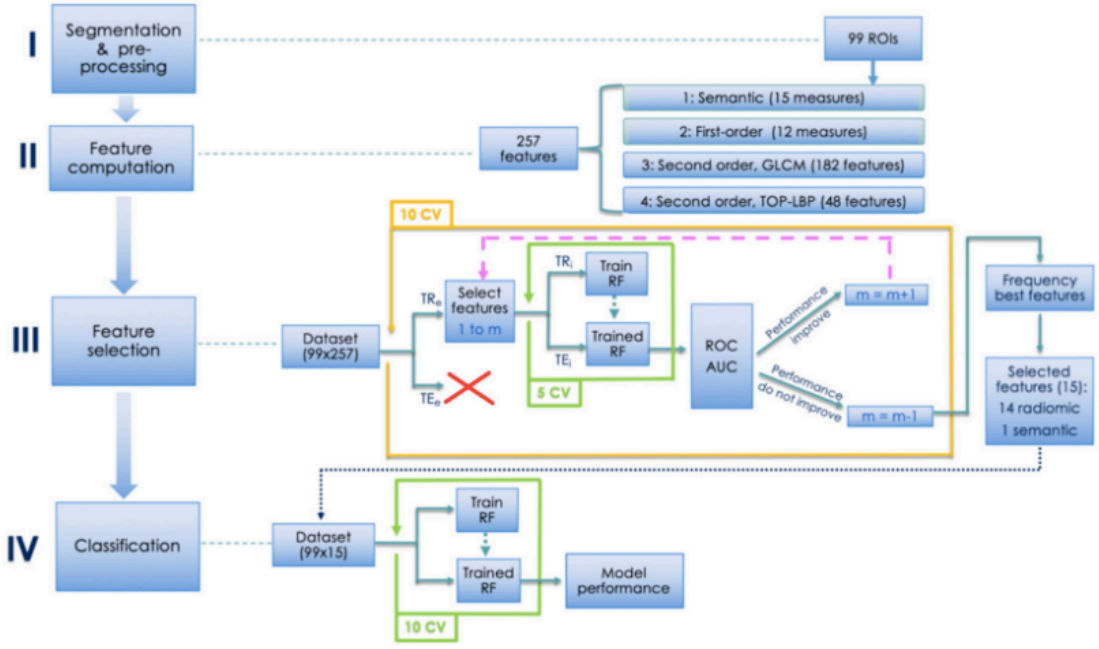

Figure 11: Schematic representation of the proposed approach.
TR and TE denote the training and the test set, whilst the subscript e and i stand for external and internal cross-validation, respectively.

4.2.2 Methodology

4.2.2.1 Segmentation and pre-processing

This initial step of the method is depicted in the first block of Figure 12. This step also includes data preparation operations, such as the anonymization. The preparation of the images was performed on a personal workstation using 3D Slicer (version 4.8), an open-source software freely available online (http://www.slicer.org)[123]. Each case was identified with a progressive identification number (ID). For each tumor, the subtracted post-contrast T1w-MRI was selected. Since five post-contrast phases were available, we used the second sequences, because image lesions in the second phase (60–120 s) had the highest contrast resolution.

At this point, for each case, a label-map was generated. Using manual and assisted segmentation techniques with the thresholding technique, the lesions were manually drawn (Figure 13).

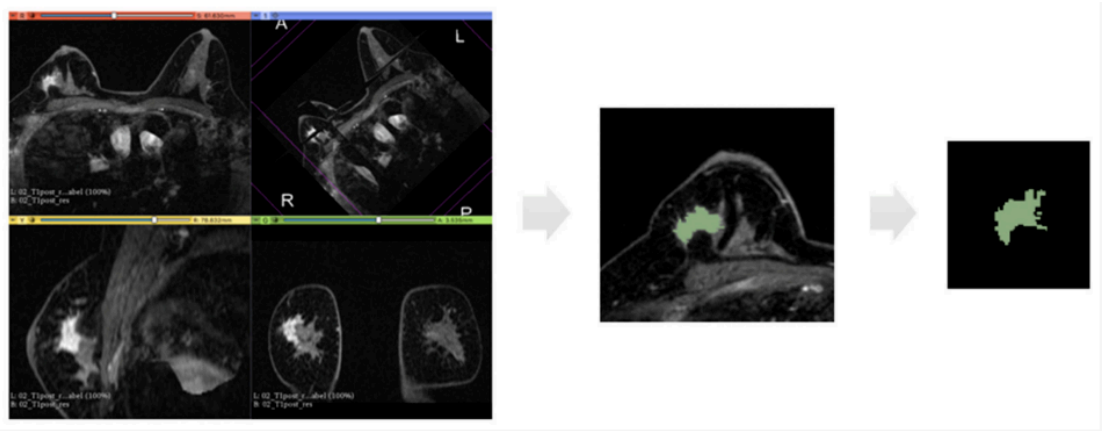

Figure 12:Representation of the extraction of a segmentation mask.
From left to right: first panel represents an example of 3D image analysed in our dataset which is the second phase of subtracted post-contrast sequence; then, for each 2D slice, we have the region of interest drawn around the tumour mass and finally the binary mask extracted from this segmentation

The segmentation was initially always performed in the axial projections and subsequently remodeled and optimized in the other projections until the lesion was contoured optimally, avoiding the necrosis when present into the lesion, as shown in Figure 14.

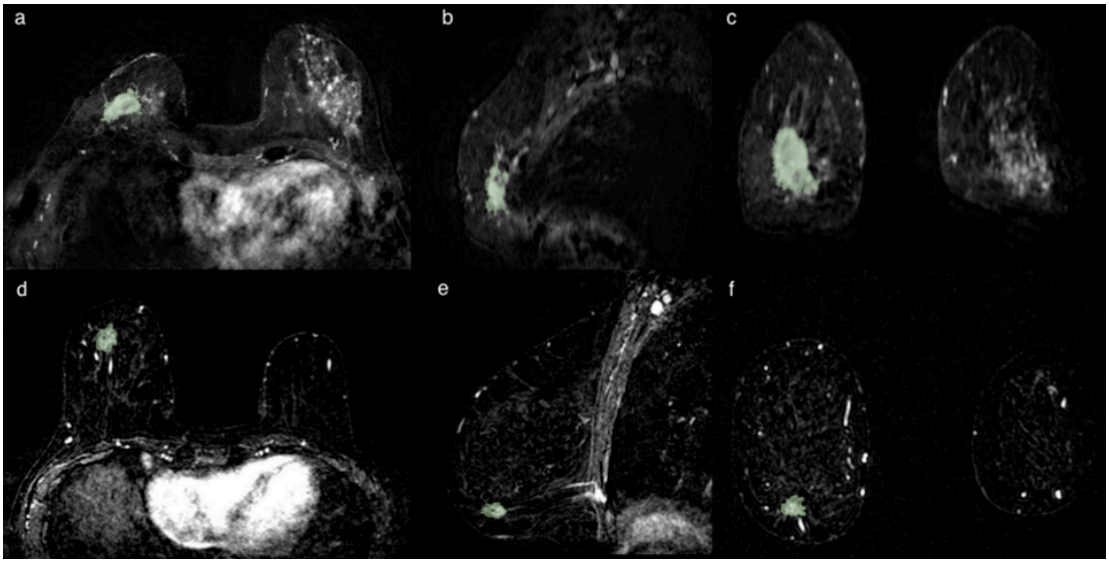

Figure 13: Two cases of breast cancer with positive and negative lymph node axillary involvement at definitive surgery, respectively.
The first (a–c) represented by a 32 mm mass, in 56yo patient with TN CDI tumor characterized by spiculated margins, heterogeneous enhancement with a necrotic core and a signal/intensity curve type 3 at MRI exam. The second (d–f) is a 13 mm nodule, in a 61 yo patient, LUMINAL A CDI, characterized by irregular margins, heterogeneous enhancement and a signal/intensity curve type 3. The segmentation has been performed in the axial image (a,d), following the margins and including the spicule characterized by contrast-enhancement. The segmentation was then optimized in the sagittal (b,e) and coronal (c,f) planes, avoiding the darker part representing the necrosis and the vessels.

Any multifocal or multicenter lesions were also segmented. In the event of a bilateral tumors, the lesions were attributed to the same patient but to different IDs, considering them one at a time.

Let us now focus on the segmented slices; we notice that the boundaries of each ROI are often coarse, affecting the quality of the features when calculated. This can be easily understood considering that many of the second-order measures presented in the following are computed from each voxel, considering also the other voxels in its neighborhood. Before the feature extraction step, we pre-process the lesions' contours to obtain a 3D volume with smoother edges, thus including large part of voxel neighborhood. This is performed in two stages. First, we iteratively compute a 3D convex hull (CH), which is the minimum-volume bounded into a

convex polygon and containing the ROI[124]; this process begins considering all not connected three-dimensional regions segmented using a default 26-connectivity. Second, all the 3D regions connected using a 26-connectivity are merged into a single volume, determining a new global CH. Such technique iterates until there are no more CHs left to merge, i.e., until the algorithm converges.

The CHs obtained so far are considered the definitive ROIs (Figure 15 depicts an example) and from these volumes we extract the features defined in the next subsection.

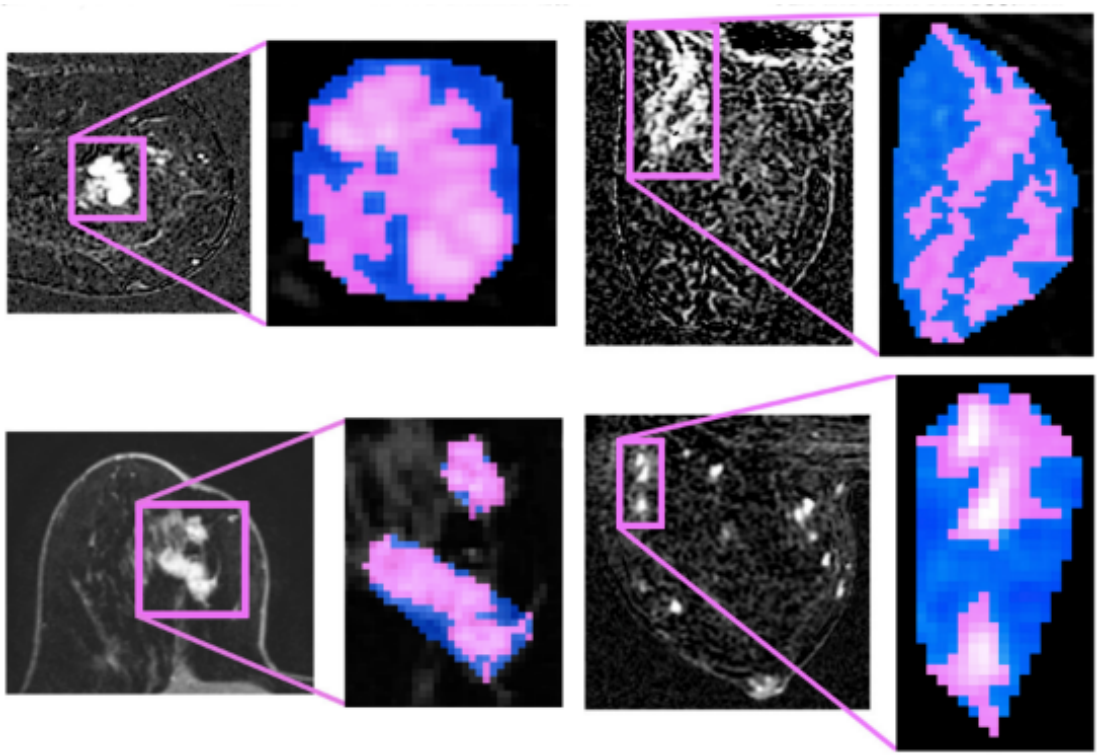

Figure 14: Four pre-processing examples.
Left side halves: Bifocal lesion localized at upper-external quadrant, characterized by spiculated margins and IS/T curve. Right side halves: in pink, we highlight the original segmentation and, in blue, the convex hulls obtained.

It worth noting the double benefit of applying of the described algorithm. Indeed, it is able to remove all the "outliers" from the ROI, i.e., all sub-volumes that are not large enough to

delineate a CH. Moreover, the final volume integrates a slightly larger region surrounding the lesion, thus including all those voxels considered after the conversion of the ROI into a three-dimensional convex shape, and such extra tissue can bring useful quantitative information on how the tumor infiltrates during growth.

4.2.2.2 Features computation

In addition to the clinical data and the histological data presented in the materials, this work leverages on several radiomic features that can be divided into first-order and texture features. According to **Figure 1**, this step is performed in the second block. Texture measures, in turn, are derived from the 3D Gray Level Co-occurrence Matrix and from the Three Orthogonal Planes-Local Binary Patterns. Next will be described each feature group.

4.2.2.3 Feature selection and model construction

First-Order Features

They are based on counting of the image voxels grouped by their grey level and, therefore, they measure the intensity density distribution of the ROIs. The literature has shown that humans are mostly sensitive to the light distribution of pixels in images and are better able to discriminate their differences when considering such characteristic rather than others: this is because the first order features are usually referred to as human inspired descriptors. In practice, we extracted the histogram from the 3D ROI and then we computed 12 characteristics up to the fourth-order moments, which are: the mean, the standard deviation, the skewness and the kurtosis. In addition, from the histogram, we also extracted the width, the entropy, the energy, the value of the histogram absolute maximum and the corresponding grey-level value, the energy around such maximum and the number of relative maxima in the histogram and their energy. More details and the definitions of all the features are available in the paper by Cordelli et al.[125].

3D Gray Level Co-Occurrence Matrix Features

Usually, healthy tissues and tumors have different textures. Therefore, a deeper focus on the relative distribution of voxels within a ROI rather than simply counting them can reveal many details that can be used as a measure of its microstructure. To this goal, for each ROI we computed the 3D grey level co-occurrence matrix (GLCM3), i.e., a well-known image transformation that captures such a distribution as it can be considered the 3D generalization of the more popular planar GLCM.

Let us denote with I a 3D greyscale image, and consider a Cartesian reference system O(x, y, z). If we position the origin of the reference system in the upper left corner of I, the position of all its voxels is given by the vector $p=pxî+pyĵ+pzk̂$, with px, py and pz ∈ N, and $î$, $ĵ$, $k̂$ are the direction vectors of the axes. We also denote a displacement vector $d=dxî+dyĵ+dzk̂$, with dx, dy and dz ∈ N. If we define m equal to the number of bit used to represent I, a GLCM3 is a square matrix of size $N = 2^m$, where each entry (gi, gj), with both gi and gj ∈ $[0, 2^m - 1]$, represents the number of times a voxel in p with intensity gi is separated by a displacement d from another voxel with intensity gj, therefore located in p + d. Denoting as dh the h-th component of d, to assess all possible directions, we considered the combination of dh ∈ {−1, 0, 1} as displacements, without considering the (0, 0, 0) vector yielding to 26 different displacement directions. More details on GLCM3 are reported by Sebastian et al.[126]. As last step, from each GLCM3 we compute 7 measures, namely the autocorrelation, homogeneity, entropy, energy, covariance, inertia and absolute contrast[127]. Concatenating such GLCM3 measures we get 26 × 7 = 182 textural descriptors for each ID.

Three Orthogonal Planes-Local Binary Patterns Features

This descriptor is a generalization of well-known planar Local Binary Pattern (LBP) and, although relatively new in radiomics, it can be applied to a three-dimensional image. Now we first present a brief summary of the original LBP's concepts and, then, we will introduce the 3D descriptor.

Denoting again a generic 2D image with *I*, if we consider each pixel *p*, it is possible to compare its intensity Ip with the intensities of all its j-th neighbours pixels laying on a circle centred in *p* and with radius r. The i-th pixel is set to 1 if Ij>Ip, 0 otherwise. The next step is to process all neighbours of *p* in a circular fashion, read the resulting sequence of 0 s and 1 s as a binary string and encode the value of *p* to the equivalent decimal value. In practice, all the pixels in *I* are processed following this procedure to obtain an image encoding all the intensity distributions of the pixels with respect to their neighbours. Using this descriptor, we thus grasp part of textural information contained in *I*. Finally, notice that, if we denote with P the number of local neighbours that surround the central point *p*, the number of patterns for this planar LBP implementation is 2^P.

In order to extend this approach to a three-dimensional environment, in [128] the authors presented a solution that considers an helicoidal neighbourhood for each voxel. However, this produces 2^{3P+2} patterns; to avoid this computational burden, we introduce another 3D implementation of LBP transformation that considers the co-occurrence on three orthogonal planes crossing the center of the analysed volume, as detailed in [128]. This method is named Three-Orthogonal-Planes LBP (TOP-LBP) and computes three 2D LBPs, one for each plane, and the obtained histograms are concatenated to obtain a unique representation for the specific volume. This conspicuously alleviates the computational load, since the number of patterns for TOP-LBP is 3×2^P. Furthermore, in our LBP implementation, we consider two more variants to cope with other two issues of 2D LBP definition. First, we computed rotation invariant LBP, i.e., all binary strings obtained as the circular shift of a fundamental string are considered the same. Second, we implemented a uniform version of LBP, i.e., all binary strings containing more than two crossings from 0 to 1 or from 1 to 0 are considered not uniform and coded with a specific string. In our case, setting P = 8, we get 48 features by computing first-order measures from each histogram of the three 2D LBP.

Feature Selection

One of the main goals of radiomics is to find a signature, i.e., the set of all those features with the highest discriminative power for the task at hand. Moreover, this is also beneficial for the

learning phase of a classification algorithm, as reducing the number of features to be considered reduces the risk of the curse of dimensionality. To this goal, we searched for the most discriminative features by using a wrapper (third block of **Figure 1**). This feature selection method is known for being able to find dependencies between the descriptors and, at the same time, it exploits the interaction between the subset of features to be found and the model selection itself [129]. In fact, after determining a search method for all possible subsets of features, it evaluates them by training and then testing a specific classification model.

We used the Random Forest (RF) as a learning paradigm evaluating the feature subsets by maximizing the area under the curve of the receiving operator (AUC-ROC). The reason for selecting the RF is its ability to work with both qualitative and quantitative features, its capacity to reduce the risk of data overfitting, and the fact that it can also handle datasets with a large number of characteristics. Furthermore, for all its parameters, we used the default values provided by the Matlab library, without any fine tuning. Indeed, we were not interested in the best absolute performance and, also, it has been empirically observed that in many cases the use of tuned parameters cannot significantly outperform the default values of a classifier suggested in the literature [130].

With regard to the subset exploration strategy, it is worth noting that we set the wrapper to work with the best first search by proceeding with nested cross-validation to evaluate the model. This approach ensures that the performance of the model during validation is not affected by a possible favourable split of the data into training and testing, and thus eliminates any bias in the final performance evaluation since the outer test set was never used in the wrapper model evaluation, as also shown in the third block in Figure 12.

In practice, in each cycle of external cross-validation all samples are divided into a training set and a test set. Then, in the feature selection we take into account only the training samples and apply an additional five-fold internal cross-validation to them. On each inner loop, we test several subgroups of *Fm* features, where each subgroup is composed of the first *m* features of the whole pool.

Finally, for constructing the radiomic signature, we consider only those features that are selected at least 20% of times in the nested cross-validation experiments described so far. The purpose of this approach is two-fold: first, in this way a balance is maintained between the discriminative power of *Fm* and the risk of overfitting, also reducing the curse of dimensionality which, although mitigated thanks to the use of RF, still remains partly present; second, the validity of this method has been experimentally determined in some preliminary tests, omitted here in order not to burden the discussion. In conclusion, this procedure returned the final set of 15 features shown in Table 9.

Histotype	Semantic
Energy around relative maxima	1st order
Energy around absolute maximum	1st order
Energy in direction (−1.1.0)	GLCM3
Energy in direction (0.−1.1)	GLCM3
Energy in direction (0.−1.−1)	GLCM3
Energy in direction (0.−1.1)	GLCM3
Inverse in direction (−1.0.0)	GLCM3
Inverse in direction (1.−1.1)	GLCM3
Inverse in direction (−1.0.−1)	GLCM3
Range	LBP-TOP
Range U	LBP-TOP
Range U RI	LBP-TOP
Number of relative maxima RI	LBP-TOP
Energy around relative maxima RI	LBP-TOP

Table 9: The final feature set

Classification and Model Construction

The fourth and last block in Figure 12 depicts the final classification stage, which provides the resulting performance when using a RF as classifier. As before, we set its parameters to the

default values. In order to cope with the class imbalance in the dataset we introduce a misclassification cost making the algorithm cost-sensitive. Let us recall that cost sensitive learning is one of the three main approaches that can be used to address class imbalance, a.k.a. class skew, which is the problem of having a disproportionate training set among different classes. This issue arises since traditional learning algorithms are designed to minimize errors over the majority samples, ignoring or paying less attention to instances of the minority classes, and this usually results in poor predictive accuracy over the minority ones. The predicted label will assume values of 0 in case of negative axillary involvement, and 1 in case of positive axillary involvement. On this basis, in the following, we consider an error cost matrix that sets to 0.59 the cost of false negative errors (FN error = 0.59) and we set to 0.41 the cost of wrongly classifying a negative lymph node into a positive one (FP error = 0.41). Straightforwardly, no penalties are set in the case of a correct prediction. The rationale lies in observing that improperly indicating the presence of a LN metastasis is less dangerous than wrongly indicating its absence, since in the latter case the patient will be no further treated; we account for this by setting misclassification cost to 50% more than the value of the other one.

We performed all the experiments in ten-fold cross-validation, using different randomly generated partitions when cycling over the inner loop with respect to the outer cross-validation loop of the feature selection. This approach as well as not using the outer test partition in the feature selection stage, avoids any bias into the final classification model.

Finally, in order to quantify the results obtained in all the outer folds and, therefore, evaluate the overall classification performances, we built the ROC and computed the underlying area (AUC).

4.2.3 Results

In this study, 97 breast cancer patients with 99 breast lesions were enrolled (2 patients had bilateral breast cancer). The average age of the entire cohort was 55.48 years. There were 27/99 axillary cables confirmed as ALN positive metastases and the rest 72/99 as negative.

The clinical and pathological characteristics of the patients are summarized in Table 10.

Class	Group	Frequency	Percentage
Familiarity	none	62	62.6
	1 fam	33	33.3
	>1 fam	4	4.0
HT	no	89	89.9
	yes	10	10.1
Menopause	no	42	42.4
	yes	57	57.6
IS curve/T	I	3	3.0
	II	53	53.5
	III	43	43.4
Margins	regular	3	3.0
	irregular	46	46.5
	lobulated	30	30.3
	spiculated	20	20.2
Histotype	IDC	83	83.8
	ILC	12	12.1
	Medullary	4	4.0
Grading	1	12	12.1
	2	45	45.5
	3	42	42.4
Class	Luminal A	39	39.4
	Luminal B	35	35.4
	Her2	9	9.1
	TN	16	16.2

Table 10: Main frequencies for the various semantic classes analyzed.

For the sake of presentation, the proposed radiomic approach is divided into specific steps as shown in Figure 12. After the feature selection, a signature of 15 features that lead to the final results was individuated. The selected features were 14 radiomic features and one semantic.

These features were used to build the classifier, whose performance is shown in the first row of Table 11. Our classifier achieved a sensibility and specificity of 86% and 74%, respectively. To deepen the results, we perform additional experiments by investigating one group of features at a time. Hence, we obtained four features' groups (1° order, GLCM3, LBP-TOP and

semantic features) in addition to the signature used to build the proposed classifier. The sensibility and specificity calculated by each feature group were 85% and 48% for 1° order features, 79% and 52% for GLCM3 features, 69% and 41% for LBP-TOP features and 97% and 11% for semantic features, respectively.

Features	TP (%)	FP (%)	TN (%)	FN (%)	Accuracy	ROC Area
Proposed	62 (89)	7 (11)	20 (67)	10 (33)	0.828 (95% CI, 73.6% to 92.0%)	0.856 (95% CI, 77.5% to 93.7%)
Clinical	71 (75)	24 (25)	3 (75)	1 (25)	0.748 (95% CI, 70.2% to 79.4%)	0.533 (95% CI, 45.7% to 60.9%)
First order	61 (81)	14 (19)	13 (54)	11 (46)	0.748 (95% CI, 66.6% to 83.0%)	0.665 (95% CI, 58.4% to 74.6%)
GLCM3	57 (83)	13 (17)	14 (48)	15 (52)	0.717 (95% CI, 64.1% to 79.3%)	0.763 (95% CI, 65.4% to 87.2%)
LBP-TOP	50 (76)	16 (24)	11 (33)	22 (67)	0.616 (95% CI, 54.2% to 69.0%)	0.618 (95% CI, 53.0% to 70.6%)

Table 11: Metrics that describe the various accuracies considering each feature class stand alone and the proposed approach, obtained by the combination of all the features classes.
(TP = True Positive. FP = False Positive. TN = True Negative. FN = False Negative. ROC = Receiver Operating Curve. CI = Confidence Interval)

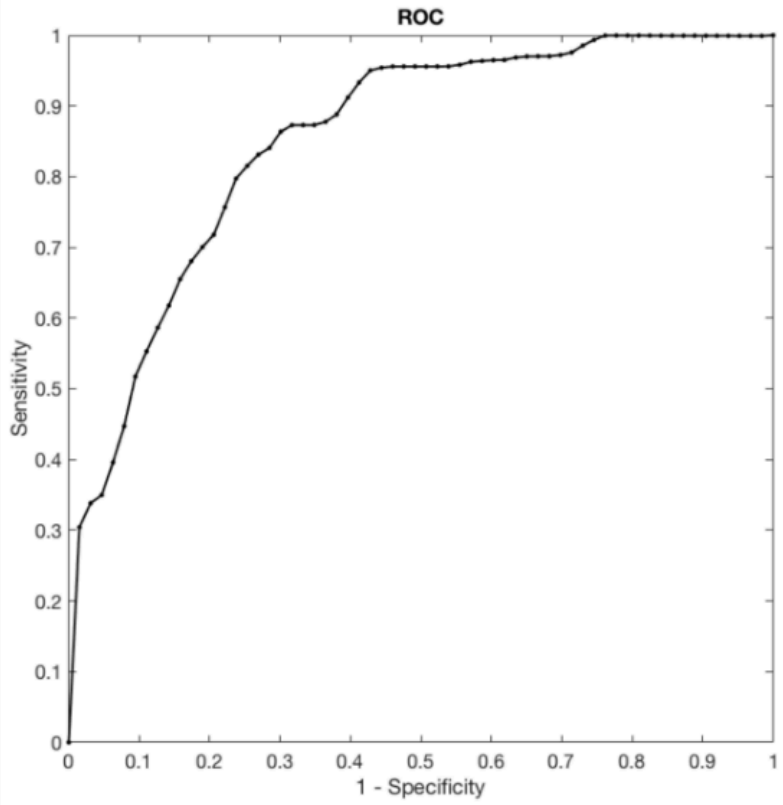

Figure 15: The chart shows the ROC curve which is representative of the classifier performance.

4.2.4 Discussion and Conclusions

In the literature, it is well known that the condition of the axillary lymph nodes in breast cancer patients is one of the most important prognostic factors in terms of loco-regional recurrence and overall survival. In this regard, the American College of Surgeons conducted an important study on 20,547 patients with local or regional disease treated with radical mastectomy, and identified a recurrence rate of 19% and 49% and 5-year survival rates without recurrence of 60% and 35%, for patients with negative and positive lymph nodes, respectively[131]. Given the importance of lymph nodes status, not only for the prognosis but also in the guide of oncological therapeutic choice, safer and non-invasive approaches have been investigated for the initial lymph node staging[38,112,132].

Today, we know that radiological images contain much more information than that which is perceivable and visible to the radiologist. This hidden information can provide several interesting data about the tissues, data that are quantifiable. As stated by Gillies RJ et al. in[8], radiomics is based on this principle because it extracts and analyses large quantities of characteristics, defined as "features", computed from medical images routinely acquired.

In this work, we present an approach for providing valid, rapid and non-invasive support in the prediction of axillary lymph node status, using radiomics features computed from post-contrast MR images associated with patient clinical information and tumor histology.

To date, there are works on breast radiomics which evaluate its effectiveness in diagnosis, identification or characterization, prognosis or response to therapy, using the imaging information produced by different techniques (US, mammography and MRI). The most recent review on radiomics and breast cancer[7] has only 17 studies. Additionally, even fewer are the studies which evaluate the use of breast radiomics in predicting axillary lymph node status. To the best of our knowledge, there are six main papers that have been published since 2017. All are monocentric retrospective studies, with a population ranged from 62 to 163 patients with only one work using 411 samples. Our study includes 97 patients with 99 breast lesions, being in the average of the current literary trend. Chai et al. analyse the features with and without contrast medium agent sequences and compare the two derived accuracy and AUC[118], while Dong et al. use only the T2-weighted and diffusion sequences with the aim of identifying an instrument that prevents the use of the intravenous contrast medium, obtaining AUC values of 0.770 and 0.787, respectively[111]. Our contribution and all the other work in the literature[37,110,112,119,131] use only contrast enhanced sequences: this choice is motivated by the results obtained from studies of both breast radiomics[7,75] and breast MRI[110,115], which identify in the contrast enhanced sequence the main sequence for detection and characterization of the breast lesion.

Regarding the choice of the contrast enhanced phase, there is currently no consensus in defining in which phase the extraction of features offers the best forecast. Liu et al. have applied, for the same purpose of this paper, radiomic features extracted from the first contrast enhanced phase

obtaining an AUC of 0.806[119]. In our study, instead, the second contrast enhanced phase was used.

The higher contrast resolution of the contrast enhanced sequences among all the sequences allows a high definition of the morphological lesion details[115]. Even more, the use of a high-field magnet, 3 Tesla here, guarantees the production of high temporal and spatial resolution image[133]. Obtaining high quality images not only allows a more adequate definition and selection of the features, but reduces the classifier noise.

In our sample of 99 breast tumors, the selection method applied identified 14 quantitative radiomic and 1 semantic features, represented by tumor histological class, correlating significantly with the lymph node status.

Among the 15 selected features, the most numerous were first and second order ones, which describe the intensity and textural characteristics of the tumor, representing the intratumoral heterogeneity and the subtle alterations of the morphology of the tissues. The selection of tumor histological class and not grading or dimension, as we can it can be expected, may be explained to the type of wrapper used for the selection. The reason of such behaviour is probably relied to the multivariate approach of the chosen feature selection algorithm. In fact, when considering multiple features pools, the discrimination power of each tested subset is usually outperforming the capability of predicting the correct class using single features. Therefore, when considering a large set of features mainly computed from the images, where only a smaller percent belongs to a semantic nature (i.e., the clinical and histological data), the final signature is reasonably expected to contain only those semantic features useful to maximize the performance when used in conjunction with the radiomics features. Another motivation is related to the physiologic tumor behavior since, as demonstrated by other studies, ductal carcinoma is characterized by morphological and enhancement patterns which are expression of more aggressive behaviour.

Then, the combination of the selected features made it possible to train a classifier capable of predicting the lymph node status in patients with breast cancer, with an AUC of 0.86. The AUC reported in Figure 15 shows the relation between the true positive (sensitivity) and the false

positive (1–specificity) rates. Our classifier identifies 20/27 TN patients with a specificity of 74%, which is considered a very good result in comparison with US overall specificity for palpable and non-palpable LN, which it is based on the LN size criteria and it ranges between 44.1% and 97.9%. It is interesting to note that the specificity achieved using only semantic feature is of 11%, indicating the relevance of radiomics as supporting tool.

An example of how the classifier correctly works is reported in the two cases shown in Figure 17.

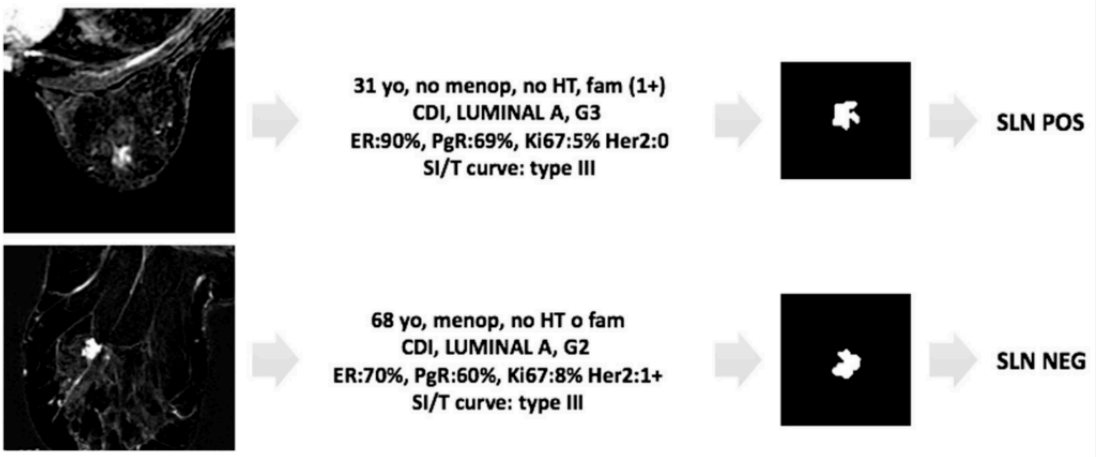

Figure 16: Example of applicability of the proposed approach.
Two cases of invasive breast cancer, characterized by very similar MRI features (morphology, margin, post-contrast intensity, IS/T curve, and size and position of the two lesions are comparable) as shown in the left panels. The first case (first row) is a ductal G3 LUMINAL A carcinoma, in 31yo patient with familiarity, the second one (second row) is also a ductal G2 LUMINAL A carcinoma, but in 68yo patient without familiarity. Despite the common MRI and histological features, the produced classifier correctly categorizes the two patients as positive and negative LN status, respectively.

The study has several aspects worth highlighting. The set of analysed images was acquired in a single center using a single MRI device, repeating the same protocol for all patients, which allows maximum reproducibility in the extraction and analysis of radiomic characteristics. Segmentation was performed for all patients by a single radiologist, who used the same methodology; also the extraction of the features was performed starting from free commercial software, validated in many previous studies, but modified according to the specific study

needs, in particular the features analysed included also a 3D extension of Local Binary Patterns in addition to the best known first-order features, GLCM and LBP.

There are also several limitations: first of all, it is a retrospective study with a relatively small sample. The provenance of the population from a single medical center limits the possible generalization and results. A larger cohort of patients, from multiple centers, is needed for a more rigorous analysis. Secondly, ROIs were drawn manually by the radiologist, so they are time-consuming and subject to user errors and variability. An automatic, reliable and validated segmentation method is ideal but not yet available. Thirdly, it was decided to use only the contrast enhanced sequences. Although these have been demonstrated to have the best accuracy, we recognize that it could be limiting and, in any case, lose some important information provided by the other sequences.

Our results suggest that histological data and radiomics features can be combined for the prediction of lymph node metastases, guiding the treatment planning. The results achieved suggest that they have the potential to impact the clinical practice by offering to clinicians and to patients the possibility of avoiding invasive procedures such as lymphadenectomy or lymph node biopsy in unnecessary cases.

4.3 A DL approach to predict LNS with different bounding boxes using DCE sequence

Among the DL-based solutions used to analyze visual imagery, the use of convolutional neural networks (CNNs) is the most applied in medical field. A CNN consists of a set of convolutional layers capable of learning a compact hierarchical representation of the input that is well suited to the specific task to be solved.

In BC, radiomics has been studied considerably for the characterization of primary cancer, while there are still relatively few works that evaluate its application to the relationship between tumor imaging characteristics and LNS[74,134,135]. In these works, in particular, traditional radiomics systems were mainly investigated, while DL applications are limited to a few studies and ultrasound cas[136–138]. Furthermore, the majority of these studies focused on analyzing the imaging features of the primary tumor exclusively, ignoring the tissue adjacent to the tumor. Nevertheless, evidence has shown that the peritumoral region contains valuable information on the potential tumor aggressiveness and, also, lymphatic spread, in particular in cases of multifocal and multicentric tumors[73,139,140]. In our previous works[74,139], we analyzed the impact of the radiomics of the primary tumor and the peritumoral edema on the LNS prediction. This time, based on the previously cited papers demonstrating the importance of the peritumoral parenchyma, we assume that, close to the tumor lesions, there might be some data not visible to the human eye that influences the metastatic lymph node spread.

Then, in this work, we investigate the role of the lesions neighboring parenchyma features through CNNs by experimentally analyzing different tumor bounding (TB) options. The novelty of the technical contribution can be summarized as follows:

- We propose and implement different tumor bounding options to systematically assess the contribution gathered by the healthy tissue in the prediction of the axillary lymph node status;
- We propose three different CNNs whose architectures vary according to the characteristics of the volumes considered as inputs;

- We evaluate the impact of the 3D and 2D features extracted from input volumes and the influence of healthy tissue in cases of multifocal and multicentric tumors.

4.3.1 Population

All breast MRI exams were performed for preoperative evaluation at the Central Radiology Department of Policlinico Umberto I between January 2017 and January 2020 and retrospectively reviewed. A written informed consensus was obtained before the execution of a contrast-MRI for all examinations.

All patients met the following inclusion criteria: three Tesla magnetic field MRI examinations, post-contrast sequences, mass-like tumors, histopathological confirmation of invasive breast cancer, a complete histological analysis, and definitive lymph-node status of the ipsilateral axilla. In cases of BC bilateral lesions, the two lesions were evaluated separately since the two breasts can be considered a single part. Patients with an incomplete MRI examination or damaged images and patients without a complete histopathological analysis were excluded.

The patients were excluded if they had breast implants or expanders, were in follow-up neo- or adjuvant chemotherapy, or the MRI images were not of excellent diagnostic quality.

A total of 153 patients (average age 55 years; range 30–85) met the inclusion criteria. In two patients who had bilateral breast cancer, the two breasts were considered as a single, independent part. Therefore, a total of 155 malignant breast cancer lesions were included in total.

The LNS status was assessed as positive if at least one lymph node involved by metastasis was present in the definitive histopathological axillary cable sample (LN+); the LNS was considered negative if all axillary lymph nodes were safe (LN–).

4.3.1.1 MRI Examination

All MRI examinations were performed using a 3T magnet (Discovery 750; GE Healthcare, Milwaukee, WI, USA). The sequence used for the analysis was the dynamic Contrast-enhanced T1-weighted 3D sequence using fat suppression with a TR and a TE of 6.6 ms and 4.3 ms, respectively, with an ip angle of 10°, a matrix of 512 × 256, NEX 1, and a slice thickness of 2.4 mm. An amount of 0.2 mmol/kg of Gadobenate-dimeglumine (Multihance®; Bracco Imaging, Milan, Italy) was used as the contrast agent, injected through a 20G intravenous cannula at a rate of 2 mL/s plus 15 mL of saline solution at the same speed. For each acquisition, the relative subtracted images were automatically generated and used for tumor analysis.

The images were analyzed by two radiologists with 10 and 3 years of experience, respectively. The tumors were described as unifocal when only one lesion was present; multifocal when more than one tumor lesion was present in the same breast quadrant/region; and multicentric when multiple tumor lesions were present in different breast quadrants/regions. For each lesion, the target dimensions, margins (regular, irregular, lobulated, or spiculated), and intensity signal timing curve (I, II, or III, based on wash-in and wash-out) were reported.

4.3.1.2 Clinical Data

The patients' clinical data were collected, and, according to these data, the population was split into subgroups: age, familiarity (considered positive if at least one familiar member was affected by breast cancer at any age), hormone therapy (considered positive if the patient performed at least 3 continuous months of hormone therapy including any kind contraceptive, replacement, or therapeutic therapy), and menopausal status.

4.3.1.3 Histological Data

The samples were obtained by a core biopsy or surgery and analyzed by an anatomic pathologist with more than 15 years of experience. The tumor histotype classification followed the WHO classification[141]. The tumor histological grade was assigned in accordance with the NGS, and a score from one to three was given for these tumor characteristics: tubular formation, nuclear pleomorphism, and the number of mitoses. Furthermore, the estrogen receptor (ER), progesterone receptor (PgR), human epidermal growth factor receptor (HER2), and the proliferation index Ki67 were assessed for immunohistochemical analysis. A cut-off of 10% was used to consider the ER and PgR expression as positive; while HER2 was considered positive when >+2, and ki67 was considered positive when >14%.

Moreover, other histological data were collected: histotype (including ductal carcinoma (IDC) and invasive lobular carcinoma (ILC)), grading (divided into G1, G2, or G3), and tumor class, which includes the hormone receptor status and the proliferation index percentage (Luminal A: ER+, HER2– and low ki67; Luminal B: ER+, HER2 –/+ and high ki67; HER2 overexpressed; Triple Negative (TN): ER–, PgR–, HER2–).

4.3.1.4 Axillary Lymph Node Status

The axillary lymph node status was considered as the final output. The LNS was assessed after an invasive breast cancer diagnosis using definitive surgical characterization (sentinel node dissection, sampling dissection, or total lymphadenectomy, based on surgeon decision but curative in all cases) [**1**,**2**]. The LNS was simply classified as positive, if there was at least a sentinel LN involved, or negative if there was no positive lymph node. On this basis, the dataset accounts for 27 positive and 128 negative patients, which are referred to as LN+ and LN– in the following. Table 12 reports the details of the dataset used in this paper.

Patients	153
BC lesion	155
LN+	128
LN−	27
Series	3D T1-weighted DCE
Mode	3T (Discovery 750; GE Healthcare, Milwaukee, WI, USA)
Dose	0.2 mmol/kg of Gadobenate-dimeglumine
Injection flow rate	2 mL/s

Table 12: Details about the involved dataset.

4.3.2 Methodology

4.3.2.1 Pre-Processing and segmentation

The patients' clinical data and the frequencies of the tumor histological and MRI characteristics were reported. The Wilcoxon test was performed to compare these data between patients with and without LN involvement, setting the statistical significance at $p < 0.05$. The statistical analysis was performed using MatLab v. 2020b.

The images were anonymized and uploaded on a dedicated open-source software (3D Slicer, version 4.8). An identification number (ID) was assigned to each patient. Bilateral tumors were considered with two different IDs. For each case, the subtracted post-contrast T1w-MRI was selected. The second phase (60–120 s) was selected for ROI segmentation, due to its higher contrast resolution. Then, a label map was created. The lesions were manually drawn through manual and assisted thresholding segmentation techniques on the axial projection (Figure 18). When present, necrosis was avoided by segmentation. For multifocal or multicentric tumors, all lesions, even the smallest, were segmented.

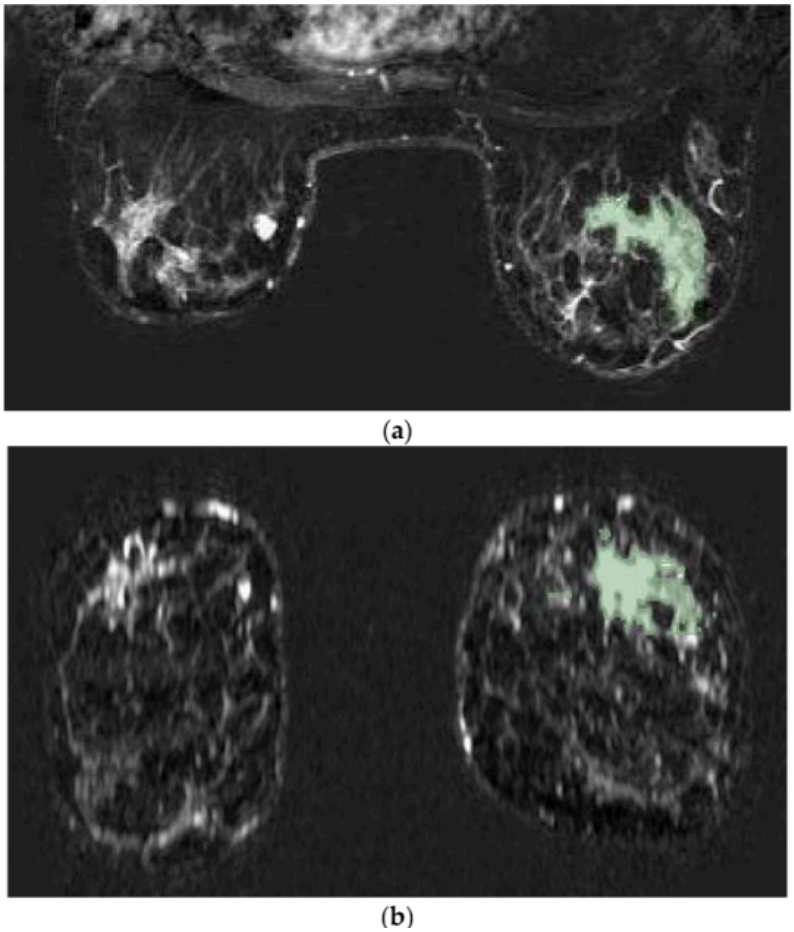

Figure 17: Tumor lesion segmentation
Tumor lesion segmentation using 3D Slicer software in axial (a), coronal (b), and sagittal (c) MRI projections during the second phase of the post-contrast sequence as demonstrated in a case involving a 56-year-old woman with right invasive ductal breast cancer with unifocal mass-like lesion characterized by spiculated margins and heterogeneous enhancement after contrast medium administration with curve SI/T type III.

4.3.2.2 Deep Learning Analysis

An assessment of axillary lymph nodes reflects inherent primary tumor features, whose examination enables the discovery of non-invasive substitutes for the sentinel node biopsy currently being utilized. Most literature proposals have focused on the DCE sequence-extracted hand-crafted characteristics of breast lesions and performed the classification with ML techniques. The absence of a well-defined, efficient collection of features in the area of BC

has, however, prompted researchers to investigate large and heterogeneous characteristics, implementing a feature selection step to pick the most discriminative ones.

The use of CNNs for the ALN metastasis prediction in this study is made possible by their capacity to automatically learn the set of features that best suits the problem at hand, thereby eliminating the need for the feature extraction stage common to ML techniques and resolving the challenge of identifying the most discriminating set of primary tumor characteristics. Furthermore, we investigate several tumor boundary alternatives that vary depending on the quantity of healthy tissue to include in order to assess how the peritumoral area influences the performance of the involved networks.

The implemented analysis consists of three main steps: the imaging data definition, used to prepare data belonging to the DCE sequences of different patients; the volume extraction and bounding options, which describe the tumor bounding options; and the architecture of the CNNs for the ALN assessment, which introduces the involved CNNs that differ according to the characteristics of the volumes considered as input.

4.3.2.3 Imaging Data Definition

The DCE acquisition consists of MRI images (3D volumes) taken before (pre-contrast) and after (post-contrast) the contrast agent (CA) injection, resulting in 4D data, with three spatial (x,y,z) dimensions and one temporal (t). Although MRI exams are acquired with the same instruments, patients may present a different number of acquired volumes, resulting in a need for the selection of a subset of them. By denoting the number of acquired volumes for a patient p with N_p, with t_i the i_{th} acquisition, where t_0 and tNp−1 are the pre-contrast and the last one, respectively, the subtractive series is obtained by considering $t_i - t_0$ with i form 1 to $N_p - 1$. To equalize the number of acquisitions for all patients, four specific subtractive volumes are selected: the first (t_1), second (t_2), and last (tNp−1) volumes and the median index (t_m) volume between the third and the second-to-last volume. In this way, information about the wash-in and wash-out of the CA due is preserved.

4.3.2.4 Volume Extraction and Bounding Options

Based on the hypothesis that peritumoral tissue may contain valuable information on tumor aggressiveness, and thus affects the ALN metastasis spread [**19**,**20**,**21**], we evaluate how the amount of the included non-tumor tissue impacts the ALN involvement, considering that patients differ in the size and number of lesions. In particular, a total of six different tumor bounding options are proposed and analyzed. They differ according to their included quantities of healthy tissue, which allows us to evaluate the contribution gathered from the area surrounding the tumor. Figure 19 shows the differences in the axial projections of the proposed bounding options.

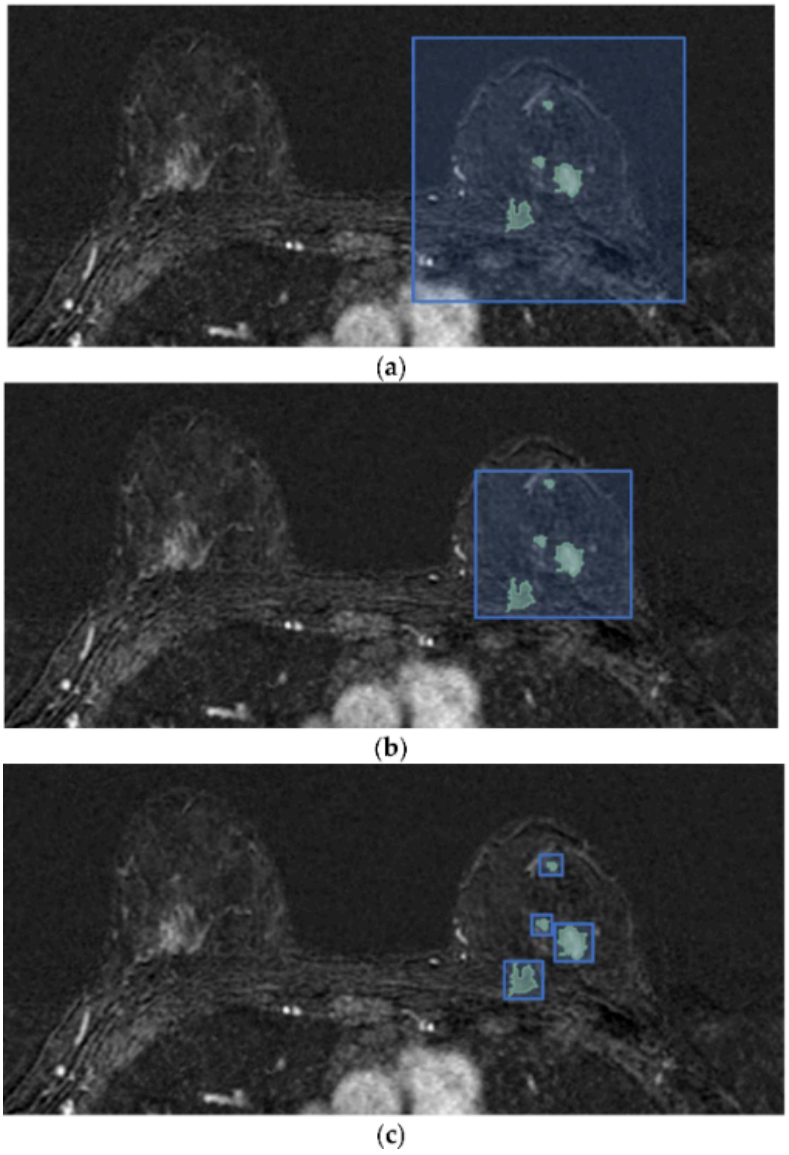

Figure 18: Differences in axial projections of the implemented tumor bounding options.
(a) In the SFB, a fixed-size 3D bounding box is used; (b) in the SVB option, the smallest 3D bounding box circumscribed to the tumor region is considered, and (c) in the SLVB, the SVB option is applied to each lesion of the patient.

In the single fixed-size box (SFB), a fixed-size 3D bounding box is centered in the tumor, completely encompassing the whole tumor region (or of all lesions, in cases of multifocal or

multicentric tumors), and is used to crop each subject. The bounding box is always the same; it is applied to all four subtractive acquisitions (t_1, t_2, t_m, tNp−1) and is patient-independent. The amount of non-tumor tissue directly depends on the tumor lesions' dimensions: in the case of a single and small lesion, the extracted volume contains a high portion of healthy tissue compared to a damaged one and vice versa.

In the single variable-size box (SVB), the smallest 3D cubical bounding box is used to crop each subject in the four subtractive selected acquisitions. In contrast to the SFB, the cubical box in the SVB option is patient-dependent and aims to limit the amount of included non-tumor tissue. As a consequence, the amount of non-tumor tissue depends on the shape of each patient's tumor region and he difference between the largest and smallest dimensions. Nevertheless, in cases of multifocal and multicentric tumors, the parenchyma between lesions is included in the extracted 4D volume.

As aforementioned, the DCE sequence of a patient is 4D data, in which each voxel is associated with information regarding its measurement in millimeters. In more detail, the attribute of pixel spacing is the physical distance between the centers of each two-dimensional pixel, specified by two numeric values representing the row and vertical spacing, while the spacing between slices measures the spacing between slices along the normal to the first image. The above attributes represent the DCE image resolution. Since the DCE sequences belonging to different subjects may have different resolutions, in the single isotropic-size box (SIB) option all the volumes are firstly re-sampled to obtain MRI images with isotropic voxels, and then the tumor bounding option SVB is applied, as described above.

In the single lesion variable-size box (SLVB), the SVB option was applied to each tumor lesion of the considered patient. A box for each tumor lesion is extracted, and, therefore, the tissue between lesions is excluded in the prediction of the ALN status with the aim of analyzing how much the parenchyma between lesions in multifocal/multicentric tumors impacts the ALN assessment.

In the single lesion isotropic-size box (SLIB), the tumor bounding box involves extracting a different box for each of the involved patient's lesions that considers a DCE sequence with 3D

volumes re-sampled to obtain isotropic voxels. The SLVB procedure is applied after having resized each acquisition.

In the two-dimensional slice (2DS) option, we propose applying the SVB procedure and then cutting the sequence of the 3D cropped volumes along the projection with the highest spatial resolution, which results in a series of two-dimensional slices with four temporal instants. For each patient, this process generates a set of 3D slices, with two spatial dimensions and one temporal, which represents the same section of tissue seen at four different time points (t_1,t_2, t_m, and tNp−1). It is worth noting that only the slices containing lesions are taken into account: this is possible since the ALN status assessment requires lesion segmentation/detection.

Table 13 summarizes the six different bounding options proposed in this paper.

Bounding Option	Details
SFB	A fixed-size 3D bounding box is used.
SVB	The smallest 3D cubical bounding box is used.
SIB	The SVB is applied after resizing the volumes to obtain MRI images with isotropic voxels.
SLVB	The SVB is applied to each lesion.
SLIB	The SLVB is applied after resizing the volumes to obtain MRI images with isotropic voxels.
2DS	The SVB is applied, and then the sequence of the 3D cropped volumes is cut along the projection with the highest spatial resolution.

Table 13: Details of the proposed bounding box options

4.3.2.5 CNN Architecture for ALN Assessment

Based on these different bounding box options, we investigate different CNNs, whose architectures were developed for turning to the characteristics of the volumes considered as inputs, for their ALN status predictions. Hence, we designed three CNNs: the SFB-NET, used for the SFB option; the VB-NET, considered when the size of the bounding box varies according to each tumor lesion, that is, for the SVB, SIB, SLVB, and SLIB; and the 2DS-NET, introduced for the 2DS bounding option. Each CNN receives four-channel volumes as input, which represents the considered acquisitions of the DCE sequence. The proposed networks consist of different reduction layers and two fully connected layers. The output of each CNN consists of a two-element vector, which, after the application of the softmax function, is interpreted as the probability that the network associates with each class, namely the positive (LN+) and negative (LN−) ones.

The SFB-NET is a 3D CNN with three reduction layers, whose architecture is shown in Figure 20a. Each reduction block consists of a convolutional layer with $5 \times 5 \times 5$ kernels (the number of kernels depends on the output channels) and a stride set to four.

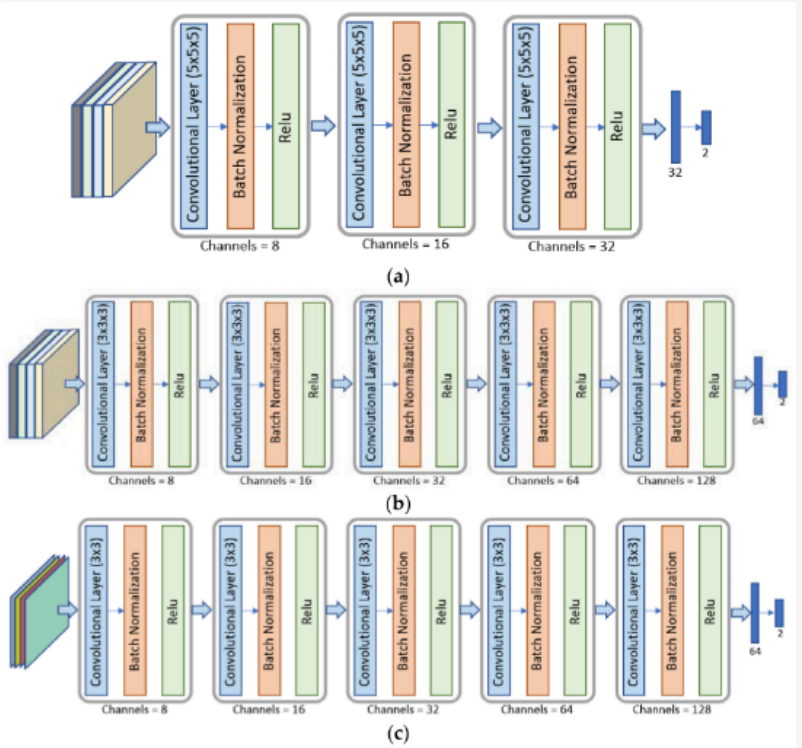

Figure 19: Architectures of the different CNNs used.
Panel (a): The SFB-NET is a 3D CNN with three reduction blocks and two fully connected layers. Panel (b): The VB-NET is a 3D CNN with five reduction blocks and two fully connected layers. Panel (c): The 2DS-NET is a 2D CNN with five reduction blocks and two fully connected layer.

The architecture of the VB-NET is shown in Figure 20b. It is a 3D CNN, consisting of five reduction blocks. The 4D input volume represents the smallest cubical box surrounding the tumor region or each lesion considered at four different time instants. Therefore, the sequence of the five convolutional layers with $4 \times 4 \times 4$ kernels aims to implement a gradual dimensionality reduction. The stride is set to two, and the padding is one, except for the last convolutional layer.

The 2DS-NET is reported in Figure 20c. It presents the same characteristics as described for the VB-NET and is implemented by replacing the 3D convolutions with the standard 2D ones. In other words, the 2DS-NET is a 2D CNN since the input is a 3D volume that represents a bidimensional slice with four temporal instants (channels). Each convolutional layer consists

of 4 × 4 kernels, and the stride is set to two. The padding is one, except for the last convolutional layer.

4.3.2.6 Experimental Setup

The proposed tumor bounding box options result in volumes with different characteristics. Indeed, in the SFB case, a 160 × 160 × 80 box centered in the lesion center is used, generating a 160 × 160 × 80 × 4 volume for each patient, which represents the input for the SFB-NET. The size of the box is chosen in such a way that, for each patient, the tumor is completely enclosed. Since the SVB, SIB, SLVB, and SLIB options consider a bounding box size that varies according to each patient's tumor region, a resize stage with the bilinear interpolation method provides a standardized size of 64 × 64 × 64 × 4, which then fed the VB-NET. The resize stage is also needed for the 2DS option. The extracted four-channel images are resized to a common size of 64 × 64 × 4, before consideration as inputs for the 2DS-NET.

As aforementioned, the SLVB, SLIB, and 2DS bounding options generate a set of volumes or 3D slices (four-channel images) for each patient. However, the aim is to provide a unique label for each subject that represents the potential risk of having axillary lymph node metastasis. As a consequence, the predicted classes of all volumes (or 3D slices) belonging to the same patient need to be combined. From all of the combining strategies, majority voting (MV) is used, in which the label for each subject is the most common predicted class over all of its volumes (or 3D slices).

We select the implemented CNN architectures by choosing the best configurations obtained by varying the size of the convolutional kernels from three to seven and the number of reduction layers from three to five. Moreover, we also take into account the architecture proposed by Nguyen et al.[19], which is the only architecture that aims to predict ALN metastasis using the primary tumor DCE-MRI features with a DL approach. In particular, we consider the solution proposed in [19] as a starting point for providing different changes.

The performance is evaluated in terms of accuracy (ACC), sensitivity (SENS), specificity (SPE), area under the ROC curve (AUC), and Cohen's kappa coefficient (K). All experiments were run by applying a patient-wise 10-fold cross-validation (CV) to better assess the generalization ability of each approach, avoiding the use of volumes belonging to the same subject both in the training and evaluation steps. We split the set of patients into 10 different folds, and, in each iteration, we use one fold for testing, one for validation, and the remaining folds for training. In particular, we ensure that each patient is included once in the validation set and once in the test set by using the i-th iteration, the i-th fold as the test, and the previous one as the validation, as reported in Figure 21.

Figure 20: Details about the implemented patient-based 10-fold CV.
In the i-th iteration, the i-th fold is selected as the test (green), and the previous one is selected as the validation (orange). The remaining folds are included in the training set.

When training the CNNs, we augmented the dataset by applying random rotation and flipping. In particular, during the training, each volume underwent a vertical and horizontal flip and a rotation of an angle of either 90, 180, or 270 degrees with a probability of 0.5. The training set was balanced by replicating some randomly chosen volumes belonging to the minority class, which relies on data augmentation operations to introduce variability among the samples. The

extracted volumes were normalized in the (0,1) range to ensure that, in the classification step, the used convolutional neural networks operate with volumes that have the same scale across different patients. During the experiments, the maximum number of epochs was set to 500; the batch size was set to 16 for the SFB-NET and VB-NET and 32 for the 2DS-NET. The learning rate for the cross-entropy loss was set to 10^{-6}. We used the Adam optimizer with a weight decay set to 10^{-4}. To find the appropriate hyper-parameters, we implemented a grid search by varying the batch size in (8, 64), the learning rate in (10–7, 10–3), and the weight decay in (0, 10–4).

All experiments were carried out using Pytorch (version 1.10, developed by Meta AI and now part of the Linux Foundation umbrella), and the pre-processing step, including the different bounding box options, was implemented in MATLAB 2020b.

4.3.3 Results

The clinical characteristics of the patients and the pathological and MRI features of each tumor are summarized in Table 14. The menopause status, grading, and class significantly differ ($p < 0.05$) between patients in cases of lymph node involvement. No other significant differences were observed between the two examined groups (LN+ vs. LN–).

Class	Group	Frequency	Percentage	LN+	LN−	*p* Value
Familiarity	none	109	70.32%	22	87	0.1314
	≥1 fam	46	29.68%	5	41	
HT	no	141	90.97%	27	114	0.0733
	yes	14	9.03%	0	14	
Menopause	no	67	43.23%	17	50	0.0233 *
	yes	88	56.77%	10	78	
IS curve/T	I	21	13.55%	3	18	0.2819
	II	69	44.52%	10	59	
	III	65	41.94%	14	51	
Margins	regular	7	4.52%	0	7	0.5504
	irregular	83	53.55%	18	65	
	lobulated	19	12.26%	3	16	
	spiculated	46	29.68%	6	40	
Histotype	IDC	129	83.23%	23	106	0.7351
	ILC	23	14.84%	4	19	
	Medullary	3	1.94%	0	3	
Grading	1	21	13.55%	1	20	0.0011 *
	2	69	44.52%	7	62	
	3	65	41.94%	19	46	
Class	Luminal A	61	39.35%	6	55	0.0013 *
	Luminal B	67	43.23%	9	58	
	Her2	12	7.74%	6	6	
	TN	15	9.68%	6	9	

Table 14: Patient and tumor feature frequencies and relative percentages are reported in relation to the final label (lymph node LN− involvement).
The difference between the two groups (LN+ vs. LN−) was reported, and the statistical significance was set at 0.05 (*). HT (hormonotherapy), IS curve/T (intensity signal curve/time), IDC (invasive ductal carcinoma), ILC (invasive lobular carcinoma), TN (triple negative).

Each of the performances of the different CNNs, based on each bounding option, is shown in Table 15. The best performances in terms of accuracy and AUC are obtained for the 2DS-NET: 78.63% and 77.86%, respectively. The 2DS-NET also showed the highest specificity; the highest sensibility was reported for the VB-NET based on the SVB and SIB as bounding

options. Moreover, the solution involving the SIB option obtained the highest performance in terms of K.

Model	Option	ACC	SPE	SENS	AUC	K
		Mean	Mean	Mean	Mean	Mean
SFB-NET	SFB	70.62%	75.92%	43.33%	69.52%	0.1349
VB-NET	SVB	76.79%	78.17%	**71.67%**	75.05%	0.3753
VB-NET	SIB	78.13%	78.82%	**76.67%**	77.13%	**0.4487**
VB-NET	SLVB	52.71%	53.24%	48.33%	53.11%	0.0124
VB-NET	SLIB	62.58%	71.79%	18.33%	47.34%	−0.0798
2DS-NET	2DS	**78.63%**	**85.75%**	46.67%	**77.86%**	0.2911

Table 15: Performance of the CNNs in LNS prediction (LN+ vs. LN−).
ACC (accuracy), SPE (specificity), SENS (sensibility), AUC (area under the curve), K (Cohen's kappa coefficient), SFB (single fixed-size box), SVB (single variable-size box), SIB (single isotropic-size box), SLVB (single lesion variable-size box), SLIB (single lesion isotropic-size box), 2DS (two-dimensional slice), and NET (network). The best test performances are evident in bold.

Figure 22 shows the ROC curves of the implemented experiments, obtained by plotting the true positive rate against the false positive rate at various threshold settings. It is possible to note that when the SLVB and SLIB options are used, the model performs worse than when a random classifier is used.

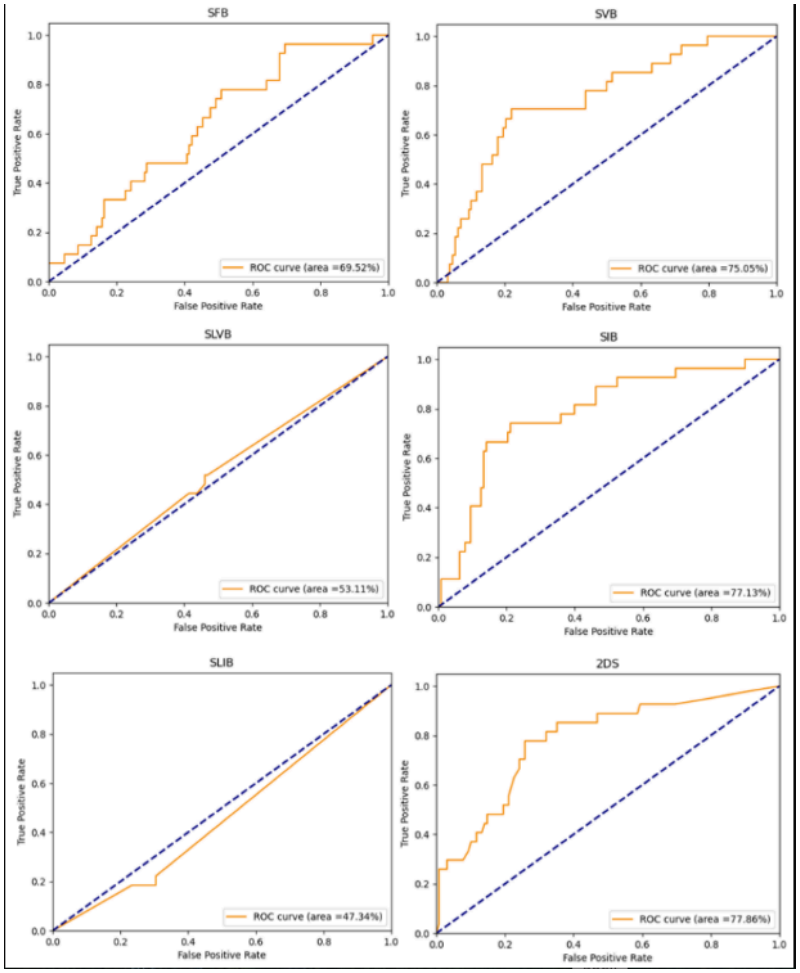

Figure 21: ROC curves of the implemented experiments. The title of each plot refers to the tumor bounding option that is used.

Figure 23 shows the precision-recall curves of the implemented experiments that report the tradeoff between the precision and recall for different thresholds. When the classes are severely imbalanced, they are an effective indicator of how well the predictions worked. In particular, considering LN+ as the positive class and denoting with *tp*, *tn*, *fp*, and *fn* the true positive, true negative, the false positive, and the false negative, respectively, the precision and recall are computed as follows: Precision=tp /(tp+fp); Recall=tp /(tp+fn).

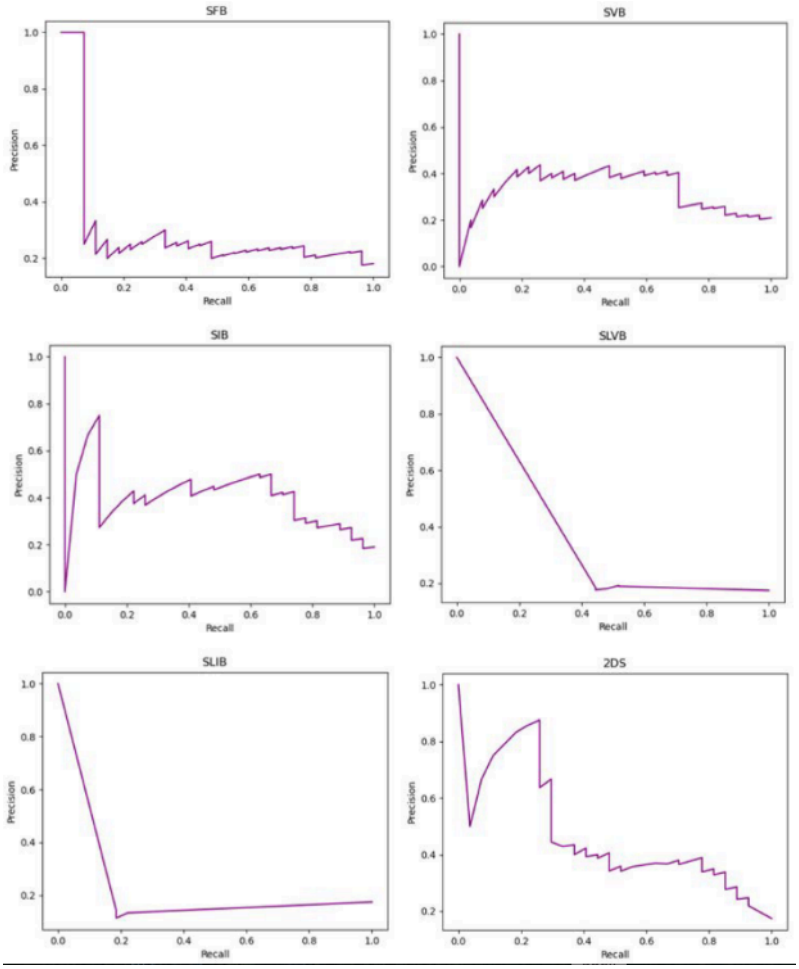

Figure 22: Precision-recall curves of the implemented experiments. The title of each plot refers to the tumor bounding option that is used.

In particular, the recall matches the sensitivity, while the precision represents the fraction of positive instances correctly classified among all the samples predicted as positive.

Finally, Figure 24 reports the confusion matrices computed by taking into account the predictions of the implemented models.

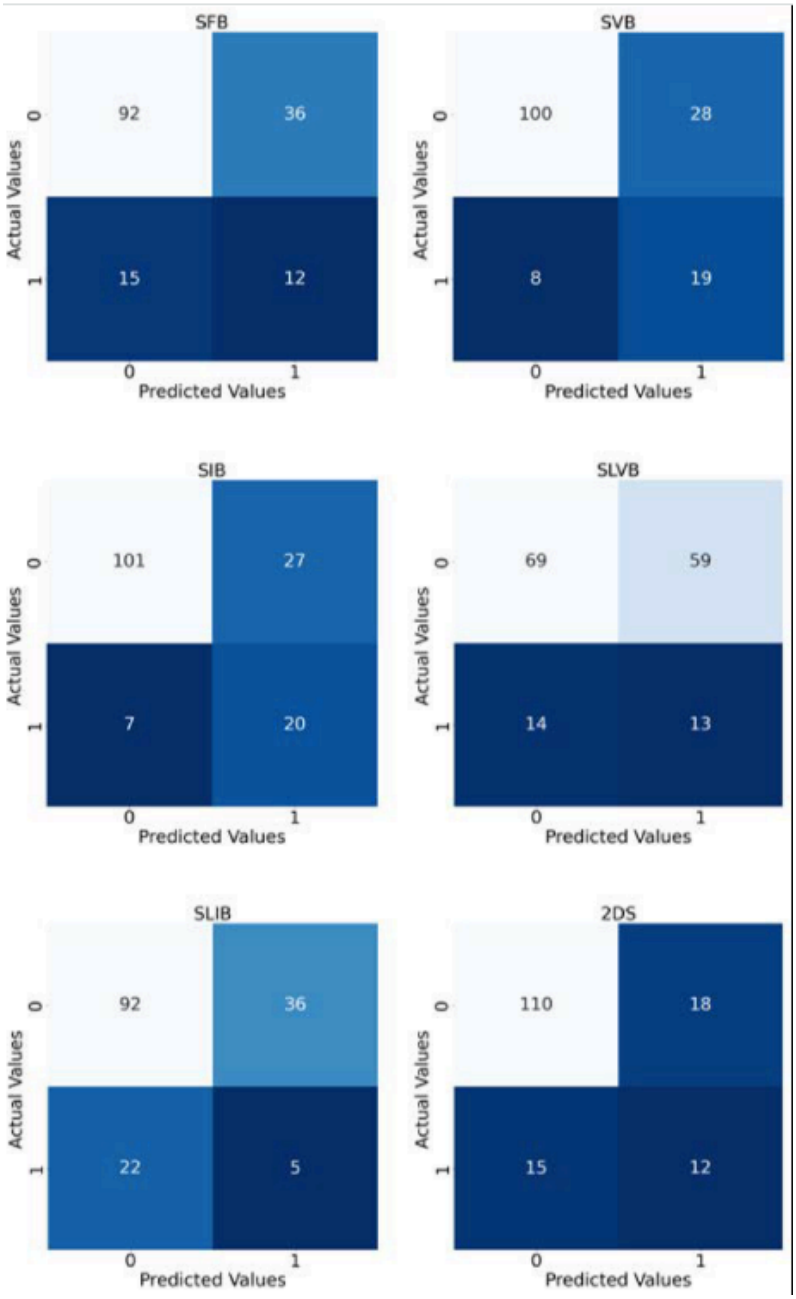

Figure 23: Confusion matrices of the implemented experiments.
The title of each matrix suggests the tumor bounding option that is used. In order to have a matrix for each experiment, we merged the predictions of the 10 folds considered as test sets. The colour depends on the number inside the square: the higher the number, the lighter the colour.

4.3.4 Discussion and Conclusions

It is now well established that the ALNS is directly dependent on tumor aggressiveness, which is reflected in a different representation of the breast lesion during DCE-MRI[39]. It has also been widely confirmed that the peritumoral tissue is influenced by tumor angio- and lympho-invasiveness[139,140,142]. Clinical and anamnestic data are actually not enough to determine the LNS with certainty, although some of these may benefit the doctors (in our case, cohort grading, class, and menopause status differ significantly between patients with lymph node involvement). In the majority of the works present thus far in the literature that exploits ML classifiers as predictors of the axillary cavity's status in patients with breast cancer, generally only the features of the primary lesion have been evaluated, without considering the breast parenchyma adjacent to the lesions and, in particular, the tissue between the multiple lesions in cases of multifocality and multicentricity[44,74,135].

Furthermore, these works are highly heterogeneous in terms of the involved dataset sizes, feature extraction/selection methods employed, and trained classifiers. Most literature proposals extract the shape, texture, morphological, and first-order features, while pharmacokinetic parameters are exploited only by Liu et al. [117]. In our previous work[73,134] 3D extension of local binary patterns (LBPs) was also explored as features to enrich texture description. Regarding the proposed classifiers, they are predominantly support vector machines, logistic regressions, linear discriminant analyses, and random forests.

In our previous work, we combined patients' clinical data, primary breast tumor histological information, and MRI radiomic features (First-Order, 3D Gray-Level Co-Occurrence Matrix, Three Orthogonal Planes-Local Binary Patterns) to predict the ALNS. In more detail, in a very innovative way, in this previous work, we considered both breast lesions and peritumoral regions for feature extraction considering a convex hull algorithm, which is the minimum volume bounded into a convex polygon and contains the ROI. The high dimensionality of the problem to solve (257 features) makes the wrapper features selection method necessary before using a random forest (RF) classifier to provide the prediction.

In this work, we make a step forward by exploiting DL approaches and CNNs, in particular, for ALN status prediction. Their ability to autonomously learn the set of features that fits the specific task to solve makes the feature extraction step unnecessary, which is typical of ML techniques, thus overcoming the problem of finding the most discriminating set of primary tumor characteristics. Moreover, we also evaluate how the peritumoral region affects the performance of the involved networks by investigating several tumor bounding options that differ according to the amount of healthy tissue to be included.

The implemented methodology consists of three main steps: the *imaging data definition*, used to prepare data belonging to the DCE sequences of different patients; the volume extraction and bounding options, which describe the tumor bounding options; and the CNN architecture for the ALN assessment, which introduces CNNs that differ according to the characteristics of the volumes considered as input.

Among all of the experiments, the results obtained involving the SIB option and the VB-NET showed the best performance. We argue that the obtained result is completely in accordance with the specific problem to be solved. In the SFB option, the fixed-sized bounding box may result in the inclusion of an excessive amount of healthy tissue with respect to the lesioned one, especially in patients with a small tumor lesion. As a consequence, the SFB-NET may focus on areas of the input volumes that do not contain useful information for the prediction and do not reflect the tumor's intrinsic behavior. The SVB and SIB options consider the smallest 3D cubical bounding box circumscribed to the tumor region, which limits the amount of the included peripheral and healthy tissue. As a result, the VB-NET directly takes the region of interest as input, leading to an increase in performance. However, the SVB extracts volumes whose voxels have different dimensions (in terms of mm) along the three spatial axes. In more detail, the spacing between slices is usually greater than the pixel spacing, resulting in a 3D cubical box whose dimension greatly depends on the x- or y-axes. As a consequence, the SVB option may introduce a high amount of healthy tissue along the z-axis, resulting in the need to introduce the SIB procedure that considers volumes with isotropic voxels. The considerations made between the SVB and SIB can be also used for the SLVB and SLIB. However, the results shown in Table 15 suggest that approaches that deal with a box for each lesion may not be the

best solution. We argue that the presence of multiple lesions may be relevant information for the prediction of axillary lymph node metastasis. Therefore, if the lesions are split into different boxes, such information may be lost when considering, in particular, the multifocal and multicentric tumors as a single entity and not as different tumors. Finally, the 2DS option that creates a set of 3D slices for each patient, has the advantage of increasing the size of the dataset. However, when comparing the results with the solution involving the SIB procedure, it is possible to note a high difference in terms of sensitivity that makes us confirm the SIB as the preferred bounding option.

Moreover, Table 15 shows that the different bounding box options affect the results. In particular, the solutions involving the SIB and SVB options are the approaches with the more balanced results of the two classes, providing also the largest performances in terms of sensitivity (76.67% and 71.67%, respectively). On the contrary, the other bounding box options tend to favor the negative class (LN–); we deem that this cannot be considered a limitation of the work, but rather a result. Indeed, an explanation can be found in the following reasons:

- In the SFB option, the fixed-sized bounding box may result in an excessive amount of healthy tissue with respect to the lesioned one;
- We argue that the presence of multiple lesions may be relevant information for the prediction of axillary lymph node metastasis. Therefore, when the SLVB and SLIB options are used, such information may be lost;
- The 2DS option considers 2D slices and, thus, does not exploit volumetric features.

To the best of our knowledge, this paper represents the first work exploring the use of different primary tumor bounding options and convolutional neural networks to evaluate how the peritumoral region affects the ALNS assessment.

The work proposed by Nguyen et al. [**25**] represents the only one that has aimed to predict ALN metastasis using primary tumor DCE-MRI features with a DL approach. In that work, a 3D CNN is implemented to process the DCE-MRI images using a subtractive approach that works with the third, fourth, and fifth post-contrast volumes. A 3D cuboidal bounding box of the size $50 \times 50 \times 50$, encompassing the tumor region, is used to crop the DCE-MRI data.

Nguyen et al. obtained the best values with 64.64%, 69.29%, 37.37%, and 58.56% for ACC, SPE, SENS, and AUC, respectively; all are lower compared to our best results (78.06%, 78.91%, 74.07%, and 75.84%), although they are not directly comparable since the datasets are different.

In our opinion, the fixed-size bounding box used in Nguyen et al. reduces the generalization capability of the implemented model. We also consider the absence of significant differences between our architectural model and that used by Nguyen et al. Furthermore, the authors included patients from two different hospitals, which may influence the imaging acquisition homogeneity.

Our work demonstrates that the peritumoral parenchyma, and, in particular, ones located between multiple lesions in the case of a multicentric or multifocal tumor, contains information not visible to the radiologist's eye that correlates with the metastatic spread to the lymph nodes of the axillary cable. Parenchyma near the tumor lesions in current MRI clinical practice is not considered. In detail, in our work, the bounding box that has shown the best results is that which includes the parenchyma closest to the lesions and between multiple lesions but excludes the parenchyma furthest from the lesions themselves. These data suggest that there is an invisible cellular diffusion near the tumor lesions that artificial intelligence can help to reveal.

In our previous works, we have shown how the features contained in the primary tumor, first shown in [74,117], and in the peritumor edema, then explored in [139], significantly influence the LNS. The demonstration that the contribution of the peritumoral parenchyma is significant paves the way for new research, which has had little or no exploration in the current literature.

The main limitation of our study is the small number of patients that belong to a single medical center. Although the size of the dataset is similar to that used in other works[110,117,119,136], which analyzed samples in the range of 146–164 patients, a future perspective aim is to include data collected from different centers and also to evaluate how different image modalities contribute to the prediction of axillary lymph node metastasis. Another important limitation is the imbalance of class priors (22 patients with LN+ vs. 128 patients with LN–), which, in several

cases, resulted in sensitivity values lower than the specificity ones, although we implemented standard techniques to cope with this issue.

In our future work, we want to implement data and test CNNs on different datasets and validate the bounding box technique on different kinds of images and MRI sequences. Furthermore, we want to investigate the role of the individual histological parameters reported in this article in relation to the radiomics data in the prediction of LNS.

However, these results are absolutely noteworthy in relation to the role played by the peritumoral parenchyma.

In conclusion, we demonstrated that a selective inclusion of peritumoral tissue increases the performance of the CNNs in predicting lymph node involvement and that the tissue localized among different tumoral lesions, especially in cases of multifocal and multicentric cancers, is strongly related to angio- and lymph-invasiveness and to lymph node metastasis. This tissue is barely evaluated by the radiologist's eye on all MRI sequences and, in this way, an AI approach may represent a valid supporting tool.

Of course, more studies are needed to confirm these results. Future perspectives include increasing the number of patients, in particular patients with a positive LNS, to test the reproducibility of the results on different samples (different MRI and samples) and to include semantic features in the CNN analysis.

4.4 The DL contribute to predict LNS using a multimodal approach

To our knowledge only few works used DL approaches to predict lymph node status in BC patients using MR images.

Gao et al evaluated DCE-MRI of 941 patients and proposed a DL model containing a 3D Deep residual network (ResNet) architecture and a convolutional block attention module called RCNet. Three RCNet models were used based on tumor, ALN and combined tumor-ALN regions. The latter showed the highest auc 0.907 in the internal test cohort and 0.852 in the external test cohort[18].

Nguyen et al[19] study the preoperative DCE MRI 4D of 357 patients from two hospitals through CNN 2D 3D 4D and 4D that integrate multiple types of data in order to prevent lymph node metastases, with a 10 Fold cross validation and reaching a sensitivity of 72% and an AUROC of the 71%.

The first used only features extracted from DCE-MR image. On the other hand, Nguyen et al for the first time used a multimodal approach including histological features. In this work, we introduce an MDL-based methodology for ALNS evaluation, exploiting three MRI sequences (DCE, T2, and DWI), and clinical and histological characteristics of the primary tumor. In particular, we implement a Multi-Input Single-Output 3D CNN where the TF makes the different modality-specific DNNs influence each other in the extraction of features constituting the shared representation, thus improving the integration of different image modalities.

4.4.1 Population

4.4.1.1 MRI Examination

All the MRI examinations of the breast performed at the Central Radiology Department of Policlinico Umberto I, from January 2017 until January 2020 performed for tumor staging, were retrospectively evaluated. A total of 153 subjects (average age 55 years; range 30–85),

with 155 malignant BC lesions, were included in the study. The MRI investigations were performed with a 3 Tesla magnetic field using a Discovery 750 machine (by GE Health- care, Milwaukee, WI, USA), using an 8-channel coil (8 US TORSOPA) dedicated for breast study, with the patient in prone position. A first sequence was carried out along the three orthogonal planes of the space for framing. Then the following sequences were performed:

- T2-weighing single shot fast spin eco with Dixon technique (IDEAL) for the intravoxel separation of the water/fat interface (TR/TE 3500–5200/120–135 ms, matrix 352 × 224, FoV 370 × 370, NEX 1, slice thickness 3.5 mm)
- Axial, single shot, fat suppressed planar echo, diffusion-weighted (DWI) sequences with a diffusion-sensitized gradient applied along the x, y, z axes and with b values of 0 and 1000 sec/mm^2 (TR/TE 2700/58 ms, matrix 100 × 120, FOV 360 × 360, NEX 6, slice thickness 5 mm). From each DWI sequence a relative quantitative (ADC) map was generated[143].
- Axial T1-weighted 3D dynamic gradient echo sequences (DCE), with suppression of fat signal (VIBRANT), before and 4-6 times after intra- venous contrast medium administration (TR/TE 6.6/4.3 ms, flip angle 10°, matrix 512 × 25, NEX 1, slice thickness 2.4 mm). The contrast agent used during the examination was gadobenate dimeglumine (Multihance®; Bracco s.p.a, Milan, Italy), administered at a concentration of 0.2 mmol/kl through intravenous injection with a 20 Gauge cannula at a rate of 2 ml /sec and followed by the infusion of 15 ml of physiological saline at the same rate. Therefore, subtracted images were automatically obtained using the post-contrast images.

Each exam was transferred to a workstation (Advantage Windows Work- station 4.4; GE Medical System, Milwaukee, WI, USA) for post-processing analysis. The images were analyzed by two radiologists with 10 and 3 years of experience and the following characteristics were recorded: tumor localization based on breast quadrant; tumor distribution (unifocal, multifocal, multicentric); diameter of the target lesion; tumor margins (regular, irregular, lobulated, or spiculated); lesion intensity signal timing curve (IS/T curve) on DCE

sequence; visibility on T2 of the lesion; visibility on DWI of the lesion; apparent diffusion coefficient values (ADC).

For each patient, specific anamnestic-clinical data was previously collected, on the basis of which the patients were divided into groups as follows: age; menopausal status (pre- or post-menopausal); hormonal therapy (patients who have performed at least 3 continuous months of hormone therapy of any type, namely contraceptive, replacement therapy or therapeutic); familiarity (patients with at least 2 female or male family members affected by breast cancer at any age).

4.4.1.1 Histological Data

The histological examination was performed on tumor material obtained through core-biopsy or surgery and analyzed by a pathologist with more than 15 years of experience. The tumor histological grade value was assigned in accordance with the Next-generation Sequencing (NGS) for which a score from 1 to 3 was given. The immunohistochemical analysis was conducted, evaluating the estradiol (ER), the progesterone (PgR), the herceptin2 receptor (HER2) and the proliferation index Ki67.

The state of the axillary cavity was histologically determined after the diagnosis of breast cancer. The patients in our study were simply classified as positive (LN+) or negative (LN-), depending on whether there was, in the former case, at least one lymph node involved, or, in the latter case, no positive lymph nodes. On this basis, the dataset accounts for 27 positive and 128 negative samples.

4.4.1.1 Images analysis

The images were anonymized and uploaded on a dedicated open-source soft- ware (3D Slicer, version 4.8, November 2012). An identification number (ID) was assigned to each patient. Bilateral tumors were considered with two different IDs. For each case, the subtracted post-contrast T1-MRI was selected. The second phase (60–120 s) was selected for ROI segmentation, due to its higher contrast resolution. Then, a label map was created. The lesions were manually drawn through manual and assisted thresholding segmentation techniques on the axial projection. When present, necrosis was avoided by segmentation. For multifocal or multicentric tumors, all lesions, even the smallest, were segmented.

The features collected for each patient can be grouped in three main classes, namely, clinical, histological, and image-derived. For the sake of simplicity, the three classes are indicated only with clinical features (CL) and are summerized in Table 16.

Class	Details
Clinical	age, familiarity, hormone therapy, menopausal status
Histological	ER, PgR, HER2, ki-67, grading, tumor class, histotype
Image-Derived	visibility on T2, visibility on DWI, apparent diffusion coefficient values (ADC), signal timing curve, dimensions, margins breast quadrant, multifocality

Table 16: Details about the collected clinical features.

4.4.2 Methodology

The presence of clinical features and multiple sequences acquired during the MRI scan results in the need of merging information coming from heterogeneous sources. Multimodal learning provides fusion techniques that vary according to when the integration is performed, exploiting different levels. The early fusion acts at data-level, using a single model for the specific task to

solve. As a consequence, it only requires a single training stage, moving the complexity of the solution to the data integration step, where different modalities are fused in a common representation. The late approach represents a decision-level method, requiring a model for each data source, with a high computational cost. The intermediate fusion is the most flexible solution, since it exploits the layers in a deep-based architecture, enabling the integration of the learned representations at various levels of abstraction [144]. Indeed, it consists of the presence of different modality-specific paths that learn the features representation of each data source, with the aim of providing a shared representation processed by a shared path. The involvement of deep learning approaches, such as CNNs, and the adopted training strategy with backpropagation make the architecture find the best shared representation for the specific task to solve, overcoming the limitation of the early fusion method regarding the integration of data in common features set that is determined before the use of a model. Moreover, in contrast to the late fusion, in the IF technique, the heterogeneous sources simultaneously contribute to the decision, avoiding the implementation of a final decision step. In the case of CNNs, the multiple modality-specific paths provide features maps, directly extracted from each data source, that are defined during training iterations to be well suited to the task to be solved. The aim of this paper is to propose a transfer module that improves the integration of heterogeneous modalities, implementing the features maps calibration during their extraction in the modality-specific paths. The calibration consists of allowing the features maps to influence each other through an alteration process aimed at reducing the contribution of less significant features. The backpropagation algorithm in the training phase allows the creation of a calibration customized for the task under consideration.

The implemented methodology consists of two main steps: the Pre-processing, used to prepare data belonging to different patients, including both MRI sequences and clinical information, and the IF approach for ALN status assessment, introducing the implemented methodology based on intermediate fusion technique with the proposal of a Transfer Model (TM), that contributes into making the features maps extraction process dependent on heterogeneous data modality.

4.4.2.1 Pre-processing

In MRI examination, the DCE sequence consists of the intravenous injection of a contrast agent (CA), whose absorption and release determine the specific wash-in and wash-out times respectively. Indeed, the DCE involves the acquisition of 3D volumes at different times, considering MRI images taken before (pre-contrast) and after (post-contrast) CA injection. The result is a 4D data with three spatial and one temporal dimension, that can be interpreted as a 3D image with several channels. Denoting with t_i the i_{th} acquisition, where t_0 is the pre-contrast volume, the subtractive series is obtained by considering $t_i - t_0$ with i raging from 1 to $T^p - 1$, that corresponds to the number of acquisitions for the patient p. Following the methodology proposed in [142], four specific subtractive volumes are selected: the first, the second, the last volume, and the median index volume between the third and the second-to-last volume, with the aim of preserving information about the wash-in and wash-out of CA owing.

Differently from DCE, T_2 and DWI scans consist of the acquisition of a 3D volume without the temporal information, which can be considered as a 3D image with 1 channel. Since the breast lesion is segmented considering the second subtractive volume, the T_2 and DWI acquisitions are aligned to the segmentation mask generated from the T1 volume exploiting the Slice Location attribute that corresponds to the relative position of each slice. In particular, slices belonging to different acquisitions with the same Slice Location value are considered aligned. As a consequence, it is possible to apply the segmentation mask to all the sequences by taking into account only the aligned slices. Moreover, before applying the information about the lesion localization, the DWI scan is co-registered to the DCE volume considering the second post-contrast acquisition as a reference and using mutual information as a similarity metric[145,146].

In the study proposed in [142], different tumour bounding options are analyzed considering the characteristics of the DCE dataset for the assessment of ALN status. In particular, the work described in [142] evaluates how the amount of the included non-tumour tissue impacts the ALN assessment, taking into account that patients differ in size and number of lesions. Among all the proposed tumour bounding options, the Single Isotropic-size Box (SIB) presents the best

performance for the evaluation of the metastasis. It considers the fact that different patients may show acquisitions that vary in terms of resolution, namely the information regarding the measurement of each voxel in millimetres (mm). As a consequence, all the volumes are re-sampled to obtain isotropic voxels with dimension $1 \times 1 \times 1 mm^3$, before selecting the smallest box surrounding the tumour area. According to the results re- ported in [142], the SIB bounding option is selected and applied to DCE, T_2, and DWI volumes.

Information coming from images needs to be merged with the clinical features (CL) acquired as described in Sections 3. In particular, all the collected data are detailed in Table 16 consisting of 19 features. However, a subset of them is selected by the domain expert (D.S), obtaining the subset CL-S1. The basic idea is to remove all the features which can be derived from the images, retaining those that provide additional information to the acquisitions. Inspired by [19], the CL-S1 includes age, familiarity, hormone therapy, menopausal status, dimensions, ER, PgR, HER2, ki-67, and grading, resulting in a set of 10 features.

4.4.2.2 IF approach for ALN status assessment

In this paper, we propose a methodology based on intermediate fusion as its flexibility enables overcoming all the limitations observed in early and late methods. Moreover, the implemented solution focuses on the proposal of a module improving the integration of heterogeneous modalities and on the definition of a neural network exploiting both images and tabular data. The resulting network is a Multi Input - Single Output CNN-based architecture able to process both MRI sequences and clinical information.

4.4.2.3 Transfer Module (TM)

We aim to improve the integration of different data modalities at an intermediate level, proposing a Transfer Module (TM) that makes the specific modality paths influence each other

while extracting the features maps. In particular, inspired by the works presented in [53; 60], TM is able to perform a cross-modality calibration by implementing a gating procedure to reduce the importance of the less significant components. It can be inserted be- tween layers belonging to different convolutional cores to take into account the complementary characteristics of the images in the determination of the resulting features maps.

For each modality j, we denote the output of the L_j^i as $F_j^i \in R^{X_j^i \times Y_j^i \times Z_j^i \times C_j^i}$, where $X_j^i \times Y_j^i \times Z_j^i$ is the spatial dimension, while C_j^i is the number of channels. The TM inserted in the layer i is a multi-input multi-output module since it considers as input M features maps F_j^i and provides M outputs $\tilde{F}_j^i$, corresponding to the calibrated versions of the features maps, that are obtained by applying the gating procedure. TM is organized in two different steps, namely the Shared vector computation and the Multimodal calibration, as summarised in Figure 2.

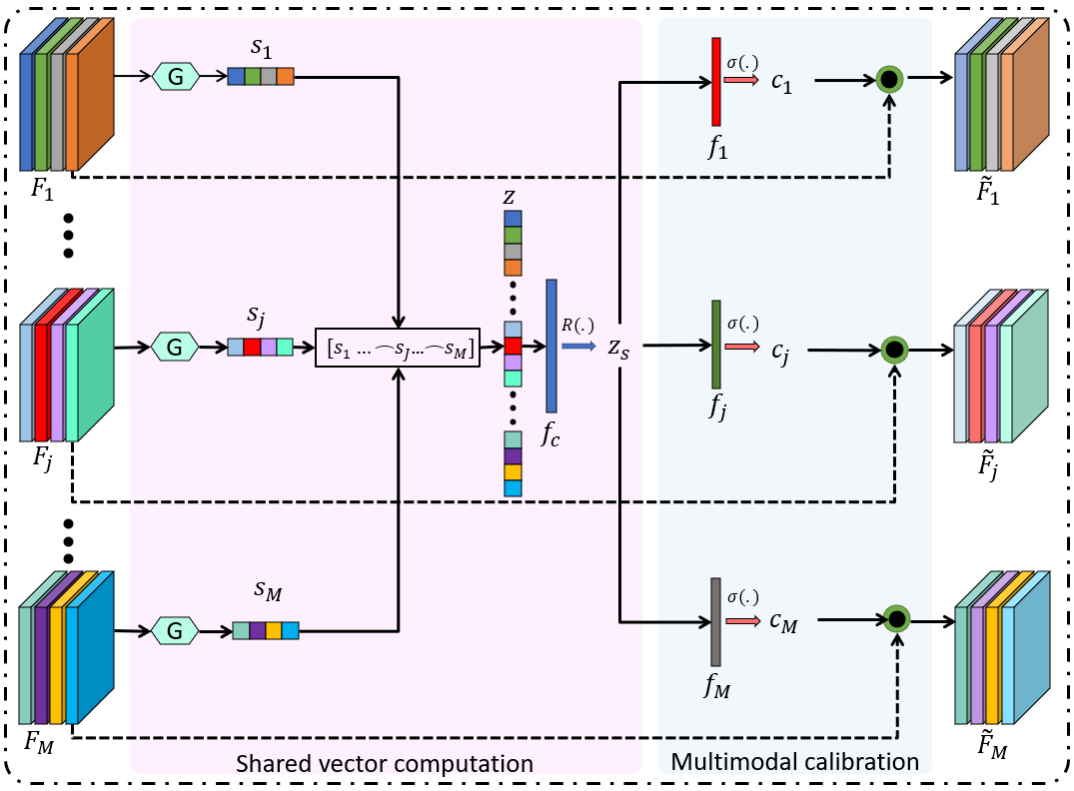

Figure 24: Architecture of the proposed TM module for three different image modalities.
The input consists of the features maps coming from the modality-specific paths, that are further processed by considering two steps acting in a pipeline, namely the Shared vector computation and the Multimodal calibarion. It is worth noting that the index i of the layer in which the module is inserted is omitted to avoid overly complex notation in the image.

The first one aims to determine the shared representation z_s^i of the i – th layer considering the vectors computed from the features maps. The channel descriptor vector $s_j^i \in R^{1\times C_j^i}$ for each features map is obtained by using a Global Average Pooling operation $G(\cdot)$, that acts in each channel computing the mean value and providing a channel-wise descriptor. The concatenation of all the s_j^i determines the vector $z^i \in R^{1\times C^i}$, where $C^i = \sum_{j=1}^{M} C_j^i$. Then, z^i is further processed by considering it as input for the fully connected layer f_c, followed by the ReLU function $R(\cdot)$, thus marking the first difference from the module presented in [53]. Indeed, the presence of ReLU function introduces a nonlinear map between the elements of the input vector z^i, consisting of descriptors relating to the different image modalities. The result is the shared representation $z_s^i \in R^{1\times C^t}$, where $C^t = C^i/r$, and r is the reduction ratio. The Shared vector computation can be summarized as follows:

1. s_j^i, with j from 1 to M
2. $z^i = [s_1^i \frown s_2^i ... \frown s_j^i \frown s_M^i]$
3. $z_s^i = R(f_c(z^i))$

The Multimodal calibration uses the shared representation z_s^i to calibrate the features maps F_j^i, thus exploiting information coming from different data sources. z_s^i is considered as input for M fully connected layers f_j, each for a specific modality j. The Sigmoid activation function $\sigma(\cdot)$ is used to force the output in the range [0,1], resulting in M calibration vectors $c_j \in R^{1\times C_j^i}$. The features maps $\tilde{F}_j^i$ are obtained by implementing a channel-wise product between the

input F_j^i and the corresponding c_j, creating a gating mechanism, where the contribution of the selected filters is reduced. This allows each F_j^i to be influenced by the others during the features extraction step. In contrast to the solution proposed in [53], we do not change the output range of the Sigmoid function with the multiplication by a scalar value, obtaining a self-gating mechanism, that is able to emphasize the most informative characteristics while suppressing the less useful ones [60]. The Multimodal calibration step can be summarised as follows:

1. $c_j = \sigma(fc_j(z_s^i))$ with j from 1 to M
2. $\tilde{F}_j^i = c_j \odot F_j^i$ with j from 1 to M , where $\odot$ represents the channel-wise product

The proposed transfer module exploits the underlying connections between information coming from heterogeneous and complementary data sources, fully extending the work proposed in [60], which studies the relationship be- tween channels, into a multimodal scenario. Indeed, the implemented gating mechanism acts on the channels of the extracted features maps, leveraging the shared representation to highlight in each data source the characteristics that contribute the most to the task under consideration. When it is inserted between the i–th layers of the convolutional cores, each $\tilde{F}_j^i$ is affected by the other M – 1 modalities, allowing the integration of information from the first levels of the network and enhancing the modality-specific paths dependency. The proposed TM can be used to implement a cross-modality calibration in application with a generic number of image modalities. In this paper, we consider three different image sources (M = 3), with $j \in \{DCE, DW\ I, T_2\}$.

4.4.2.4 CNN Architectures

In this paper, we focus on four different data sources, namely three MRI sequences, and the set of clinical information CL-S1. As a consequence, the proposed network is a Multi Input - Single output CNN, whose architecture is presented in Figure 3 and consists of 3 convolutional cores, the Multilayer Perceptron (MLP) for the clinical features, and two concatenating

operations. In particular, two different CNNs are exploited for the definition of the convolutional (ConvC) and the classification (CC) parts, regardless of the specific data modality considered. Such two networks are named in the following as BasicNet and ResNet.

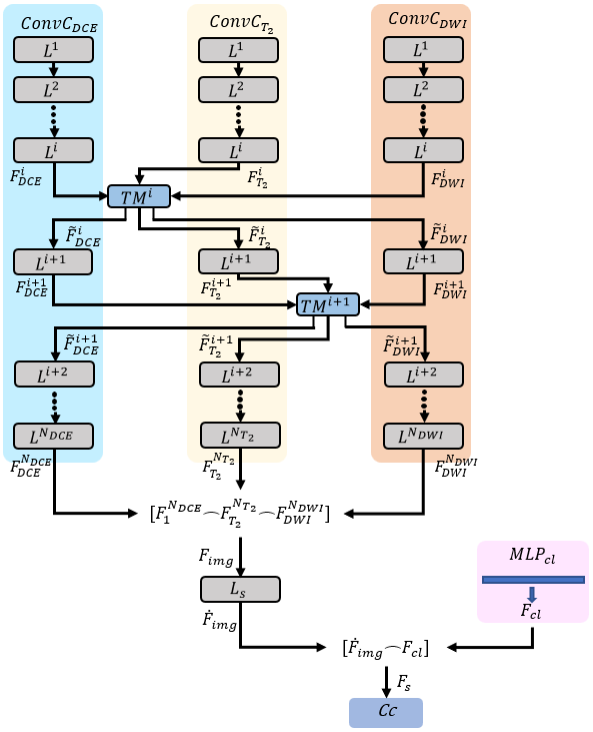

Figure 25: Architecture of the IF approach including TM between layers belonging to different convolutional cores and clinic features

The ConvC of the BasicNet, responsible for features maps extraction, consists of five reduction layers, including blocks with a 3D-convolutional operation, followed by batch normalization and ReLU function. Each convolutional layer consists of a 3D operation with 4×4×4 kernels and a stride set to 2 in order to extract features from the input volume while having a gradual dimensionality reduction. The padding is set to 1 in each layer, excluding the last one where it is set to 0. Moreover, each layer doubles the number of channels, while the convolutional operation in the first block presents output channels. In the BasicNet architecture, the classification core consists of two fully connected layers, resulting in the architecture shown in Figure 26a.

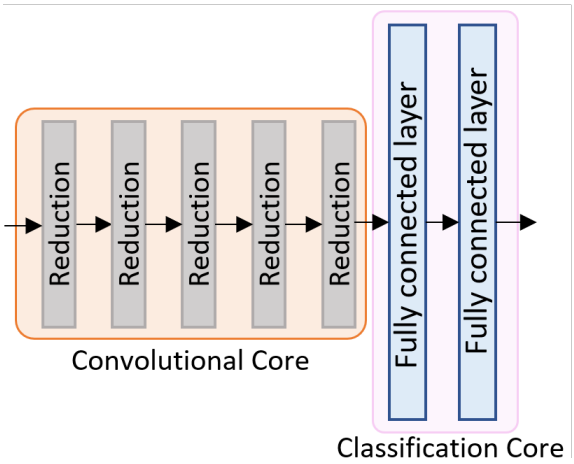

Figure 26a: Architecture of the CNNs used to de ne the convolutional and classification cored of the proposed network.
BasicNet architecture: the network consists of five reduction layers, representing the Convolutional Core (Conv-C), followed by two fully connected layers spaced by ReLU function, constituting the Classification Core (CC).

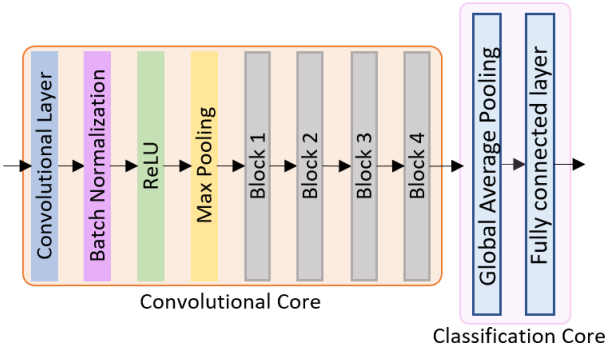

Figure 27b: Architecture of the CNNs used to de ne the convolutional and classification cored of the proposed network.
ResNet architecture: the Convolutional Core (Conv-C) consists of a first convolutional layer, followed by batch normalization, ReLU and Max Pooling layers, and a chain on four layers containing the blocks implementing the residual network, while the Global Average Pooling and the fully connected layer represent the Classification Core (CC).

The ResNet involved in this paper is inspired by the state-of-art net- work proposed in [147], considering the work described in [148], where 3D ResNet variants are implemented, offering a set of architectures pre-trained with medical images for segmentation tasks. The architecture consists of a ConvC, which is an encoder with residual layers, and a set of decoders for the generation of the segmentation masks. Hence, the ResNet is adapted for the classification task

as proposed in [148]: the encoder is retained while the set of decoders is replaced with a CC consisting of a global average pooling and a fully connected layer. Figure 26b shows the resulting ResNet architecture highlighting the ConvC, consisting of a first convolutional layer, followed by batch normalization, ReLU and Max Pooling layers, and a chain on four blocks containing the layers implementing the residual network as proposed in .

In both networks, the number of input channels of the first convolutional layer in the ConvC depends on the imaging modality. In particular, the first layer in the $ConvC_{DCE}$ presents 4 channels, while $ConvC_{T2}$ and $ConvC_{DWI}$ consider 1 channel volumes as input. The CNN proposed in this paper exploits for each $ConvC_j$ only the convolutional part of BasicNet or ResNet.

The MLP used to process CL-S1 presents a fully connected layer with 10 input and 4 output features, followed by the ReLU function.

In the multimodal scenario, described in Figure 3, the first concatenation operation integrates F^{NDCE},F^{NT2},F^{NDWI}, obtaining the F_{img} features map, DCE T2 DWI with C channels, where $C = \sum_{j=1}^{M} C_j$ and C_j is the number of channels of the features map related to the modality j at the end of its convolutional core. Then, the shared image features map F'_{img} is generated by the block L_S that includes a convolutional operation with a number of input and output channels set to C, a $1 \times 1 \times 1$ kernel, values of stride and padding set to 1 and 0 respectively, followed by batch normalization and ReLU function. Moreover, when the ResNet architecture is exploited for the ConvC, L_S also includes a Global Average Pooling layer to obtain a features vector. The second concatenation operation merges the F'_{img} with features vector F_{cl} coming from the MLP that processes the clinical information. The resulting representation F_S is considered as input for the CC, which consists of two fully connected layers spaced by ReLU function. In particular, the first layer in the classification core considers an input features vector of C + 4 elements, with C/3 output features, while the second one is responsible of the prediction, presenting two output neurons.

4.4.3 Results

This section reports in Table 17 the results of the implemented experiments, considering both the proposed solution (IF-MT) and different fusion techniques. In particular, each configuration is detailed according to the involved network, the ML model, the used data modality, and the fusion approach in columns Net, Model, Modality, and App respectively. Moreover, the performance is evaluated in terms of ACC, Spe, Sens and AUC, reporting the average rate computed adopting a 10-fold CV and the standard deviation after the ± symbol. Table 17 is organized into seven main sections.

Net	Model	Modality	App	Acc	Spe	Sens	AUC
-	SVM	CL	U	**76.77±0.12**	78.13±0.13	**70.37±0.19**	71.53±0.21
	MLP	CL-S1	U	75.97±0.11	**78.74±0.10**	62.96±0.22	**75.61±0.17**
BasicNet	-	DCE	U	78.06±0.11	78.91±0.12	74.07±0.15	78.53±0.09
		T2	U	74.84±0.11	78.13±0.11	59.26±0.28	62.36±0.21
		DWI	U	79.87±0.13	85.04±0.12	55.56±0.25	66.00±0.23
		ALL	IF - MT	**87.01±0.08**	**88.19±0.08**	**81.48±0.17**	**90.14±0.10**
		ALL	IF	84.42±0.11	85.04±0.12	**81.48±0.17**	81.28±0.15
	NN	ALL	EF(CL-S1)	79.87±0.13	81.10±0.14	74.07±0.15	80.87±0.15
	SVM	ALL	EF(CL)	62.34±0.15	62.99±0.14	59.26±0.22	61.13±0.19
	SVM	ALL	LF	84.52±0.11	86.72±0.11	74.07±0.14	78.62±0.15
ResNet10	-	DCE	U	85.16±0.07	87.50±0.07	74.07±0.17	81.34±0.12
		T2	U	85.16±0.08	89.84±0.09	62.96±0.08	76.71±0.14
		DWI	U	83.77±0.08	88.98±0.09	59.26±0.11	71.60±0.15
		ALL	IF - MT	**90.91±0.08**	92.91±0.07	**81.48±0.17**	87.17±0.14
		ALL	IF	90.26±0.08	92.13±0.08	**81.48±0.17**	87.43±0.12
	NN	ALL	EF(CL-S1)	84.42±0.09	86.61±0.09	74.07±0.15	85.62±0.11
	SVM	ALL	EF(CL)	85.06±0.08	**98.43±0.08**	22.22±0.25	60.32±0.23
	SVM	ALL	LF	89.03±0.07	92.19±0.07	74.07±0.05	**92.85±0.10**
ResNet18	-	DCE	U	84.52±0.08	86.72±0.08	74.07±0.15	80.93±0.10
		T2	U	78.71±0.13	82.03±0.16	62.96±0.08	72.60±0.16
		DWI	U	81.17±0.11	85.83±0.12	59.26±0.11	71.30±0.13
		ALL	IF - MT	**89.61±0.06**	**91.34±0.06**	**81.48±0.17**	85.04±0.14
		ALL	IF	86.36±0.10	88.19±0.10	77.78±0.16	84.54±0.14
	NN	ALL	EF(CL-S1)	83.12±0.10	85.04±0.10	74.07±0.15	84.60±0.12
	SVM	ALL	EF(CL)	66.88±0.10	67.72±0.12	62.96±0.14	65.34±0.19
	SVM	ALL	LF	83.23±0.08	85.94±0.08	70.37±0.10	**86.98±0.10**
ResNet34	-	DCE	U	78.71±0.10	80.47±0.12	70.37±0.15	74.13±0.12
		T2	U	76.77±0.10	80.47±0.13	59.26±0.19	67.85±0.16
		DWI	U	74.03±0.08	77.95±0.10	55.56±0.20	64.36±0.13
		ALL	IF - MT	**87.66±0.10**	**89.76±0.09**	**77.78±0.16**	83.79±0.15
		ALL	IF	87.01±0.07	88.98±0.05	**77.78±0.16**	**84.49±0.12**
	NN	ALL	EF(CL-S1)	83.77±0.08	86.61±0.09	70.37±0.15	77.72±0.14
	SVM	ALL	EF(CL)	74.03±0.07	81.89±0.10	37.04±0.25	59.46±0.21
	SVM	ALL	LF	81.29±0.10	82.03±0.11	**77.78±0.14**	82.96±0.19
ResNet50	-	DCE	U	78.06±0.08	79.69±0.08	70.37±0.20	82.26±0.11
		T2	U	67.10±0.10	68.75±0.13	59.26±0.11	64.06±0.09
		DWI	U	72.08±0.12	75.59±0.13	55.56±0.25	64.01±0.17
		ALL	IF - MT	**84.42±0.11**	**85.83±0.11**	**77.78±0.16**	83.96±0.16
		ALL	IF	81.82±0.11	82.68±0.12	77.78±0.16	81.31±0.13
	NN	ALL	EF(CL-S1)	79.87±0.09	81.89±0.10	70.37±0.15	81.16±0.12
	SVM	ALL	EF(CL)	67.53±0.15	68.50±0.15	62.96±0.21	65.73±0.17
	SVM	ALL	LF	80.00±0.11	81.25±0.11	74.07±0.11	**85.19±0.15**
CNN[43]	-	DCE	U	63.76±0.16	69.67±0.19	41.67±0.38	65.21±0.22
	-	DCE+CL-4	IF	**68.18±0.15**	**70.87±0.21**	**55.56±0.33**	**67.07±0.20**

Table 17: Performance of the implemented experiments evaluated in 10-fold CV setting considering different models and multimodal data fusion techniques.

The first one does not consider any CNN since it focuses on the clinical features represented in the form of tabular data. In particular, the first and second rows exploit the SVM classifier with the CL set and the MLP model with CL-S1 set respectively. The experiment involving the CL features presents a slightly better performance in terms of ACC and Sens, gaining 76.77±0.12% and 70.37±0.19% respectively. On the other hand, the MLP model achieves 78.74±0.10% and 75.61±0.17% of SPE and AUC respectively, using the CL- S1 set.

In the following parts, CNNs are exploited, presenting the proposed IF-MT architecture and several variants. In the configurations considering the U approach, the network is trained for each imaging modality, leveraging the ne-tuning in the case of ResNet [61]. In the EF(CL-S1) and EF(CL), all data modalities are involved, exploiting the network as a features extractor and performing the classification using the SVM and NN respectively. The solution with the LF merges the predictions coming from the CNNs trained with the unimodal approach and the experiment presented in the first row of the table with the SVM model and CL set.

The results obtained with the BasicNet are reported in the second section. It is possible to note that the configuration using the IF-MT approach outperforms the others in all metrics, achieving 87.01±0.08% in ACC, 88.19±0.08% in Spe, 81.48±0.17% in Sens and 90.14±0.10% in AUC. More- over, the EF(CL-S1) exceeds the EF(CL) by a large margin (at least 15%) in each metric, while the LF shows an increase in terms of ACC and Spe compared to EF solutions.

The third part of the table presents the results involving the ResNet10. The proposed approach (IF-MT) reports the best performance in terms of ACC and Sens, obtaining 90.91±0.08% and 81.48±0.17% respectively, improving the U approaches in all metrics. It is worth noting that although the EF(CL) presents the highest specificity, it shows the lowest value of Sens (22.22±0.25%). Furthermore, the experiment with the LF outperforms the EF solutions, achieving also achieving the highest AUC (92.85±0.10%).

The fourth section focuses on the ResNet18, where the methodology considering the IF-MT obtains 89.61±0.06%, 91.34±0.06%, 81.48±0.17%, and 85.04±0.14% in terms of ACC, Spe,

Sens, and AUC respectively. The experiments with the U approach present lower values in all metrics than both the IF and IF-MT solutions, while the LF continues to outperform both the EF(CL-S1) and EF(CL) configurations, showing 83.23±0.08%, 85.94±0.08%, 70.37±0.10%, and 86.98±0.10% in terms of ACC, Spe, Sens, and AUC respectively.

The performance obtained with the ResNet34 is presented in the fifth part of the table, where the methodology proposed in this paper still reports the best performance, achieving 87.66±0.10% in accuracy, 89.76±0.09 in specificity, and 77.78±0.16 in sensitivity. On the other hand, the IF approach shows the same value in terms of Sens and slightly better performance in AUC, obtaining 84.49±0.12%. Moreover, the EF(CL-S1) exceeds the EF(CL) in each metric, reporting an improvement of at least 30% in the sensitivity. The methodology involving the LF gains the same sensitivity of IF and IF-MT, with lower values in ACC, Spe, and AUC.

The sixth part focuses on the ResNet50, where the IF-MT outperforms the unimodal approaches, the EF, and LF techniques, achieving 84.42±0.11%, 85.83±0.11%, 77.78±0.16% in accuracy, specificity, and sensitivity. More- over, the LF exceeds both the EF configurations, obtaining the highest AUC in the section, that is 85.19±0.15

The implemented methodology is also compared with the solution proposed in the literature by Nguyen et al.[19] that represents the first attempt to apply multi-modal learning for ALN metastasis prediction. The authors propose a solution that exploits a CNN, DCE sequence of the primary tumor and four clinical features, namely age, ER, ki-67 and HER2, that in this paper repre- sent the CL-4 set. In [19] a 3D CNN is implemented to process DCE-MRI images using a subtractive approach that works with the third, fourth and fifth post-contrast volumes. A 3D cuboidal bounding box of size 50×50×50, encompassing the tumor region, is used to crop each DCE-MRI data. The CL-4 set is inserted in the first fully connected layer of the classification core of the implemented network. We consider two different configurations, exploiting only the DCE sequence and the DCE volumes with clinical features, respectively. It is possible to note that the implemented methodology out- performs the solution proposed by Nguyen et al.[19] in all metrics, regardless of the specific CNN.

Finally, Table 18 reports a comparison between the CL-S1 set and features selected from the solution involving the SVM and the CL in the unimodal approach. It is possible to note that the features determined by the domain experts are included in the results of the backward features selection.

Set of features	Details
CL-S1	age, familiarity, hormone therapy, menopausal status dimensions, ER, PgR, HER2, ki-67, and grading
SVM features	age, familiarity, hormone therapy, menopausal status dimensions, ER, PgR, HER2, ki-67, grading apparent diffusion coefficient (ADC), signal timing curve margins, tumor class, histotype, multifocality

Table 18: Comparison of the features selected by the domain expert (D.S), and those obtained with the backward features elimination and SVM.

4.4.4 Discussion and Conclusions

In this paper, we propose a Multi Input-Single Out CNN, able to process breast MRI sequences and patients' clinical information for axillary lymph node metastasis prediction, implementing an intermediate fusion approach and enhancing the integration of data coming from heterogeneous sources with the involvement of a transfer module. In particular, the implemented model can perform a cross-modality calibration, with a process aiming to modify the extracted features maps in the convolutional cores, highlighting the most useful characteristics while reducing the effects of the less informative ones. The results reported in Table 18, it is possible to note that, for each network, the solutions exploiting the IF-MT configuration achieve the best performance, outperforming by a wide margin, the experiments involving a single data modality (U). Moreover, the LF performs better than the EF approaches, thus supporting the preference of the former over the latter. Indeed, the LF leverages different models, each trained for a specific data modality. As a consequence, the four classifiers learn to extract features that reflect the distinctive characteristics of each modality, delaying the combination of the results in a post-processing step and also exploiting the uncorrelated nature

of errors performed by models. On the other hand, in both EF solutions, the networks trained in the U approach are used as features extractors, creating a high-dimensionality shared representation, that may make it di cult to capture the complementary characteristics of each data modality. In solutions involving the IF and IF-MT configurations, the shared representation is created by concatenating features coming from the convolutional cores at an intermediate level, thus preserving the distinctiveness of the different image modalities, which is then exploited in the classification core. Moreover, the training strategy exploiting the backpropagation algorithm allows the definition of a representation well suited for the specific task to solve, improving the integration of heterogeneous data compared to the EF and LF experiments. In particular, the presence of the proposed transfer module makes the layers of the convolutional cores influence each other during the features extraction step, enabling the CNN to self-select the characteristics in each modality that best contribute to the problem to be solved. Indeed, the gating mechanism reduces the impact of the selected features according to a shared representation that considers all data sources.

From Table 18 it is also possible to note that, when the ResNet architectures are exploited, the performance decreases as the complexity of the network increases. We argue that this characteristic can be explained by the limited amount of data used in the study. Although ne-tuning is exploited, the small number of MR images may affect the convergence of the involved CNN, limiting its generalization ability.

When the work proposed by [19] is considered, it is possible to note that the solution implementing the IF technique achieves better performance than the unimodal configuration. However, the proposed methodology with the transfer module outperforms by a wide margin the approach described by Nguyen et al. We argue that different aspects contribute to the explanation of this result. Since authors in [43] used a dataset consisting of 357 patients, the performance is reasonably affected by the size of the dataset involved in our work (153 patients). Moreover, the fixed-size bounding box used in [43] reduces the generalization capability of the implemented model as reported in [7]. In this paper, we also exploit different complementary imaging modalities that contribute to the characterization of the primary tumor, leveraging the transfer module to improve the integration of data.

In conclusion, in this study, we performed different experiments with the aim of showing the effectiveness of the proposed transfer module inserted in a CNN implementing the intermediate fusion in a multimodal scenario. Despite the role of magnetic resonance imaging of primary tumors in predicting the involvement of the axillary state has been strongly demonstrated in the literature, it still remains unused in international guidelines [16; 17; 18]. The results presented in this work intend to highlight not only, as previously demonstrated in other works [43; 65], the importance of the role of DL approaches in predicting the state of the axillary lymph node, but also the contribution that pre-contrast sequences can offer. In particular, the T2w sequence allows to highlight the morpho-structural properties of the tumor while the DWI describes the intrinsic properties and tumor aggressiveness. As demonstrated in our previous works, both the peritumor edema evaluated in the T2w sequences and the ADC value, a quantitative expression of the cellular restriction of the tumor lesion, correlate significantly with the state of the axilla [14; 66].

Despite the promising results, we argue that the limitations of our work are represented by the reduced size of the dataset and the involvement of a single medical center. Although the number of patients is similar to that used in [21; 27; 28], future works will focus on the evaluation of the proposed methodology on data collected from different centers, increasing also the size of the considered population.

5 METASTASIS: Artificial Intelligence in Metastases prediction

Only about 5% of breast cancers have metastases at the time of diagnosis. Most diagnoses of metastatic disease occur during the follow-up and treatments for localized disease. The risk of recurrence over time mainly depends on the stage at diagnosis and the molecular subtype. The latter is also usually associated with a different pattern of metastatic sites. After documenting a recurrence of the disease, a complete restaging should be performed.

5.1 Artificial Intelligence in Bone Metastases: an MRI and CT Imaging Review

Bone is the third most frequent site for metastatic localization, after lung and liver[149], with breast and prostate cancer accounting for almost 70% of primary tumors[150]. In most cases, bone metastases influence a patient's short-term prognosis, as bone lesions can rarely be completely eradicated. Patients with bone metastases have the option of undergoing palliative care to reduce the size of the lesions, slow their growth, or allow for improvement in symptoms. Bone metastases lead to a sharp reduction in life expectancy: average survival in patients with bone metastases from melanoma is 6 months; from breast cancer, 19–25 months; and from prostate cancer, 53 months[151].

The improvement of therapeutic strategies to deal with the various forms of cancer has led to an increase in life expectancy and, consequently, a lengthening of the time a patient can coexist with metastatic disease[152]. The most frequent site of bone metastasis is the axial skeleton because of its high red marrow content[149,150,153,154], which is therefore frequently responsible for the increased morbidity and decreased quality of life of patients.

The spectrum of clinical manifestations is very heterogeneous, ranging from complete absence of symptoms to severe pain, reduced mobility, pathologic fractures, spinal cord compression, bone marrow aplasia, and hypercalcemia. Hypercalcemia is in turn responsible for constipation, polyuria, polydipsia, and fatigue[151,155]. In the final stages, hypercalcemia may lead to cardiac arrhythmias and acute renal failure[149].

Therefore, to identify a proper course of treatment, it is essential to differentiate metastatic lesions from any primary or benign lesions of the bone. In order to assess the patient's prognosis and choose the most appropriate medical treatment according to their life expectancy, bone metastases should be diagnosed at the time of the diagnosis of the primary tumor: the aim is to reduce the incidence of complications and improve the quality of life.

Bone scintigraphy, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) are all capable of assessing the presence of bone metastases[156]. The sensitivity and specificity of bone scintigraphy are 78% and 48%,

respectively, but despite its relatively low specificity which may require further imaging examinations, it is still the most widely available technique and the most suggested by the guidelines for the study of bone disease. The CT exam, with a sensitivity and specificity of 74% and 56%, respectively, can be used as a guide during interventional diagnostic procedures. In addition, CT allows simultaneous evaluation of bone and systemic staging, reducing the burden of imaging for patients. MRI shows a sensitivity and specificity of 95% and 90%, respectively. It is a radiation-free technique and is considered the imaging modality of choice for assessing metastatic spread in the bone marrow. 18F FDG-PET (fluorodeoxyglucose) has a sensitivity and specificity of 98% and 56%, respectively: the sensitivity may vary among different histologies, as some well-differentiated tumors can go undetected because of their low metabolism[157].

Radiomics is an emerging branch of artificial intelligence (AI) that involves converting digital medical images that contain information related to tumor pathophysiology, also known as features, into measurable and quantifiable data. These data, combined with clinical and qualitative imaging-derived data, can improve medical decision making[158].

The field of radiomics is constantly and rapidly evolving. The purpose of AI is to aid the physician in the assessment of lesions beyond subjective visual interpretation in order to obtain additional information about tumor behavior and pathophysiology that is otherwise not inferable by the human eye with currently used techniques.

As a topic of relatively recent emergence and application, there is considerable variability in the workflow that determines the results of radiomics-related studies. For traditional radiomics approaches, the workflow is divided into specific steps: data selection, medical imaging evaluation/segmentation, feature extraction, exploratory analysis, and modeling. The acquisition technical specifications and medical image reading modalities, the software and how the segmentation of the regions of interest (ROIs) is produced, the feature extraction, and the algorithm of the predictive model are all subject to numerous factors, making the research, and therefore the literature on it, highly heterogeneous. The radiomics quality score (RQS) was

introduced in order to evaluate the past and future radiomics studies by achieving homogeneity in study reporting[159].

The purpose of this review is to investigate the potential role of radiomics as a decision-supporting tool in predicting bone disease status, differentiating benign from malignant bone lesions, and characterizing malignant lesions at the genetic level.

5.1.1 Methodology

MEDLINE databases, such as PubMed and Web of Science, were employed for the research, using the following strings: ((“radiomics” OR “machine learning”) AND (metastases OR metastasis) AND (“bone” OR “spine” OR “spinal”)).

No limitations were applied to the search strategy. The following criteria were used for the inclusion of the studies: (a) imaging analysis involved only CT and MRI modalities; (b) the studies addressed the ability of radiomics to predict, diagnose, or characterize bone lesions; (c) the studies involved humans only; (d) the articles were accessible through our institution; and (e) the publications were in English.

Case studies, abstracts, reviews, letters to editors, editorials, and commentaries were excluded. We completed the search by manually reviewing the bibliography of all selected articles.

Two reviewers conducted the search, selected the studies, and extracted data from each study independently. From a total of 100 articles, 13 research articles were considered suitable and then collected.

The quality of the methodology was assessed according to the RQS as described by Lambin et al.[159].

Each of the 16 criteria, covering individual aspects of the radiomics workflow, was assigned a different maximum score in relation to its importance. The absence of feature selection and validation results in a reduction in the final score by –3 and –5 points, respectively. The two

reviewers assigned, in agreement, the RQS to the selected studies in absolute and percentage values (maximum value of 36, representing 100%).

The following data were extracted from each study: title, authors, year and journal of publication, study objective, study design (retrospective or prospective), number of patients, CT and MRI technical information, software used for segmentation and feature selection, number and type of radiomics features considered, algorithms used for classification, summary of results, and RQS.

5.1.2 Results

Our search found 13 publications on radiomics as a decision support tool regarding bone lesions. All studies were published between 2018 and 2021 and were retrospective in design. Study characteristics, as recorded by the reviewers, are shown in Table 19.

Authors	Publication Year	Objective	Journal	Number of Patients	Imaging Modality	Segmentation	Technique Used for Feature Selection	Validation	Classification	Features	Best Results	Calibration Statistics	Decision Curve Analysis	RQS
Jiang X et al. [12]	2021	Detect EGFR mutation in spinal metastasis in patients with primary lung adenocarcinoma	*Journal of Magnetic Resonance Imaging*	97	3T MRI T1W, T2W, T2W-FS	Manual, ITK-SNAP	Mann–Whitney U-test, LASSO, 10-fold cross-validation	Y	Logistic regression models	Handcrafted features: first-order, shape- and size- based, texture, filtered features	Fusion features: AUC = 0.771, ACC = 0.550, SEN = 0.750, SPE = 0.875	Y	Y	10/36 = 27.7%
Ren M et al. [13]	2021	Detect EGFR mutation in spinal metastasis in patients with primary lung adenocarcinoma	*Medical Physics*	110	3T MRI T1W, T2W, T2W-FS	Manual, ITK-SNAP	Intraclass correlation coefficient (ICC) analysis, Mann–Whitney U, LASSO, 10-fold cross-validation	Y	Logistic regression, random forest, neural network, and support vector machine	First-order, shape-based, and texture (1967)	Fusion features: AUC = 0.803 (0.682–0.921), SEN = 0.700, SPE = 0.818; nomogram, AUC = 0.882 (0.693–0.976), ACC = 0.808, SEN = 0.866, SPE = 0.846	Y	Y	11/36 = 30.5%
Fan Y et al. [14]	2021	Detect EGFR mutation in spinal metastasis in patients with primary lung adenocarcinoma	*Physics in Medicine & Biology*	94	3T MRI, T1W, T2W-FS	Manual, ITK-SNAP	Mann–Whitney U-test, LASSO, 10-fold cross-validation	Y	Logistic regression models	First-order, shape- and size- based, texture, high-dimensional features (1595)	Multiregional radiomics signature: AUC = 0.777 (0.612–0.967), ACC = 0.688, SEN = 0.615, SPE = 0.947	N	N	8/36 = 22.2%
Ran C et al. [15]	2020	Detect EGFR mutation subtypes in exons 19 and 21 in spinal metastasis in patients with primary lung adenocarcinoma	*Academic Radiology*	76	3T MRI, T1W, T2W-FS	Manual	Mann–Whitney U, LASSO, 10-fold cross-validation	Y	Logistic regression models	First-order, shape-based, and texture (1967)	T1W: AUC = 0.728 (0.526 0.903), ACC = 0.692, SEN = 0.692, SPE = 0.769; T2W-FS AUC = 0.852 (0.736 0.968), ACC = 0.731, SEN = 0.866, SPE = 0.769; nomogram, AUC = 0.821 (0.692–0.929), SEN = 0.667, SPE = 0.909	Y	Y	12/36 = 33.3%
Wang Y et al. [16]	2019	Pretreatment prediction of bone metastasis in patients with prostate cancer	*Magnetic Resonance Imaging*	176	3T MRI T2W, T1W DCE	Manual, IBEX	Linear regression, ridge regression, logistic regression models	Y	Linear regression, ridge regression, logistics regression models	Shape, intensity, intensity histogram, GLCM, gray-level run (976)	Combined T2W and DCE: AUC = 0.898 (0.833–0.938), ACC = 0.821, SEN = 0.667, SPE = 0.782	Y	N	8/36 = 22.2%
Hayakawa T et al. [17]	2020	Investigate the potential prognostic value of clinical risk factors, image features, and radiomics of pelvic bone metastasis in newly diagnosed prostate cancer patients	*Japanese Journal of Radiology*	69	CT	Manual, 3D Slicer	N	N	Not available	Shape-based, first-order statistics, texture (105)	Maximum 2D diameter and least axis were detected as risk factors for OS (HR 1.037 and 1.013, respectively)	N	N	0/36 = 0%
Zhang W et al. [18]	2020	Pretreatment prediction of bone metastasis in patients with prostate cancer	*BMC Medical Imaging*	116	3T MRI, T2W-FS, DWI, DCE T1W	Manual, AK software	ANOVA	Y	Logistic regression models	Not available (206)	AUC = 0.84	Y	Y	14/36 = 38.8%
Sun W et al. [19]	2021	Distinguish between benign and malignant bone tumors	*Cancer Imaging*	206	CT	Manual, ITK-SNAP	LASSO	Y	Logistic regression models	Shape, statistical, texture, wavelet (1130)	Radiomic model, AUC = 0.781 (0.663–0.918); nomogram, AUC = 0.917	Y	Y	12/36 = 33.3%
Xiong X et al. [20]	2021	Differentiating between multiple myeloma and different tumor metastasis lesions of the lumbar vertebra	*Frontiers in Oncology*	107	3T MRI, T1W, T2W-FS	Manual	LASSO, 10-fold cross-validation	Y	Support vector machine, k-nearest neighbor, random forest, artificial neural networks (ANNs), and naïve Bayes	Histogram features, GLCM, GRLM, and an autoregressive model (282)	Differentiating myeloma and metastasis, ANN T2W-FS: AUC = 0.813, SEN = 0.879, SPE = 0.790; differentiating myeloma and metastasis subtypes, ANN T2W-FS: AUC = 0.648, SEN = 0.714, SPE = 0.775	N	N	8/36 = 22.2%
Yin P et al. [21]	2018	Differentiation between primary sacral chordoma, sacral giant cell tumor, and sacral metastatic tumor	*Journal of Magnetic Resonance Imaging*	167	3T MRI, T2W-FS, T1W CE	Manual, ITK-SNAP	ANOVA, LASSO, Pearson correlation, random forest	Y	Random forest	Histogram features, form factor features, Haralick, GLCM features, RLM (385).	Combined T2W and T1W CE: AUC = 0.773, ACC = 0.71; T2W, AUC = 0.678, ACC = 0.541; T1W CE, AUC = 0.592, ACC = 0.568	Y	N	9/36 = 25%
Zhong X et al. [22]	2020	Differentiating of cervical spine osteoradionecrosis from metastasis after radiotherapy in nasopharyngeal carcinoma	*BMC Medical Imaging*	123	1.5 MRI, T1W CE	Manual, MaZda	Intraclass correlation coefficient (ICC) analysis, combination feature selection algorithm (combination of Fisher coefficient, classification error probability combined with average correlation coefficients, and mutual information), LASSO, 10-fold cross-validation	Y	Logistic regression models	Histogram, gray-level co-occurrence matrix, run-length matrix, absolute gradient, autoregressive model, and wavelet (279)	Nomogram: AUC = 0.720 (0.573–0.867), ACC = 0.723, SEN = 0.800, SPE = 0.640	Y	Y	11/36 = 30.5%
Filograna L et al. [23]	2019	Differentiate between metastatic and nonmetastatic vertebral bodies in patients with bone marrow metastatic disease	*La Radiologia Medica*	8	1.5 MRI, T1W, T2W-FS	Not available	Wilcoxon test	N	Logistic regression models	Statistical / histogram, morphological, and textural features (89)	T1W: AUC = 0.831 (0.685–0.942); T2W: AUC = 0.911 (0.829–0.993)	N	N	2/36 = 5.5%
Lang N et al. [24]	2019	Differentiate metastatic cancer in the spine originated from lung cancer and other nonlung tumors	*Magnetic Resonance Imaging*	61	3T MRI DCE	Manual, automatic	Random forest algorithm	N	Logistic regression models	Texture, histogram (33 × 3 maps)	3 features, histogram + texture: ACC = 0.68; 5 features, histogram + texture ACC = 0.71;	N	N	1/36 = 2.7%

Table 19: Characteristics of the selected radiomics studies.

Eleven (85%) studies were MRI-based, whereas two (15%) were CT-based. Four (30%) studies were focused on whether radiomics could predict epidermal growth factor receptor (EGFR) mutation in spinal metastases of primary lung adenocarcinoma. Three (23%) studied bone metastases from prostate cancer: two aimed to predict the presence of bone metastases from

prostate cancer, one studied the prognostic role in terms of overall survival (OS) and cause-specific survival (CSS) of radiomics in prostate cancer patients with bone metastases. Four (30%) studies aimed to differentiate bone metastases from other pathological conditions: two studies evaluated the ability of radiomics to differentiate bone metastases from benign vertebral bone disease, and two studies evaluated the ability of radiomics to differentiate bone metastases from other pathological bone lesions. One (7%) study aimed to differentiate between metastatic and non-metastatic vertebral bodies, and one aimed to differentiate between metastatic lesions in the spine originating from lung cancer and other non-pulmonary cancers.

5.1.2.1 EGFR Mutation Prediction in Spinal Metastasis from Primary Lung Adenocarcinoma

Jiang et al.[160] analyzed MRI-based multiparametric radiomics for EGFR mutation prediction on T2-weighted (T2W), T2-weighted fat-saturated (T2W-FS), and T1-weighted (T1W) images: both traditional handcrafted and deep learning-based features were derived from each MRI sequence. For each of the two types of approach, radiomics models showed better results using combined features from all the MRI sequences than those with features extracted from each individual sequence. A fusion model created by integrating traditional handcrafted and deep learning-based features from the three sequences achieved the best prediction performance. A radiomics nomogram was obtained by integrating the best performing radiomics features: a decision curve analysis (DCA) confirmed the potential clinical utility of the radiomics nomogram.

Ren et al.[161] produced a nomogram using an MRI-based radiomics signature and smoking status to classify patients with EGFR mutation and wild-type EGFR through analysis of spinal metastases on T2W, T2W-FS, and T1W images. In addition to the radiomics model, a deep learning approach was considered: the combined signature generated higher AUCs than either feature type alone. Four different machine learning classifiers were developed and compared, with logistic regression outperforming the others. The nomogram achieved an AUC of 0.821

(SEN = 0.667, SPE = 0.909): DCA showed that the nomogram had a higher net benefit than all treatment and nontreatment strategies when the threshold was greater than 0.013.

Fan et al.[162] proposed a predictive model that could determine the presence of EGFR mutation in spinal metastasis subregions. Spinal metastases were divided into subregions based on patient- and population-level clustering: marginal, fragmentary, and internal subregions and the total tumor region. Radiomics features were extracted from the subregions' T2W-FS and T1W images. For both sequences, the radiomics signature derived from the inner subregions outperformed other subregions or the entire tumor regions in terms of AUC: the multiregion radiomics signature derived from merging the inner subregion from T1W and T2W-FS MRI achieved the best detection capabilities. The results suggest that the inner region is biologically more aggressive than the others.

Ran et al.[163] further investigated the predictive ability of the EGFR mutation in spinal metastases by constructing a radiomics model that could identify the mutation subtype in exon 19 and exon 21. The radiomics signature derived from the T2W-FS MRI consistently outperformed the T1W-derived signature in terms of AUC, ACC, sensitivity, and specificity. A nomogram model was constructed by incorporating the combined radiomic signature, age, and CEA level, achieving an AUC of 0.881 in the validation set: a decision curve analysis (DCA) confirmed that the model potentially guides individual treatments for patients with lung adenocarcinoma.

5.1.2.2 Bone Metastasis from Prostate Cancer

Wang et al.[164] determined that multiparametric prostate MRI predicted the presence/absence of bone metastasis in prostate cancer patients using radiomics features alone and combined with free PSA level and Gleason score. The combined MRI features derived from T2W and DCE showed higher prognostic performance than features derived from the single sequence and Gleason score. The radiomics MRI model combined with clinicopathological features (free

PSA level, age, and Gleason score) yielded the highest AUC (AUC = 0.916), further improving predictive performance.

Hayakawa et al.[165] investigated the potential prognostic value of clinical risk factors (anamnestic and laboratory data and histological prostate cancer characteristics), imaging features, and radiomics of pelvic bone metastases in patients with newly diagnosed prostate cancer: patients were studied for OS and CSS. Only shape-based features were detected as risk factors for OS, and "maximum 2D diameter", defined as the largest tumor surface dimension in the axial plane, was detected as a risk factor for OS after multivariate analysis (HR = 1.007). None of the radiomics features were detected as a risk factor for CSS in the uni- and multivariate analysis. After multivariate analysis, LDH, hemoglobin, and "maximum 2D diameter" were detected as risk factors for OS, whereas total Gleason score, LDH, and maximum 2D diameter were detected as a risk factors for CSS. Radiation therapy to the prostate gland and bone metastases did not significantly improve both OS and CSS.

Zhang et al.[166] established and validated a radiomics model that combined prostate multiparametric MRI-based radiomics signature and clinical risk factors to predict bone metastasis in patients with prostate cancer before treatment. The radiomics signature constructed from features extracted from DWI, T2W-FS, and DCE images showed good predictive efficiency. The nomogram, which incorporated the radiomics signature based on MRI and clinical risk factors, had an AUC of 0.92 in the validation set. DCA also demonstrated the clinical use of the radiomics model, which had better discriminatory efficiency than t-PSA or radiomics signature alone.

5.1.2.3 Differentiation of Bone Metastases from Other Bone Diseases

Sun et al.[167] proposed a CT-based nomogram able to distinguish between benign and malignant bone tumors. The nomogram, obtained by combining the radiomics signature and clinical model (consisting of demographics and CT characteristics), had higher diagnostic performance than the clinical model, but there was no statistical difference compared with the radiomics

signature (AUC = 0.823 in the validation set). The DCA showed that the nomogram had higher diagnostic performance than the clinical model and achieved greater net clinical benefits than the clinical and radiomics signature models when considered alone.

Xiong et al.[168] evaluated the discrimination ability in T1W and T2W-FS MRI sequences between bone lesions from multiple myeloma and metastasis through several machine learning models: support vector machine (SVM), k-nearest neighbor (KNN), random forest (RF), artificial neural networks (ANNs), and naïve Bayes (NB). The ANN classifier from T2W images showed the best performance, both in differentiating myeloma from metastases and for classifying metastasis subtypes.

Yin et al.[169] developed and validated a multiparametric prostate MRI-based radiomics model to differentiate primary sacral chordoma, giant cell sacral tumor, and metastatic sacral tumor. Radiomics features extracted from the combined T2W-FS and CE T1W images exceeded those from the T2W-FS or T1W images alone, but T2W-FS outperformed T1W images. The highest radiomics model AUC was achieved when clinical and imaging data were combined.

Zhong et al.[170] proposed an MRI-based radiomics nomogram to differentiate cervical spine osteoradionecrosis from metastasis in patients with nasopharyngeal carcinoma after radiotherapy. The nomogram model demonstrated good calibration and discrimination, and DCA indicated that, if the threshold probability of a lesion for diagnosis as osteoradionecrosis is >12%, the radiomics nomogram adds net benefit when compared to either the treat-all-patients scheme or the treat-none scheme.

5.1.2.4 Other Studies

The study of Filograna et al.[171] is the only study that demonstrated the ability of radiomics-based MRI to differentiate between metastatic and non-metastatic vertebral bodies in non-radiotherapy-treated cancer patients with metastatic bone marrow disease from primary tumors of different nature (three lung cancers, one prostate cancer, one esophageal cancer, one

nasopharyngeal cancer, one hepatocarcinoma, and one breast cancer). Internal cross-validation showed an AUC of 0.814 for T1W images and 0.911 for T2W images. One histogram feature (minimum gray level) and one textural feature (joint variance of the gray level co-occurrence matrix) were found to be the best-fitting features in T1W and T2W images, respectively.

Lang et al.[172] aimed to differentiate metastatic spine cancer derived from primary lung cancer and other no pulmonary cancers (breast, thyroid, prostate, liver, kidney) using an ROI-based model, radiomics, and deep learning. The accuracy of the radiomics model when histogram and texture features were combined was higher than that when histogram and texture features were evaluated alone. By increasing the number of features from three to five, the accuracy showed slightly higher values (from 0.68 to 0.71 in the histogram + texture model). The accuracy of the radiomics model was worse than that of the hot-spot ROI-based (ACC = 0.79) and deep learning (ACC = 0.71 ± 0.043 and 0.81 ± 0.034) methods.

The application of radiomics in the diagnosis and characterization of bone lesions is recent and constantly evolving, as is the entire field of radiomics. The articles identified by our two researchers are few and were all published within the period between 2018 and 2021, with approximately 70% in the period immediately after 2020. Reflecting the relative freshness of this area of research, all studies are retrospective, performed at a single center, and with a small study population, ranging from 8 to 176 patients.

Radiomics can not only predict the presence of bone metastases and differentiate skeletal regions without lesions from those containing metastases, but its application is able to determine the primary tumor, differentiate metastases from other bone lesions (both benign and malignant), and predict mutation status (such as EGFR). Apart from MRI and FDG-PET, which have high predictive values, the other imaging methods have relatively low sensitivity and specificity values, although they are easily accessible and widespread[156,157]. Despite the predictive capabilities of the traditional imaging methods, there is some clinical information regarding bone metastases, including the genetic status or the primary tumor, that the naked eye is not able to perceive, due to similar clinical and imaging manifestations. Complete pathological confirmation and histological analysis are currently only possible by sampling

through bone biopsy, which is associated with relatively high procedural risks (such as vertebral artery or spinal cord damage)[173]. Radiomics models, by inferring quantifiable data from the features, allow obtaining information that, once applied in the clinical setting, can be decisive for the specific therapeutic treatment choice. Because data are extracted from noninvasive methods, and in most cases radiation-free methods, radiomics is a further step towards the reduction in a patient diagnostic burden, and at the same time towards a patient-centered medicine. Some studies have also constructed nomograms in order to graphically represent the mathematical relationship between radiomics features and other prognostic factors, both clinical and diagnostic, in order to improve the clinical applicability of a field still difficult for nonexperts to interpret.

All articles included among their limitations the relatively small sample size (<200 patients), the single-center nature of the study, and the selection bias introduced by the retrospective design. Even in studies in which validation was performed on an internal dataset, the absence of external validation leads to reduced evidence of the possible clinical application of the research: multicenter studies are necessary to validate radiomics models and nomograms. Some articles complained about the tediousness of manual segmentation, which, in addition to being time-consuming, is not free of human error despite the option of multiple segmentation: the hope is that the spread of automatic, or semiautomatic, segmentation will speed up the process and further reduce the margin of error.

Our review confirms the considerable heterogeneity in current radiomics research, as evidenced by the relatively low RQS value obtained when analyzing the reviewed studies (22.6%). There is not a homogeneous protocol used for MRI sequences among the different studies, although the highest predictive ability was always obtained in T2W-FS. Wide variability also exists in the software used for image segmentation and feature extraction; the number and the type of features explored, with and without feature selection method application; and even the models used to classify the final features. All of these elements contribute to the reduced reproducibility of the results, even if none of them are considered integral to the RQS assessment.

As described above, the most critical limitation concerns the small sample size, which leads to selection bias. A possible way to overcome this important limit is to increase the number of patients under investigation or to extend the research and results validation to other centers. In fact, it is well known that, after a first validation of a radiomic model, a subsequent path of validation through multicenter studies is necessary to allow radiomics to get closer and closer to widespread clinical applicability, even and especially through prospective studies.

This review has some important limitations. To our knowledge, no other review has exclusively investigated the role of radiomics in the analysis and prediction of bone metastases, particularly the spine localization. Even within the field of radiomics, this is a niche subfield, as is evident from the low number of studies analyzed. This novelty, in addition to the high variability of the included studies, both in methodology and in objectives, prevented us from pursuing a robust meta-analysis. We expect that as radiomics evolves and becomes more widespread, there will be an increase in the number of patients included and more extensive validation of existing datasets. Another critical issue at this early stage of research is the ability to share data across public datasets that have already been validated, as currently none of the papers publicly released their data.

In addition, we have deliberately eliminated from the research the studies based on scintigraphy and PET (we have not detected studies that have used ultrasound) and papers in non-English language or not accessible from our institution, reducing the number of the articles included. Due to an implicit publication bias, most articles on this topic focus on the use of MRI. This implies that many other methods, on which there are no current studies, do not result in a significant contribution to research in the radiomics field, a phenomenon that introduces further bias into our review. Furthermore, at the time of publication, it is safe to assume that there are additional feature extraction software and classification models currently in development that we are unaware of in the literature, which are therefore protected from our review.

5.1.2.5 RQS Assessment and Study Limitations

The average recorded RQS was 22.6% (0–38.8%). This low score confirms what has been reported in other reviews in the field of radiomics, representing a relatively low quality of research methodology[7,74,173–177]. None of the reviewed studies were prospective in design, no external validation on a dataset from another institution was performed, no cost-effectiveness of the clinical application of the radiomic models was reported, and no datasets were made publicly available (although four authors allowed access to the datasets upon request). Two articles (15%) did not perform any validation of their results. In only four (30%) of the articles, multiple segmentations were performed to assess the robustness of features to segmentation variabilities. The majority of articles (12/13, 92%) performed a feature reduction to decrease the risk of overfitting. Eight (61%) studies reported discrimination statistics (such as ROC curve and/or AUC), and six (46%) studies reported calibration statistics. Six articles (46%) reported on the potential application of the model in a clinical setting with a DCA.

5.1.3 Discussion and Conclusions

The application of radiomics in the diagnosis and characterization of bone lesions is recent and constantly evolving, as is the entire field of radiomics. The articles identified by our two researchers are few and were all published within the period between 2018 and 2021, with approximately 70% in the period immediately after 2020. Reflecting the relative freshness of this area of research, all studies are retrospective, performed at a single center, and with a small study population, ranging from 8 to 176 patients.

Radiomics can not only predict the presence of bone metastases and differentiate skeletal regions without lesions from those containing metastases, but its application is able to determine the primary tumor, differentiate metastases from other bone lesions (both benign and malignant), and predict mutation status (such as EGFR). Apart from MRI and FDG-PET, which have high predictive values, the other imaging methods have relatively low sensitivity and specificity values, although they are easily accessible and widespread[156,157]. Despite the

predictive capabilities of the traditional imaging methods, there is some clinical information regarding bone metastases, including the genetic status or the primary tumor, that the naked eye is not able to perceive, due to similar clinical and imaging manifestations. Complete pathological confirmation and histological analysis are currently only possible by sampling through bone biopsy, which is associated with relatively high procedural risks (such as vertebral artery or spinal cord damage)[173]. Radiomics models, by inferring quantifiable data from the features, allow obtaining information that, once applied in the clinical setting, can be decisive for the specific therapeutic treatment choice. Because data are extracted from noninvasive methods, and in most cases radiation-free methods, radiomics is a further step towards the reduction in a patient diagnostic burden, and at the same time towards a patient-centered medicine. Some studies have also constructed nomograms in order to graphically represent the mathematical relationship between radiomics features and other prognostic factors, both clinical and diagnostic, in order to improve the clinical applicability of a field still difficult for nonexperts to interpret.

All articles included among their limitations the relatively small sample size (<200 patients), the single-center nature of the study, and the selection bias introduced by the retrospective design. Even in studies in which validation was performed on an internal dataset, the absence of external validation leads to reduced evidence of the possible clinical application of the research: multicenter studies are necessary to validate radiomics models and nomograms. Some articles complained about the tediousness of manual segmentation, which, in addition to being time-consuming, is not free of human error despite the option of multiple segmentation: the hope is that the spread of automatic, or semiautomatic, segmentation will speed up the process and further reduce the margin of error.

Our review confirms the considerable heterogeneity in current radiomics research, as evidenced by the relatively low RQS value obtained when analyzing the reviewed studies (22.6%). There is not a homogeneous protocol used for MRI sequences among the different studies, although the highest predictive ability was always obtained in T2W-FS. Wide variability also exists in the software used for image segmentation and feature extraction; the number and the type of features explored, with and without feature selection method

application; and even the models used to classify the final features. All these elements contribute to the reduced reproducibility of the results, even if none of them are considered integral to the RQS assessment.

As described above, the most critical limitation concerns the small sample size, which leads to selection bias. A possible way to overcome this important limit is to increase the number of patients under investigation or to extend the research and results validation to other centers. In fact, it is well known that, after a first validation of a radiomic model, a subsequent path of validation through multicenter studies is necessary to allow radiomics to get closer and closer to widespread clinical applicability, even and especially through prospective studies.

This review has some important limitations. To our knowledge, no other review has exclusively investigated the role of radiomics in the analysis and prediction of bone metastases, particularly the spine localization. Even within the field of radiomics, this is a niche subfield, as is evident from the low number of studies analyzed. This novelty, in addition to the high variability of the included studies, both in methodology and in objectives, prevented us from pursuing a robust meta-analysis. We expect that as radiomics evolves and becomes more widespread, there will be an increase in the number of patients included and more extensive validation of existing datasets. Another critical issue at this early stage of research is the ability to share data across public datasets that have already been validated, as currently none of the papers publicly released their data.

In addition, we have deliberately eliminated from the research the studies based on scintigraphy and PET (we have not detected studies that have used ultrasound) and papers in non-English language or not accessible from our institution, reducing the number of the articles included. Due to an implicit publication bias, most articles on this topic focus on the use of MRI. This implies that many other methods, on which there are no current studies, do not result in a significant contribution to research in the radiomics field, a phenomenon that introduces further bias into our review. Furthermore, at the time of publication, it is safe to assume that there are additional feature extraction software and classification models currently in development that we are unaware of in the literature, which are therefore protected from our review.

In spite of the variability in the radiomics method application, the similarity of results and conclusions observed is encouraging. Furthermore, all six studies that have measured the possible application of the radiomics model in the clinical setting through DCA have shown a net benefit compared to the use of the other strategies alone, confirming the promising role of radiomics in guiding the choice of treatments for individual cancer patients.

5.2 The DL approach applied to BC DCE images to predict distant metastasis status

Building a ML system requires engineering and expertise to design an extractor that transforms the features into an appropriate representation that the learning subsystem can detect or classify. For this reason, there has been an emergence of the Deep Learning (DL) approach, which improves this process by automatically discovering the representations needed for the specific task from the raw data{Formatting Citation}. A key role is played by Convolutional Neural Networks (CNNs), which are a set of deep architectures that receive raw image data as input and extract features to learn discriminative properties using hierarchical representations of images.

Deep Learning has already been applied to BC-MRI in the lesion diagnosis, prediction of tumor molecular subtypes, pathological complete response, and axillary lymph node status [13,14,15,16,17]. To our knowledge, there are no articles that have studied the role of AI through DL in predicting the risk of distant metastasis from primary BC. Therefore, the purpose of this study is to develop a DL-CNN model to predict the risk of distant metastasis using Contrast-Enhanced 3T-MRI.

5.2.1 Population

This is a retrospective observational study where only existing information collected from human participants was used, with no identifiers linking individuals to the data/samples. All methods and procedures meet institutional and research committee ethical standards in accordance with the 2013 Declaration of Helsinki.

All breast 3T-MRI exams performed at the Radiology Department of our hospital, from January 2011 to July 2022 for pre-operative evaluation, were retrospectively reviewed. The inclusion criteria were as follows: (a) pre-operative breast 3T-MRI with Dynamic DCE-MRI sequences; (b) diagnostic confirmation of invasive breast cancer by histopathological analysis;

(c) complete histological analysis including molecular receptor status (estrogen receptor ER; progesterone receptor PgR; epidermal growth factor receptor HER2) and proliferation index Ki-67; (d) patients undergoing clinical and imaging follow-up at our institution for at least seven years or until distant metastases diagnosis.

The presence of breast implants, post-chemotherapy follow-up patients, exams for neo-adjuvant treatment evaluation, and images that were not of good diagnostic quality were all considered as exclusion criteria.

Written informed consent was obtained for all patients before MRI.

Following the mentioned criteria, a total of 157 breast cancer patients with 160 breast lesions were included in the study. Patients clinical data (age, menopausal state, family history, hormone therapy) were collected.

5.2.1.1 MRI Examination

All MRI exams were performed on a 3T magnet (Discovery 750; GE Healthcare, Milwaukee, WI, USA). Patients were positioned prone and a dedicated eight-channel breast coil (8US TORSOPA) was used. Three orthogonal localizer sequences were employed, then images were acquired following this protocol:

- T2-weighted axial single-shot fast spin echo sequence with a modified Dixon technique (IDEAL) for intravoxel fat-water separation (TR/TE 3500–5200/120–135 ms, slice thickness 3.5 mm).
- Diffusion-weighted axial single-shot echo-planar with fat suppression sequence.
- (TR/TE 2700/58 ms, slice thickness 5 mm) with diffusion-sensitizing gradient with a b-value of 0, 500, and 1000 s/mm2.
- Dynamic 3D-T1w axial and sagittal gradient echo sequence with fat suppression after injection of 0.1 mmol/kg body weight of Gadoteric acid (Dotarem®, Guerbet S.p.A., Villepinte France, or Claricyclic®, GE Healthcare S.r.l, Chicago,

IL, USA) at a rate of 2 mL/sec followed by a bolus of 15 mL saline flush (TR/TE 4/2 ms, slice thickness 2.4 mm), before, and five to ten times after intravenous contrast medium injection.

Subtracted images were automatically produced in post-processing from the images after contrast medium administration for a more accurate tumor analysis. The largest lesion was considered to be the index lesion for statistical analysis purposes. Reading was performed without access to the original reports and clinical data.

The following MRI characteristics were collected for each lesion using DCE sequences as reference images for tumor detection and characterization.

- Location on the breast quadrant;
- Margins: regular, irregular, lobulated, speculated, non-mass;
- Size (mm);
- Morphology: round, oval, or irregular.

For each index lesion, a signal-intensity-to-time curve (SI/T) was automatically generated by placing a region of interest (ROI) within the lesion on a subjectively recognized area of maximal contrast enhancement and evaluating all the acquired DCE series. The kinetics curves were classified as I (progressive wash-in), II (plateau) or III (rapid wash-out), as reported in the current ACR BI-RADS guidelines [18].

ADC values were calculated for quantitative analysis by superimposing the subtracted images on the ADC map. An ROI with a diameter of 3–6 mm was manually drawn on the slice where the lesion reached its greatest diameter. ADC measurements were performed only on the enhanced solid portion to avoid areas of T2 shine-through.

Distant metastasis status for each patient was recorded using clinical and imaging follow-up, with definitive histological characterization confirming breast cancer cells as a dichotomous result: positive, if there was at least one organ with metastases, or negative, if there were no metastases. Radiological examinations (MRI, CT, and PET) and medical records of the Patients

were collected up to a maximum of seven years after the initial diagnosis of breast cancer; if distant metastases were found, their organ site was recorded.

5.2.1.1 Histologic Characteristics

All breast lesions were characterized on the histological specimen obtained by a core biopsy and on the histological definitive sample after surgery by two pathologists. Histological diagnosis was performed according to the WHO classification.

The histopathological grade was evaluated according to the Nottingham Grading System, considering tubule formation, pleomorphism, and mitotic count through a scoring system. The total score ranges from 3 to 9: 3–5 corresponds to grade 1 (G1), 6 or 7 to grade 2 (G2), and 8 or 9 to grade 3 (G3).

Immunohistochemical (IHC) analysis was performed to evaluate molecular receptor status (ER, PgR, and HER2) and to calculate the Ki-67 index. Evaluation of ER and PgR status was performed by IHC using Dako monoclonal antibody, 1:100 dilution. The monoclonal antibody Mib-1 (1:200 dilution; Dako, Glostrup, Denmark) was used to assess the Ki-67 index, which was reported as the percentage of immune-reactive cells out of 2000 tumor cells in randomly selected high-power fields surrounding the tumor core. HER2 status was re-evaluated using the Hercep test (Dako, Glostrup, Denmark), following published guidelines. Samples that gave an equivocal IHC result were subjected to fluorescence in situ hybridization (FISH) analysis. A ratio of HER2 gene signals to chromosome 17 signals greater than 2.2 was used as a cut-off value to define HER2 gene amplification. ER and PgR status were considered to be positive if the expression was >1% and negative if the expression was <1%. HER2 expression was classified as 0, 1+, 2+ or 3+; only tumors reaching a score of 3+ were considered to be HER2-positive.

5.2.2 Methodology

5.2.2.1 Pre-Processing and segmentation

All breast lesions were characterized on the histological specimen obtained by a core biopsy and on the histological definitive sample after surgery by two pathologists. Histological diagnosis was performed according to the WHO classification.

The histopathological grade was evaluated according to the Nottingham Grading System, considering tubule formation, pleomorphism, and mitotic count through a scoring system. The total score ranges from 3 to 9: 3–5 corresponds to grade 1 (G1), 6 or 7 to grade 2 (G2), and 8 or 9 to grade 3 (G3).

Immunohistochemical (IHC) analysis was performed to evaluate molecular receptor status (ER, PgR, and HER2) and to calculate the Ki-67 index. Evaluation of ER and PgR status was performed by IHC using Dako monoclonal antibody, 1:100 dilution. The monoclonal antibody Mib-1 (1:200 dilution; Dako, Glostrup, Denmark) was used to assess the Ki-67 index, which was reported as the percentage of immune-reactive cells out of 2000 tumor cells in randomly selected high-power fields surrounding the tumor core. HER2 status was re-evaluated using the Hercep test (Dako, Glostrup, Denmark), following published guidelines. Samples that gave an equivocal IHC result were subjected to fluorescence in situ hybridization (FISH) analysis. A ratio of HER2 gene signals to chromosome 17 signals greater than 2.2 was used as a cut-off value to define HER2 gene amplification. ER and PgR status were considered to be positive if the expression was >1% and negative if the expression was <1%. HER2 expression was classified as 0, 1+, 2+ or 3+; only tumors reaching a score of 3+ were considered to be HER2-positive.

5.2.2.1 Volumes Extraction

To limit the amount of non-tumor tissue to be included, the smallest 3D rectangle circumscribed by the tumor region was considered for each patient in Single Variable-size Box (SVB), as shown in Figure 27

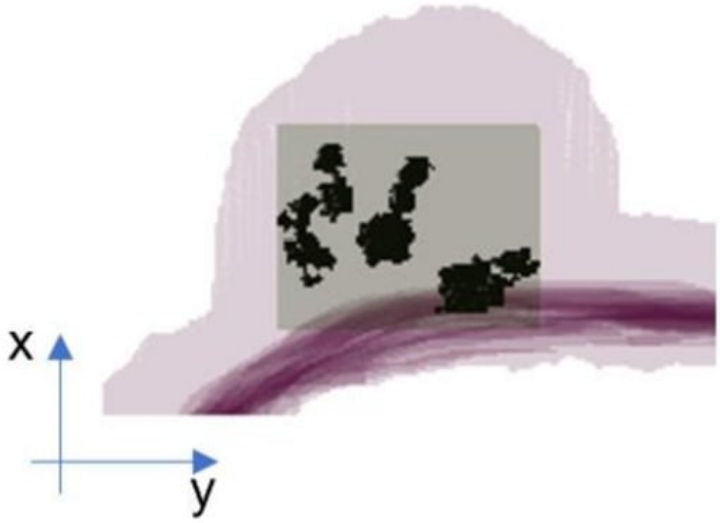

Figure 28: Representation of the SVB option, only the smallest 3D bounding box confined to the tumor region is considered.

Denoting with dxp × dyp × dzp the size of the bounding rectangle of patient p, a 3D cubic box of size dmp × dmp × dmp is used, where dmp = max(dxp, dyp, dzp), to crop the subject. The cubic box of the SVB option depends on the patient. Therefore, the amount of non-tumor tissue depends on the shape of each patient's tumor region and the difference between the largest (dmp) and maximum (dxp, dyp, dzp) size, and the difference between the largest (dmp) and smallest (dmp) size.

However, in the case of multifocal and multicentric tumors, the parenchyma between lesions was included in the 4D volume extracted. Each voxel was associated with its size information in millimeters (mm). During the acquisition of the DCE sequence for patient p, the Pixel Spacing attribute was the physical distance between the centers of each two-dimensional pixel, specified by two numerical values (ˆxp, ˆyp) that represented the row spacing and vertical spacing, respectively. In addition, the attribute spacing between slices specified by the

numerical value ˆzp was the distance between slices measured along the normal to the first image.

5.2.2.1 Metastasis Prevision Assessment

We exploited CNNs for the prediction of extra-nodal metastasis. The voxel-based (VB)-NET model was used, considering that the size of the selection rectangle varies according to the tumor region of each patient.

The proposed networks consist of several reduction layers and two fully connected layers. A reduction layer is a block with a convolutional layer, followed by a normalization layer and the ReLU function. Each convolutional layer reduces the dimensionality of the input feature map and doubles the number of channels, making pooling operations (maximum or average pooling) unnecessary.

The architecture of VB-NET is shown figure 28. We used a 3D CNN, consisting of five reduction blocks. The input volume (4D data) represented the smallest cubic box surrounding the tumor region or lesion of each patient. Because the SVB considered a selection rectangle size that varies according to each patient's tumor region, resizing was used to give the volumes a common size of 64 × 64 × 64, before feeding them to the VB-NET.

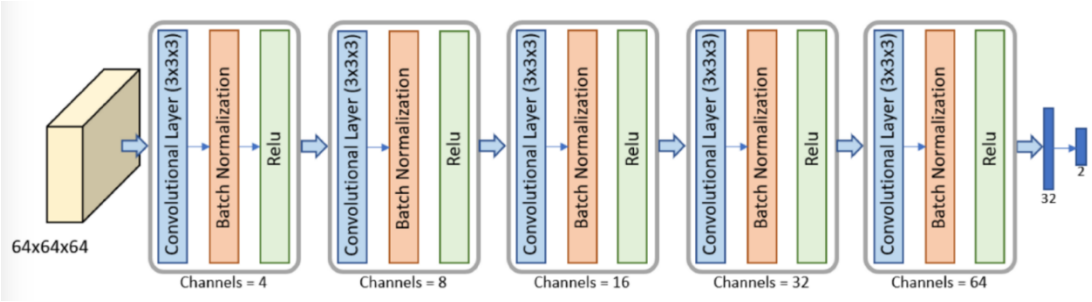

Figure 29: Architecture of the VB-NET used for the SVB. The VB-NET is a 3D CNN with five reduction blocks and two fully connected layers.

Performance was evaluated in terms of accuracy, sensitivity, specificity, and area under the ROC curve (AUC). The experiment was performed in a 10-fold cross-validation (CV) setting. A patient-based cross-validation was performed to reliably estimate performance, avoiding the use of 3D volumes or slices of the same patient during the training and evaluation phase. The experiment was conducted using Pytorch (version 1.10, Meta AI, Astor Place, New York, NY, USA), while the pre-processing phase, including the different bounding box options, were implemented in MATLAB 2020b.

5.2.2.1 Statistical Analysis

Descriptive statistics were carried out using the statistical software SPSS© version 25.0. Statistical significance was set at $p < 0.05$. Spearman's rank-order correlation was evaluated to assess whether there was a correlation between the presence of distant metastases during follow-up and categorical variables, clinical (menopausal status, hormone therapy, family history), MRI (site, margins, size, morphology, kinetic curves), and histologic features (histologic type, grading, expression of ER, PgR and HER2). The Kolmogorov–Smirnov test was performed to determine whether age, tumor size, Ki-67 index, and ADC values followed a normal distribution.

Statistical comparisons between the presence of distant metastases during follow-up and age, tumor size, Ki-67 index, and ADC values were performed using the Kruskal–Wallis H test.

5.2.3 Results

In this study, 157 breast cancer patients with 160 lesions were included. Three patients had bilateral breast cancer, but only one patient with bilateral BC presented distant metastases at follow-up. The mean age of the patients was 55 years (range 30–85 years). The patients were divided into two groups:

- Patients with distant metastases at follow-up (39/157 patients, 40 lesions);
- Patients negative for distant metastasis (control group, 118/157 patients, 120 lesions).

Metastasis sites are represented as follows: 3 (5%) brain, 17 (31%) lung, 19 (34%) bone, 14 (25%) liver, 1 (2%) gluteal subcutis, and 1 (2%) brachial plexus.

Anamnestic and clinical data of the study population were collected. The mean age of patients with metastases at follow-up was 54.3 years (range 30–84 years), while the mean age of the control group was 55.2 years (range 30–85 years). The average time to occurrence of distant metastasis was 22.8 months (range: 1–84 months).

The mean diameters of the measured lesions in the study population, the group with distant metastases, and the control group were 22.6 mm (range 6–90 mm), 33.3 mm (range 7–90 mm), 19.7 mm (range 6–80 mm), respectively.

The main imaging, histological, and clinical characteristics of the patients are shown in Table 20, 21 and 22. Two examples are reported in Figure 29 and 30.

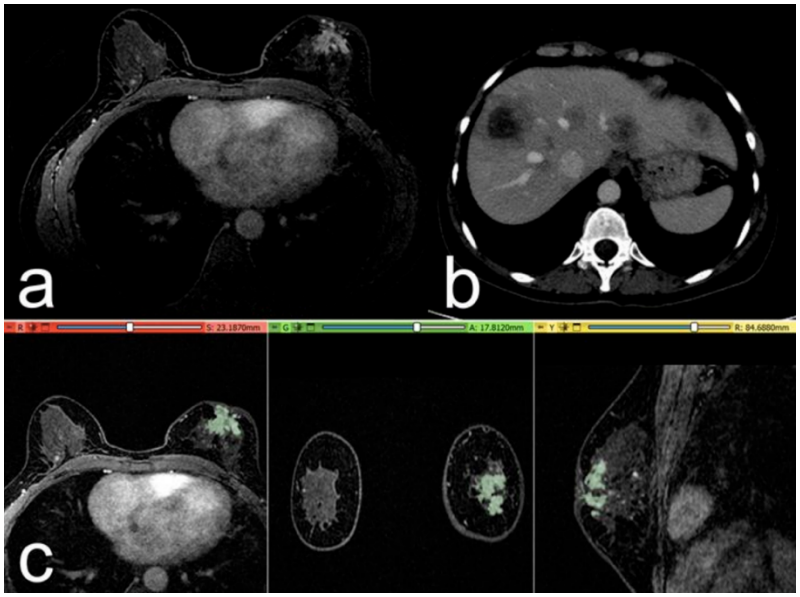

Figure 30: Case 1
Case of a 52-year-old woman with a G3 triple negative invasive ductal carcinoma, Ki-67 of 30%. (a) The post-contrast image shows a retroareolar irregular non-mass-enhancing lesion in the left breast. (b) At 1-year follow-up: post-contrast axial CT image, with some metastatic nodules in the liver parenchyma. (c) Representation of the extraction of the segmentation mask.

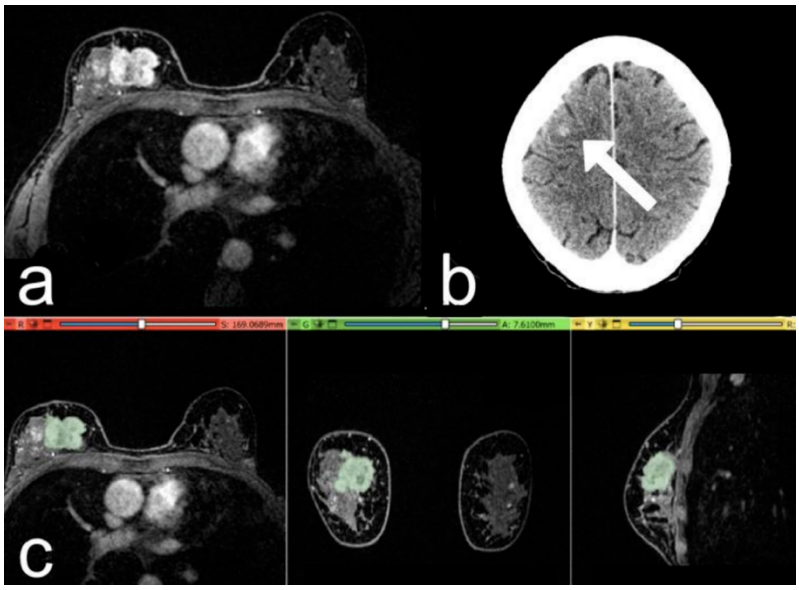

Figure 31: Case 2
Case of a 53-year-old woman with a G2 Luminal B invasive ductal carcinoma, Ki-67 of 25%. (a) The post-contrast image shows an irregular mass-enhancing lesion in the upper quadrants of the right breast. (b) At 3-year follow-up: post-contrast axial CT image, with a metastatic nodule in the right frontal lobe (arrow). (c) Representation of the extraction of the segmentation mask.

Variation			Study Population	Patients with Metastasis	Control Group	*p*-Value
Kinetic Curve	I	*n*	21	7	14	
		%	13.1%	17.5%	11.7%	
	II	*n*	71	15	56	0.962
		%	44.4%	37.5%	46.7%	
	III	*n*	68	18	50	
		%	42.5%	45.0%	41.7%	
Margins	Regular	*n*	4	0	4	
		%	2.5%	0.0%	3.3%	
	Irregular	*n*	86	21	65	
		%	53.8%	52.5%	54.2%	
	Lobulated	*n*	25	7	18	0.349
		%	15.6%	17.5%	15.0%	
	Spiculated	*n*	33	6	27	
		%	20.6%	15.0%	22.5%	
	Non-mass	*n*	12	6	6	
		%	7.5%	15.0%	5.0%	

Table 20: Description of the extracted MRI characteristics.

Variation			Study Population	Patients with Metastasis	Control Group	*p*-Value
Histology	IDC	*n*	127	30	97	
		%	79.4%	75.0%	80.8%	
	ILC	*n*	33	10	23	0.433
		%	20.6%	25.0%	19.2%	
Molecular subtype	Luminal A	*n*	59	11	48	
		%	36.9%	27.5%	40.0%	
	Luminal B	*n*	69	18	51	
		%	43.1%	45.0%	42.5%	
	HER2+	*n*	13	3	10	0.079
		%	8.1%	7.5%	8.3%	
	Triple negative	*n*	19	8	11	
			11.9%	20.0%	9.2%	
ER Status	Negative	*n*	31	12	19	
		%	19.4%	30.0%	15.8%	
	Positive	*n*	129	28	101	0.195
		%	80.6%	70.0%	84.2%	
PgR Status	Negative	*n*	58	23	35	
		%	36.3%	57.5%	29.2%	
	Positive	*n*	102	17	85	0.001 *
		%	63.7%	42.5%	70.8%	
HER2 Status	Negative	*n*	137	32	105	
		%	85.6%	80.0%	87.5%	
	Positive	*n*	23	8	15	0.044 *
			14.4%	20.0%	12.5%	
Grade	1	*n*	19	4	15	
		%	11.9%	10.0%	12.5%	
	2	*n*	71	15	56	0.225
		%	44.4%	37.5%	46.7%	
	3	*n*	70	21	49	
		%	43.8%	52.5%	40.8%	

Table 21: Description of the extracted histologic characteristics.
*** indicates statistical significance (p < 0.05). IDC: Invasive Ductal Cancer; ILC: Invasive Lobular Cancer.**

Variation			Study Population	Patients with Metastasis	Control Group	*p*-Value
Menopause	Pre-	*n*	71	17	54	
		%	44.4%	42.5%	45.0%	
	Post-	*n*	89	23	66	0.784
		%	55.6%	57.5%	55.0%	
Hormone Therapy	None	*n*	109	39	35	
		%	90.8%	97.5%	29.2%	
	Positive	*n*	11	1	85	0.168
		%	9.2%	2.5%	70.8%	
Family History	No relatives	*n*	118	32	86	
		%	73.8%	80.0%	71.7%	
	≥1 relative with BC	*n*	42	8	34	0.303
			26.3%	20.0%	28.3%	

Table 22: Description of the extracted clinical characteristics.

Using Spearman's rank test, no correlation was found between the presence of distant metastasis at follow-up and menopausal status, family history, hormone therapy, lesion margins and morphology, kinetic curve, histologic type, histologic class, grading, and ER expression ($p > 0.5$). A significant correlation was found between the presence of distant metastasis at follow-up and the expression of PgR and HER2. The Kruskal–Wallis H test demonstrated a significant correlation between the presence of distant metastasis at follow-up and lesion size (p value < 0.001). No correlation was found between the presence of distant metastasis at follow-up, patient age, ADC values, and Ki-67 index.

The VB-NET model achieved a sensitivity, specificity, accuracy, and AUC of 52.50%, 80.51%, 73.42%, and 68.56%, respectively.

5.2.1 Discussion and Conclusions

Conflicting evidence exists regarding the improved survival of metastatic patients from BC. Some studies have found that the development and spread of adjuvant systemic therapies have resulted in a reduction in the spread of BC metastases, but not an increase in survival, due to the shift from lesions developing in sites with higher survival rates, such as the bone, to more aggressive sites such as the brain and liver[178]. However, according to other studies[179–182], adjuvant therapy has increased patient survival.

BC can be characterized by clinical cancer dormancy[183,184], and a wide window of relapse, which can range from months to decades after treatment of the primary tumor. This is related to the characteristic histologic heterogeneity of breast tumors; due to a higher incidence rate of metastasis, basal-like, and HER2-positive BC having a worse prognosis than tumors that express the estrogen receptor[185]. Therefore, in order to further individualize BC treatment for each patient, it is crucial to identify a potential risk marker for metastasis development.

AI has the ability to infer data from the images that are not visible to the subjective eye of the radiologist. Our aim was to test in our sample whether using a 3T-MRI-based Deep Learning model we can non-invasively predict the risk of distant metastasis from BC. We also studied whether clinical, MRI, or histologic features could play a similar role in this task.

Ours is the first study that attempts to build a predictive model based on CNN Deep Learning for predicting distant metastasis of BC. However, our results demonstrated low AUC, accuracy, and sensitivity values of 68.56%, 80.51%, and 52.50%, respectively. These results reflect a currently insufficient capacity of the Deep Learning approach with CNN in its ability to predict the development of distant metastasis. In our previous studies, we have shown that both a radiomic and a DL approach can predict with greater than 80% accuracy the lymph node status of patients with BC[73,142]. We assume that this is due to the presence of lymphovascular invasion phenomena, some of which are not perceptible to the human eye. In the radiomic approach with a convex hull optimized segmentation, and in DL with the SVB bounding option, mammary regions adjacent to the tumor were also analyzed, obtaining information on possible

lymphatic spread. For the same reason, we found that radiomics applied to the study of peritumoral edema also contribute significantly to predicting tumor histology and prognosis[186].

Furthermore, in this study no correlation was found between the risk of distant metastasis and almost all MRI, histologic, and clinical features considered. Tumor size, and therefore tumor stage, have a very good correlation with the risk of metastasis, as already found in the literature[187,188]; however, Sopik et al. determined that there is a linear relationship between metastatic risk and primary tumor size only for lesions between 7 and 60 mm, while both very small and very large tumors do not respect a linear relationship. No correlation was found between ADC values and the risk of metastasis; on the contrary, Kim et al. found that patients with lower minimum ADC values and higher ADC difference values had worse distant metastasis-free survival[189]. Tumors with these characteristics had greater cellularity and intratumoral heterogeneity, corresponding to greater aggressiveness.

With a DL approach, which is more complex than conventional radiomics, we did not find a meaningful performance of the model for several possible factors. BC is a heterogeneous disease, characterized by considerable pathophysiologic and histologic variability: the different molecular subtypes each have a different rate of aggressiveness and a different mean period free of recurrence or relapse. In addition, each molecular subtype has preferential sites of metastasis development. In a study by Xiao et al., HR+ BCs were correlated with the risk of bone metastasis, HER2+ subtypes were significantly associated with higher rates of liver, brain, and lung metastasis, and triple-negative tumors had a higher rate of brain metastasis but a significantly lower rate of bone metastasis[190]. These differences imply variability in the risk of metastasis as well as their clinical manifestation. This necessarily implies that surgical and pharmacological treatment must be particularly tailored to the patient, an element that may lead to further variation in the risk of distant disease and the time required for the onset of metastasis. A recent study by Cheng et al. investigated and validated the radiomic approach for predicting the risk of metastasis from BC: in this study, patients with the same molecular subtype (triple negative BCs) and with metastasis only in the brain site were selected. We expect that achieving similar homogeneity of the selected patient cohort may result in increased performance of the AI model.

Our study has some limits: this is a retrospective study, with a relatively small number of patients with highly heterogeneous BC types. In addition, we could not explore the possible differences between the different histologic and molecular subtypes, nor the risks of developing metastases for each organ site, due to the small sample size. Segmentation was performed by one radiologist using the same methodology for all lesions, but with a manual assisted approach, which is a method that is time-consuming as well as not error-free. Only the second post-contrast phase was used as the segmentation mask, which, although it is recognized as the phase with higher contrast resolution for BCs, could cause the loss of additional information available from the other sequences. In particular, by considering only the second post-contrast phase, we lost information on the temporal dimension of the post-contrast tumoral enhancement progression, which was instead included in our previous DL study on the prediction of loco-regional lymph node status, which might explain the higher accuracy obtained[142].

Therefore, several steps are needed to improve the current performance of our model: enrollment of more patients to alleviate histological differences in BCs, sharing of different datasets among multiple institutions, development of prospective and multicenter studies, and validation and comparison of additional AI models. Thus, the study population should be as homogeneous as possible, with equal histological and molecular subtypes, and metastases in the same sites. In this regard, in this study we found a significant correlation between HER2, PgR expression, and metastasis risk, so the first successive step should be to select a cohort of patients with the expression of these two specific molecular receptors.

Our study demonstrated a currently insufficient ability of the Deep Learning approach in predicting distant metastasis status in patients with BC using CNNs. Our results suggest that further studies are necessary to investigate the role of AI in this specific task.

6 CONCLUSIONS

This thesis concerns the works we published in about 3 years on the role of artificial intelligence in the evaluation of the breast cancer TNM through magnetic resonance imaging. Nevertheless, we dealt with three major topics: the breast cancer primary tumor, the axillary lymph node involvement and the distant metastasis.

Although the growing interest demonstrated in literature in recent years, in all the above-mentioned topics we noticed that several issues were open, associated to a large heterogeneity in terms of methodology and results in the published papers.

Through our reviews deepened in Chapters 4.1 and 5.1, we realized how the present literature only focuses on the so-called "traditional machine learning systems", with an exceptional shortage in terms of papers addressing the use of Deep learning approaches.

We may boast the absolute novelty of having conducted several experiments, among which the most significant are the evaluation of the role of tumor associated edema in the prediction of tumor histological aggressiveness and the evaluation of a multimodal approach through a deep neural network.

However, the absence of comparable studies could represent a significant limitation since our results require external validation and can be hardly defined as generalizable.

It is noteworthy how in the field of axillary lymph node prediction, both the application of machine learning systems and the deep Learning algorithms led to such encouraging results and how the multimodal approach demonstrated the significant role of pre-contrast sequences, reinforcing the results obtained in the first works on the pre-contrast sequences about the primary tumor, furtherly explained in Chapter 3.1 and 3.2.

Conversely, the same neural network used in the prediction of distant metastases did not perform as well as in the prediction of the axillary lymph node. This essentially means that, from a clinical point of view, it is not always immediate to seek an explanation of the tumor behavior by moving away from the primary site of the breast tumor itself. But on the other hand, from a technical point of view, it indicates the difficulty of finding the optimal algorithm which can predict different outputs.

Surely, one of the main limits of our studies both concerning lymph node status and metastatic diffusion, is represented by the unbalance of input categories. However we tried to overcome this bias trough some tricks, for example by setting a misclassification cost, to make results more reliable and generalizable results.

In conclusion, if the scientific literature of artificial intelligence in the field of breast cancer is in a pioneering phase, the application of radiomics and deep learning in this context are still in an embryonic state.
Research in this topic presents undisputed difficulties due to the absence of consolidated data but at the same time enormous application potential.
The horizon of the medical clinic can be broadened by artificial intelligence if research has shared objectives and transversal collaboration.

7 REFERENCES

1. Kumar V, Abbas A.K, Fausto N AJC. Robins e Cotran. Le basi patologiche delle malattie. Malattie degli organi e degli apparati. Vol. 2. Elsevier. http://www.libreriauniversitaria.it/robbins-cotran-basi-patologiche-malattie/libro/9788821431753. Published 2008. Accessed July 6, 2014.

2. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. C. W. Elston & I. O. Ellis. Histopathology 1991; 19; 403-410. *Histopathology*. 2002;41(3A):151-152, discussion 152-3. http://www.ncbi.nlm.nih.gov/pubmed/12405945. Accessed April 18, 2019.

3. Sotiriou C, Neo S-Y, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci*. 2003;100(18):10393-10398. doi:10.1073/pnas.1732912100

4. Crabb SJ, Cheang MCU, Leung S, et al. Basal breast cancer molecular subtype predicts for lower incidence of axillary lymph node metastases in primary breast cancer. *Clin Breast Cancer*. 2008;8(3):249-256. doi:10.3816/CBC.2008.n.028

5. AIOM associazione italiana di oncologia. Linee guida NEOPLASIA DELLA MAMMELLA. 2018.

6. Limkin EJ, Sun R, Dercle L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol Off J Eur Soc Med Oncol*. 2017;28(6):1191-1206. doi:10.1093/annonc/mdx034

7. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. *Breast Cancer Res Treat*. 2018;169(2):217-229. doi:10.1007/s10549-018-4675-4

8. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577. doi:10.1148/radiol.2015151169

9. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev*. 2000;44(1-2). doi:10.1147/rd.441.0206

10. LeCun, Y., Bengio, Y., Hinton G. Deep learning. nature 521 (7553): 436. *Nature*. 2015;521.

11. Chan TH, Jia K, Gao S, Lu J, Zeng Z, Ma Y. PCANet: A Simple Deep Learning Baseline for Image Classification? *IEEE Trans Image Process*. 2015;24(12). doi:10.1109/TIP.2015.2475625

12. Chen Y, Lin Z, Zhao X, Wang G, Gu Y. Deep learning-based classification of hyperspectral data. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2014;7(6). doi:10.1109/JSTARS.2014.2329330

13. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans Pattern Anal Mach Intell*. 2018;40(4). doi:10.1109/TPAMI.2017.2699184

14. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(6). doi:10.1109/TPAMI.2016.2577031

15. King BF. Guest editorial: Discovery and artificial intelligence. *Am J Roentgenol*. 2017;209(6). doi:10.2214/AJR.17.19178

16. Goodfellow I, Bengio Y, Courville A. Deep Learning. https://www.deeplearningbook.org/. Accessed December 1, 2022.

17. Hecht-Nielsen R. Theory of the backpropagation neural network. 1989:593-605. doi:10.1109/IJCNN.1989.118638

18. Gao J, Zhong X, Li W, et al. Attention-based Deep Learning for the Preoperative Differentiation of Axillary Lymph Node Metastasis in Breast Cancer on DCE-MRI. *J Magn Reson Imaging*. 2022. doi:10.1002/JMRI.28464

19. Nguyen S, Polat D, Karbasi P, et al. Preoperative Prediction of Lymph Node Metastasis from Clinical DCE MRI of the Primary Breast Tumor Using a 4D CNN. *Med Image Comput Comput Assist Interv*. 2020;12262:326-334. doi:10.1007/978-3-030-59713-9_32

20. Matsubayashi RN, Fujii T, Yasumori K, Muranaka T, Momosaki S. Apparent Diffusion Coefficient in Invasive Ductal Breast Carcinoma: Correlation with Detailed Histologic Features and the Enhancement Ratio on Dynamic Contrast-Enhanced MR Images. *J Oncol*. 2010;2010. doi:10.1155/2010/821048

21. Westra C, Dialani V, Mehta TS, Eisenberg RL. Using T2-weighted sequences to more accurately characterize breast masses seen on MRI. *AJR Am J Roentgenol*. 2014;202(3):W183-90. doi:10.2214/AJR.13.11266

22. De Felice C, Cipolla V, Guerrieri D, et al. Apparent diffusion coefficient on 3.0 Tesla magnetic resonance imaging and prognostic factors in breast cancer. *Eur J Gynaecol Oncol*. 2014;35(4):408-414. http://www.ncbi.nlm.nih.gov/pubmed/25118482. Accessed October 23, 2016.

23. Cipolla V, Santucci D, Guerrieri D, Drudi FM, Meggiorini ML, de Felice C. Correlation between 3T apparent diffusion coefficient values and grading of invasive breast carcinoma. *Eur J Radiol*. 2014;83(12):2144-2150. doi:10.1016/j.ejrad.2014.09.015

24. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep Learning for Identifying Metastatic Breast Cancer. 2016:1-6.

25. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science (80-)*. 2011;331(6024). doi:10.1126/science.1203543

26. Bassett LW, Dhaliwal SG, Eradat J, et al. National trends and practices in breast MRI. *AJR Am J Roentgenol*. 2008;191(2):332-339. doi:10.2214/AJR.07.3207

27. Sardanelli F, Boetes C, Borisch B, et al. Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group. *Eur J Cancer*. 2010;46(8):1296-1316. doi:10.1016/j.ejca.2010.02.015

28. Ballesio L, Savelli S, Angeletti M, et al. Breast MRI: Are T2 IR sequences useful in the evaluation of breast lesions? *Eur J Radiol*. 2009;71(1). doi:10.1016/j.ejrad.2008.03.025

29. Arponen O, Masarwah A, Sutela A, et al. Incidentally detected enhancing lesions found in breast MRI: analysis of apparent diffusion coefficient and T2 signal intensity significantly improves specificity. *Eur Radiol*. 2016;26(12). doi:10.1007/s00330-016-4326-2

30. Partridge SC, McDonald ES. Diffusion Weighted Magnetic Resonance Imaging of the Breast: Protocol Optimization, Interpretation, and Clinical Applications. *Magn Reson Imaging Clin N Am*. 2013;21(3). doi:10.1016/j.mric.2013.04.007

31. Tamura T, Murakami S, Naito K, Yamada T, Fujimoto T, Kikkawa T. Investigation of the optimal b-value to detect breast tumors with diffusion weighted imaging by 1.5-T MRI. *Cancer Imaging*. 2014;14(1). doi:10.1186/1470-7330-14-11

32. Shi R yang, Yao Q ying, Wu L ming, Xu J rong. Breast Lesions: Diagnosis Using Diffusion Weighted Imaging at 1.5T and 3.0T—Systematic Review and Meta-analysis. *Clin Breast Cancer*. 2018;18(3). doi:10.1016/j.clbc.2017.06.011

33. Peters NHGM, Vincken KL, Van Den Bosch MAAJ, Luijten PR, Mali WPTM, Bartels LW. Quantitative diffusion weighted imaging for differentiation of benign and malignant breast lesions: The influence of the choice of b-values. *J Magn Reson Imaging*. 2010;31(5). doi:10.1002/jmri.22152

34. Dorrius MD, Dijkstra H, Oudkerk M, Sijens PE. Effect of b value and pre-admission of contrast on diagnostic accuracy of 1.5-T breast DWI: a systematic review and meta-analysis. *Eur Radiol*. 2014;24(11). doi:10.1007/s00330-014-3338-z

35. Gordon Y, Partovi S, Müller-Eschner M, et al. Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion. *Cardiovasc Diagn Ther*. 2014;4(2):147-164. doi:10.3978/j.issn.2223-3652.2014.03.01

36. Kuhl CK. Breast MR imaging at 3T. *Magn Reson Imaging Clin N Am*. 2007;15(3):315-320, vi. doi:10.1016/j.mric.2007.08.003

37. Cui X, Wang N, Zhao Y, et al. Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Cancer using Radiomics Features of DCE-MRI. *Sci Rep*. 2019;9(1):2240. doi:10.1038/s41598-019-38502-0

38. Chen J, Chen J, Yang B, et al. Predicting sentinel lymph node metastasis in a Chinese breast cancer population: assessment of an existing nomogram and a new predictive nomogram. *Breast Cancer Res Treat*. 2012;135(3):839-848. doi:10.1007/s10549-012-2219-x

39. Turnbull LW. Dynamic contrast-enhanced MRI in the diagnosis and management of breast cancer. *NMR Biomed*. 2009;22(1):28-39. doi:10.1002/nbm.1273

40. Svozil D, Kvasnička V, Pospíchal J. Introduction to multi-layer feed-forward neural networks. In: *Chemometrics and Intelligent Laboratory Systems*. Vol 39. ; 1997. doi:10.1016/S0169-7439(97)00061-0

41. Bebis G, Georgiopoulos M. Feed-forward neural networks: why network size is so important. *IEEE Potentials*. 1994.

42. Wang Z, Li M, Wang H, et al. Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion with CNN Deep Features. *IEEE Access*. 2019;7. doi:10.1109/ACCESS.2019.2892795

43. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. Vol 2017-January. ; 2017. doi:10.1109/CVPR.2017.195

44. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. http://code.google.com/p/cuda-convnet/. Accessed December 2, 2022.

45. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol 07-12-June-2015. ; 2015. doi:10.1109/CVPR.2015.7298594

46. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. ; 2015.

47. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol 2016-December. ; 2016. doi:10.1109/CVPR.2016.308

48. Shah SM, Khan RA, Arif S, Sajid U. Artificial intelligence for breast cancer analysis: Trends & directions. *Comput Biol Med*. 2022;142:105221. doi:10.1016/J.COMPBIOMED.2022.105221

49. Atrey PK, Hossain MA, El Saddik A, Kankanhalli MS. Multimodal fusion for multimedia analysis: a survey. *Multimed Syst 2010 166*. 2010;16(6):345-379. doi:10.1007/S00530-010-0182-0

50. Khaleghi B, Khamis A, Karray FO, Razavi SN. Multisensor data fusion: A review of the state-of-the-art. *Inf Fusion*. 2013;14(1):28-44. doi:10.1016/J.INFFUS.2011.08.001

51. Poria S, Cambria E, Gelbukh A. Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-level Multimodal Sentiment Analysis. *Conf Proc - EMNLP 2015 Conf Empir Methods Nat Lang Process*. 2015:2539-2544. doi:10.18653/V1/D15-1303

52. Nojavanasghari B, Gopinath D, Koushik J, Baltrušaitis T, Morency L-P. Deep Multimodal Fusion for Persuasiveness Prediction. doi:10.1145/2993148.2993176

53. Zadeh A, Chen M, Cambria E, Poria S, Morency LP. Tensor Fusion Network for Multimodal Sentiment Analysis. *EMNLP 2017 - Conf Empir Methods Nat Lang Process Proc*. July 2017:1103-1114. doi:10.48550/arxiv.1707.07250

54. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Li FF. Large-scale video classification with convolutional neural networks. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. September 2014:1725-1732. doi:10.1109/CVPR.2014.223

55. Neverova N, Wolf C, Taylor G, Nebout F. ModDrop: adaptive multi-modal gesture recognition.

56. Donegan WL. Tumor-related prognostic factors for breast cancer. *CA Cancer J Clin*. 1997;47(1). doi:10.3322/canjclin.47.1.28

57. Tirada N, Aujero M, Khorjekar G, et al. Breast cancer tissue markers, genomic profiling, and other prognostic factors: A primer for radiologists. *Radiographics*. 2018;38(7). doi:10.1148/rg.2018180047

58. Breast Imaging Reporting & Data System. ACR BI-RADS Atlas 5th Edition. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads. Published 2003.

59. Uematsu T. Focal breast edema associated with malignancy on T2-weighted images of breast MRI: peritumoral edema, prepectoral edema, and subcutaneous edema. *Breast Cancer*. 2015;22(1). doi:10.1007/s12282-014-0572-9

60. Kaiser CG, Herold M, Krammer J, et al. Prognostic value of “prepectoral edema” in MR-mammography. *Anticancer Res*. 2017;37(4). doi:10.21873/anticanres.11542

61. Cipolla V, Santucci D, Guerrieri D, Drudi FM, Meggiorini ML, de Felice C. Correlation between 3T apparent diffusion coefficient values and grading of invasive breast carcinoma. *Eur J Radiol*. 2014;83(12):2144-2150. doi:10.1016/j.ejrad.2014.09.015

62. Cheon H, Kim HJ, Kim TH, et al. Invasive breast cancer: Prognostic value of peritumoral edema identified at preoperative MR imaging. *Radiology*. 2018;287(1). doi:10.1148/radiol.2017171157

63. Kwak JY, Kim EK, Chung SY, et al. Unilateral Breast Edema: Spectrum of Etiologies and Imaging Appearances. *Yonsei Med J*. 2005;46(1):1. doi:10.3349/YMJ.2005.46.1.1

64. Harada TL, Uematsu T, Nakashima K, et al. Is the presence of edema and necrosis on T2WI pretreatment breast MRI the key to predict pCR of triple negative breast cancer? *Eur Radiol*. 2020;30(6). doi:10.1007/s00330-020-06662-7

65. Baltzer PA, Benndorf M, Dietzel M, Gajda M, Camara O, Kaiser WA. Sensitivity and specificity of unenhanced MR mammography (DWI combined with T2-weighted TSE imaging, ueMRM) for the differentiation of mass lesions. *Eur Radiol*. 2010;20(5):1101-1110. doi:10.1007/s00330-009-1654-5

66. Malich A, Fischer DR, Wurdinger S, et al. Potential MRI interpretation model: Differentiation of benign from malignant breast masses. *Am J Roentgenol*. 2005;185(4). doi:10.2214/AJR.04.1073

67. Uematsu T, Kasami M, Watanabe J. Can T2-weighted 3-T breast MRI predict clinically occult inflammatory breast cancer before pathological examination? A single-center experience. *Breast Cancer*. 2014;21(1). doi:10.1007/s12282-012-0425-3

68. Dietzel M, Baltzer PA, Vag T, et al. Magnetic resonance mammography of invasive lobular versus ductal carcinoma: Systematic comparison of 811 patients reveals high diagnostic

accuracy irrespective of typing. *J Comput Assist Tomogr*. 2010;34(4). doi:10.1097/RCT.0b013e3181db9f0e

69. Kawashima H, Kobayashi-Yoshida M, Matsui O, Zen Y, Suzuki M, Inokuchi M. Peripheral hyperintense pattern on T2-weighted magnetic resonance imaging (MRI) in breast carcinoma: Correlation with early peripheral enhancement on dynamic MRI and histopathologic findings. *J Magn Reson Imaging*. 2010;32(5). doi:10.1002/jmri.22279

70. Estourgie SH, Nieweg OE, Valdés Olmos RA, Rutgers EJT, Kroon BBR. Lymphatic Drainage Patterns from the Breast. *Ann Surg*. 2004;239(2). doi:10.1097/01.sla.0000109156.26378.90

71. Blumgart EI, Uren RF, Nielsen PMF, Nash MP, Reynolds HM. Predicting lymphatic drainage patterns and primary tumour location in patients with breast cancer. *Breast Cancer Res Treat*. 2011;130(2). doi:10.1007/s10549-011-1737-2

72. Uematsu T. MRI findings of inflammatory breast cancer, locally advanced breast cancer, and acute mastitis: T2-weighted images can increase the specificity of inflammatory breast cancer. *Breast Cancer*. 2012;19(4). doi:10.1007/s12282-012-0346-1

73. Santucci D, Faiella E, Cordelli E, et al. 3T MRI-radiomic approach to predict for lymph node status in breast cancer patients. *Cancers (Basel)*. 2021;13(9). doi:10.3390/cancers13092228

74. Calabrese A, Santucci D, Landi R, Beomonte Zobel B, Faiella E, de Felice C. Radiomics MRI for lymph node status prediction in breast cancer patients: the state of art. *J Cancer Res Clin Oncol*. 2021;147(6). doi:10.1007/s00432-021-03606-6

75. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446. doi:10.1016/j.ejca.2011.11.036

76. Fan M, Li H, Wang S, Zheng B, Zhang J, Li L. Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer. Weisz A, ed. *PLoS One*. 2017;12(2):e0171683. doi:10.1371/journal.pone.0171683

77. O'Sullivan S, Heinsen H, Grinberg LT, et al. The role of artificial intelligence and machine learning in harmonization of high-resolution post-mortem MRI (virtopsy) with respect to brain microstructure. *Brain informatics*. 2019;6(1):3. doi:10.1186/s40708-019-0096-3

78. Park SH, Choi H-Y, Hahn SY. Correlations between apparent diffusion coefficient values of invasive ductal carcinoma and pathologic factors on diffusion-weighted MRI at 3.0 Tesla. *J Magn Reson Imaging*. 2013;00. doi:10.1002/jmri.24519

79. Houssami N, Lee CI, Buist DSM, Tao D. Artificial intelligence for breast cancer screening: Opportunity or hype? *Breast*. 2017;36. doi:10.1016/j.breast.2017.09.003

80. Pei Z, Cao S, Lu L, Chen W. Direct cellularity estimation on breast cancer histopathology images using transfer learning. *Comput Math Methods Med*. 2019;2019. doi:10.1155/2019/3041250

81. De Kruijf EM, Van Nes JGH, Van De Velde CJH, et al. Tumor-stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients. *Breast Cancer Res Treat*. 2011;125(3). doi:10.1007/s10549-010-0855-6

82. Sataloff DM, Mason BA, Prestipino AJ, Seinige UL, Lieber CP, Baloch Z. Pathologic response to induction chemotherapy in locally advanced carcinoma of the breast: A determinant of outcome. *J Am Coll Surg*. 1995;180(3).

83. Ogston KN, Miller ID, Payne S, et al. A new histological grading system to assess response of breast cancers to primary chemotherapy: Prognostic significance and survival. *Breast*. 2003;12(5). doi:10.1016/S0960-9776(03)00106-1

84. Sinn HP, Schmid H, Junkermann H, et al. HISTOLOGISCHE REGRESSION DES MAMMAKARZINOMS NACH PRIMARER (NEOADJUVANTER) CHEMOTHERAPIE. *Geburtshilfe Frauenheilkd*. 1994;54(10). doi:10.1055/s-2007-1022338

85. Mann RM, Cho N, Moy L. Breast MRI: State of the art. *Radiology*. 2019;292(3):520-536. doi:10.1148/RADIOL.2019182947/SUPPL_FILE/RY182947SUPPF2.JPG

86. Pinker K, Moy L, Sutton EJ, et al. Diffusion-Weighted Imaging with Apparent Diffusion Coefficient Mapping for Breast Cancer Detection as a Stand-Alone Parameter: Comparison with Dynamic Contrast-Enhanced and Multiparametric Magnetic Resonance Imaging. *Invest Radiol*. 2018;53(10). doi:10.1097/RLI.0000000000000465

87. Zhao S, Guo W, Tan R, et al. Correlation between minimum apparent diffusion coefficient values and the histological grade of breast invasive ductal carcinoma. *Oncol Lett*. 2018;15(5). doi:10.3892/ol.2018.8343

88. Azzam H, Mansour S, Salem N, El-Assaly H. Correlative study between ADC value and grading of invasive breast cancer. *Egypt J Radiol Nucl Med*. 2020;51(1). doi:10.1186/s43055-019-0124-6

89. Partridge SC, Amornsiripanitch N. DWI in the Assessment of Breast Lesions. *Top Magn Reson Imaging*. 2017;26(5). doi:10.1097/RMR.0000000000000137

90. Camps-Herrero J. Diffusion-weighted imaging of the breast: current status as an imaging biomarker and future role. *BJR|Open*. 2019;1(1). doi:10.1259/bjro.20180049

91. Yuan C, Jin F, Guo X, Zhao S, Li W, Guo H. Correlation Analysis of Breast Cancer DWI Combined with DCE-MRI Imaging Features with Molecular Subtypes and Prognostic Factors. *J Med Syst*. 2019;43(4):83. doi:10.1007/s10916-019-1197-5

92. Kizildag Yirgin I, Arslan G, Ozturk E, et al. Diffusion Weighted MR Imaging of Breast and Correlation of Prognostic Factors in Breast Cancer. *Balkan Med J*. 2016;33(3):301-307. doi:10.5152/balkanmedj.2016.140555

93. Akın Y, Uğurlu MÜ, Kaya H, Arıbal E. Diagnostic Value of Diffusion-weighted Imaging and Apparent Diffusion Coefficient Values in the Differentiation of Breast Lesions, Histpathologic Subgroups and Correlatıon with Prognostıc Factors using 3.0 Tesla MR. *J breast Heal*. 2016;12(3):123-132. doi:10.5152/tjbh.2016.2897

94. Rakha EA, Reis-Filho JS, Baehner F, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res*. 2010;12(4):207. doi:10.1186/bcr2607

95. Guler EN. Gene Expression Profiling in Breast Cancer and Its Effect on Therapy Selection in Early-Stage Breast Cancer. *Eur J Breast Heal*. 2017;13(4):168-174. doi:10.5152/ejbh.2017.3636

96. Park EK, Cho KR, Seo BK, Woo OH, Cho SB, Bae JW. Additional Value of Diffusion-Weighted Imaging to Evaluate Prognostic Factors of Breast Cancer: Correlation with the Apparent Diffusion Coefficient. *Iran J Radiol*. 2016;13(1):e33133. doi:10.5812/iranjradiol.33133

97. Guvenc I, Akay S, Ince S, et al. Apparent diffusion coefficient value in invasive ductal carcinoma at 3.0 Tesla: Is it correlated with prognostic factors? *Br J Radiol*. 2016;89(1060). doi:10.1259/bjr.20150614

98. Mori N, Ota H, Mugikura S, et al. Luminal-type breast cancer: Correlation of apparent diffusion coefficients with the Ki-67 labeling index. *Radiology*. 2015;274(1). doi:10.1148/radiol.14140283

99. Luo N, Ji Y, Huang X, et al. Changes in Apparent Diffusion Coefficient as Surrogate Marker for Changes in Ki-67 Index Due to Neoadjuvant Chemotherapy in Patients with Invasive Breast Cancer. *Acad Radiol*. 2019;26(10). doi:10.1016/j.acra.2019.01.007

100. Hatakenaka M, Soeda H, Yabuuchi H, et al. Apparent diffusion coefficients of breast tumors: Clinical application. *Magn Reson Med Sci*. 2008;7(1). doi:10.2463/mrms.7.23

101. Choi SY, Chang Y, Park HJ, Kim HJ, Hong SS, Seo DY. Correlation of the apparent diffusion coefficiency values on diffusion-weighted imaging with prognostic factors for breast cancer. *Br J Radiol*. 2012;85(1016):e474-9. doi:10.1259/bjr/79381464

102. Kim K Il, Lee KH, Kim TR, Chun YS, Lee TH, Park HK. Ki-67 as a predictor of response to neoadjuvant chemotherapy in breast cancer patients. *J Breast Cancer*. 2014;17(1). doi:10.4048/jbc.2014.17.1.40

103. Onishi N, Kanao S, Kataoka M, et al. Apparent diffusion coefficient as a potential surrogate marker for Ki-67 index in mucinous breast carcinoma. *J Magn Reson Imaging*. 2015;41(3). doi:10.1002/jmri.24615

104. Shen L, Zhou G, Tong T, et al. ADC at 3.0 T as a noninvasive biomarker for preoperative prediction of Ki67 expression in invasive ductal carcinoma of breast. *Clin Imaging*. 2018;52. doi:10.1016/j.clinimag.2018.02.010

105. Ahn S, Cho J, Sung J, et al. The prognostic significance of tumor-associated stroma in invasive breast carcinoma. *Tumor Biol*. 2012;33(5). doi:10.1007/s13277-012-0411-6

106. Dekker TJA, Van De Velde CJH, Van Pelt GW, et al. Prognostic significance of the tumor-stroma ratio: Validation study in node-negative premenopausal breast cancer patients from the EORTC perioperative chemotherapy (POP) trial (10854). *Breast Cancer Res Treat*. 2013;139(2). doi:10.1007/s10549-013-2571-5

107. Reisenbichler ES, Dupont W, Plummer WD, Hameed O. Is tumor cellularity in primary invasive breast carcinoma of prognostic significance? *Virchows Arch*. 2017;470(6). doi:10.1007/s00428-017-2120-4

108. Guenzi M, Taffurelli M. Trattamento dell'ascella. *Senonetwork Ital*. 2014. http://www.senonetwork.it/app/public/files/download/FOCUS ON SENONETWORK_TRATTAMENTO DELL'ASCELLA_5o0q3rwp.pdf. Accessed March 9, 2020.

109. Liu J, Sun D, Chen L, et al. Radiomics Analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging for the Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. *Front Oncol*. 2019;9. doi:10.3389/fonc.2019.00980

110. Liu C, Ding J, Spuhler K, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI. *J Magn Reson Imaging*. 2019;49(1):131-140. doi:10.1002/jmri.26224

111. Dong Y, Feng Q, Yang W, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI. *Eur Radiol*. 2018;28(2):582-591. doi:10.1007/s00330-017-5005-7

112. Han L, Zhu Y, Liu Z, et al. Radiomic nomogram for prediction of axillary lymph node metastasis in breast cancer. *Eur Radiol*. 2019;29(7):3820-3829. doi:10.1007/s00330-018-5981-2

113. Li H, Zhu Y, Burnside ES, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, oncotype DX, and PAM50 gene assays. *Radiology*. 2016;281(2). doi:10.1148/radiol.2016152110

114. Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L. Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. *Eur J Radiol*. 2017;94. doi:10.1016/j.ejrad.2017.06.019

115. Zhang M, Horvat J V, Bernard-Davila B, et al. Multiparametric MRI model with dynamic contrast-enhanced and diffusion-weighted imaging enables breast cancer diagnosis with high accuracy. *J Magn Reson Imaging*. 2019;49(3):864-874. doi:10.1002/jmri.26285

116. Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys*. 2017;44(10). doi:10.1002/mp.12453

117. Liu M, Mao N, Ma H, et al. Pharmacokinetic parameters and radiomics model based on dynamic contrast enhanced MRI for the preoperative prediction of sentinel lymph node metastasis in breast cancer. *Cancer Imaging*. 2020;20(1). doi:10.1186/s40644-020-00342-x

118. Chai R, Ma H, Xu M, et al. Differentiating axillary lymph node metastasis in invasive breast cancer patients: A comparison of radiomic signatures from multiparametric breast MR sequences. *J Magn Reson Imaging*. 2019;50(4):1125-1132. doi:10.1002/jmri.26701

119. Liu J, Sun D, Chen L, et al. Radiomics Analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging for the Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. *Front Oncol*. 2019;9:980. doi:10.3389/fonc.2019.00980

120. Tan H, Gan F, Wu Y, et al. Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Carcinoma Using Radiomics Features Based on the Fat-Suppressed T2 Sequence. *Acad Radiol*. 2020;27(9). doi:10.1016/j.acra.2019.11.004

121. Shan YN, Xu W, Wang R, Wang W, Pang PP, Shen QJ. A Nomogram Combined Radiomics and Kinetic Curve Pattern as Imaging Biomarker for Detecting Metastatic Axillary Lymph Node in Invasive Breast Cancer. *Front Oncol*. 2020;10. doi:10.3389/fonc.2020.01463

122. Yu Y, Tan Y, Xie C, et al. Development and Validation of a Preoperative Magnetic Resonance Imaging Radiomics-Based Signature to Predict Axillary Lymph Node Metastasis and Disease-Free Survival in Patients With Early-Stage Breast Cancer. *JAMA Netw open*. 2020;3(12). doi:10.1001/jamanetworkopen.2020.28086

123. Lei Xu, Pingfan Yan, Tong Chang. Best first strategy for feature selection. In: *[1988 Proceedings] 9th International Conference on Pattern Recognition*. IEEE Comput. Soc. Press; :706-708. doi:10.1109/ICPR.1988.28334

124. Eddy WF. A New Convex Hull Algorithm for Planar Sets. *ACM Trans Math Softw*. 1977;3(4):398-403. doi:10.1145/355759.355766

125. Cordelli E, Maulucci G, De Spirito M, Rizzi A, Pitocco D, Soda P. A decision support system for type 1 diabetes mellitus diagnostics based on dual channel analysis of red blood cell membrane fluidity. *Comput Methods Programs Biomed*. 2018;162:263-271. doi:10.1016/j.cmpb.2018.05.025

126. Bino S V, Unnikrishnan A, Balakrishnan K. GREY LEVEL CO-OCCURRENCE MATRICES: GENERALISATION AND SOME NEW FEATURES. *Int J Comput Sci Eng Inf Technol*. 2012;2(2). https://arxiv.org/pdf/1205.4831.pdf. Accessed August 7, 2020.

127. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. In: *IEEE Transactions on Systems, Man, and Cybernetics*. Vol SMC-3. ; 1973:610-621. doi:10.1109/TSMC.1973.4309314

128. Zhao G, Pietik¨aine M. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. In: *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*. ; 2007. http://www.ee.oulu.fi/research/imag/mvg/files/pdf/pdf_740.pdf. Accessed August 7, 2020.

129. Sicilia R, Cordelli E, Merone M, et al. Early Radiomic Experiences in Classifying Prostate Cancer Aggressiveness using 3D Local Binary Patterns. In: ; 2019. doi:10.1109/CBMS.2019.00078

130. Arcuri A, Fraser G. Parameter tuning or default values? An empirical investigation in search-based software engineering. *Empir Softw Eng*. 2013;18(3). doi:10.1007/s10664-013-9249-9

131. Nemoto T, Vana J, Bedwani RN, Baker HW, McGregor FH, Murphy GP. Management and survival of female breast cancer: results of a national survey by the American College of Surgeons. *Cancer*. 1980;45(12):2917-2924. http://www.ncbi.nlm.nih.gov/pubmed/7388735. Accessed January 2, 2019.

132. Qiu P, Liu J, Wang Y, et al. Risk factors for sentinel lymph node metastasis and validation study of the MSKCC nomogram in breast cancer patients. *Jpn J Clin Oncol*. 2012;42(11):1002-1007. doi:10.1093/jjco/hys150

133. Rahbar H, Partridge SC, DeMartini WB, Thursten B, Lehman CD. Clinical and technical considerations for high quality breast MRI at 3 tesla. *J Magn Reson Imaging*. 2013;37(4):778-790. doi:10.1002/jmri.23834

134. Cordelli E, Sicilia R, Santucci D, et al. Radiomics-based non-invasive lymph node metastases prediction in breast cancer. In: *Proceedings - IEEE Symposium on Computer-Based Medical Systems*. Vol 2020-July. ; 2020. doi:10.1109/CBMS49503.2020.00098

135. Lee SH, Park H, Ko ES. Radiomics in breast imaging from techniques to clinical applications: A review. *Korean J Radiol*. 2020;21(7). doi:10.3348/kjr.2019.0855

136. Lee YW, Huang CS, Shih CC, Chang RF. Axillary lymph node metastasis status prediction of early-stage breast cancer using convolutional neural networks. *Comput Biol Med*. 2021;130. doi:10.1016/j.compbiomed.2020.104206

137. Zhou LQ, Wu XL, Huang SY, et al. Lymph node metastasis prediction from primary breast cancer US images using deep learning. *Radiology*. 2020;294(1). doi:10.1148/radiol.2019190372

138. Zheng X, Yao Z, Huang Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-15027-z

139. Kim SA, Cho N, Ryu EB, et al. Background parenchymal signal enhancement ratio at preoperative mr imaging: Association with subsequent local recurrence in patients with ductal carcinoma in situ after breast conservation surgery. *Radiology*. 2014;270(3). doi:10.1148/radiol.13130459

140. Sun Q, Lin X, Zhao Y, et al. Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. *Front Oncol*. 2020;10. doi:10.3389/fonc.2020.00053

141. Sinn HP, Kreipe H. A brief overview of the WHO classification of breast tumors, 4th edition, focusing on issues and updates from the 3rd edition. *Breast Care*. 2013;8(2). doi:10.1159/000350774

142. Santucci D, Faiella E, Gravina M, et al. CNN-Based Approaches with Different Tumor Bounding Options for Lymph Node Status Prediction in Breast DCE-MRI. *Cancer*. 2022;14(19). doi:10.3390/cancers14194574

143. Woodhams R, Matsunaga K, Iwabuchi K, et al. Diffusion-weighted imaging of malignant breast tumors: the usefulness of apparent diffusion coefficient (ADC) value and ADC map for the detection of malignant breast tumors and evaluation of cancer extension. *J Comput Assist Tomogr*. 2005;29(5):644-649. http://www.ncbi.nlm.nih.gov/pubmed/16163035. Accessed April 27, 2014.

144. Ramachandram D, Taylor GW. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process Mag*. 2017;34(6). doi:10.1109/MSP.2017.2738401

145. Mattes D, Haynor DR, Vesselle H, Lewellyn TK, Eubank W. Nonrigid multimodality image registration. In: *Medical Imaging 2001: Image Processing*. Vol 4322. ; 2001:1609-1620.

146. Rahunathan S, Stredney D, Schmalbrock P, Clymer BD. Image registration using rigid registration and maximization of mutual information. In: *13th Annu. Med. Meets Virtual Reality Conf.* ; 2005.

147. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. ; 2016:770-778.

148. Chen S, Ma K, Zheng Y. Med3D: Transfer Learning for 3D Medical Image Analysis. *arXiv Prepr arXiv190400625*. 2019.

149. Coleman RE. Metastatic bone disease: Clinical features, pathophysiology and treatment strategies. *Cancer Treat Rev*. 2001;27(3). doi:10.1053/ctrv.2000.0210

150. Cecchini MG, Wetterwald A, van der Pluijm G, Thalmann GN. Molecular and biological mechanisms of bone metastasis. *EAU Updat Ser*. 2005;3(4). doi:10.1016/j.euus.2005.09.006

151. Selvaggi G, Scagliotti G V. Management of bone metastases in cancer: A review. *Crit Rev Oncol Hematol*. 2005;56(3). doi:10.1016/j.critrevonc.2005.03.011

152. Yu HHM, Tsai YY, Hoffe SE. Overview of diagnosis and management of metastatic disease to bone. *Cancer Control*. 2012;19(2). doi:10.1177/107327481201900202

153. Coleman RE, Seaman JJ. The role of zoledronic acid in cancer: Clinical studies in the treatment and prevention of bone metastases. *Semin Oncol*. 2001;28(2 SUPPL. 6). doi:10.1053/sonc.2001.24155

154. Bussard KM, Gay C V., Mastro AM. The bone microenvironment in metastasis; what is special about bone? *Cancer Metastasis Rev*. 2008;27(1). doi:10.1007/s10555-007-9109-4

155. Cuccurullo V, Lucio Cascini G, Tamburrini O, Rotondo A, Mansi L. Bone Metastases Radiopharmaceuticals: An Overview. *Curr Radiopharm*. 2013;6(1). doi:10.2174/1874471011306010007

156. O'Sullivan GJ. Imaging of bone metastasis: An update. *World J Radiol*. 2015;7(8). doi:10.4329/wjr.v7.i8.202

157. Talbot JN, Paycha F, Balogova S. Diagnosis of bone metastasis: Recent comparative studies of imaging modalities. *Q J Nucl Med Mol Imaging*. 2011;55(4).

158. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5(1):4006. doi:10.1038/ncomms5006

159. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12). doi:10.1038/nrclinonc.2017.141

160. Jiang X, Ren M, Shuang X, et al. Multiparametric MRI-Based Radiomics Approaches for Preoperative Prediction of EGFR Mutation Status in Spinal Bone Metastases in Patients with Lung Adenocarcinoma. *J Magn Reson Imaging*. 2021;54(2). doi:10.1002/jmri.27579

161. Ren M, Yang H, Lai Q, et al. MRI-based radiomics analysis for predicting the EGFR mutation based on thoracic spinal metastases in lung adenocarcinoma patients. *Med Phys*. 2021;48(9). doi:10.1002/mp.15137

162. Fan Y, Dong Y, Yang H, et al. Subregional radiomics analysis for the detection of the EGFR mutation on thoracic spinal metastases from lung cancer. *Phys Med Biol*. 2021;66(21). doi:10.1088/1361-6560/ac2ea7

163. Cao R, Dong Y, Wang X, et al. MRI-Based Radiomics Nomogram as a Potential Biomarker to Predict the EGFR Mutations in Exon 19 and 21 Based on Thoracic Spinal Metastases in Lung Adenocarcinoma. *Acad Radiol*. 2022;29(3). doi:10.1016/j.acra.2021.06.004

164. Wang Y, Yu B, Zhong F, et al. MRI-based texture analysis of the primary tumor for pre-treatment prediction of bone metastases in prostate cancer. *Magn Reson Imaging*. 2019;60. doi:10.1016/j.mri.2019.03.007

165. Hayakawa T, Tabata K ichi, Tsumura H, et al. Size of pelvic bone metastasis as a significant prognostic factor for metastatic prostate cancer patients. *Jpn J Radiol*. 2020;38(10). doi:10.1007/s11604-020-01004-5

166. Zhang W, Mao N, Wang Y, et al. A Radiomics nomogram for predicting bone metastasis in newly diagnosed prostate cancer patients. *Eur J Radiol*. 2020;128. doi:10.1016/j.ejrad.2020.109020

167. Sun W, Liu S, Guo J, et al. A CT-based radiomics nomogram for distinguishing between benign and malignant bone tumours. *Cancer Imaging*. 2021;21(1). doi:10.1186/s40644-021-00387-6

168. Xiong X, Wang J, Hu S, Dai Y, Zhang Y, Hu C. Differentiating Between Multiple Myeloma and Metastasis Subtypes of Lumbar Vertebra Lesions Using Machine Learning–Based Radiomics. *Front Oncol*. 2021;11. doi:10.3389/fonc.2021.601699

169. Yin P, Mao N, Zhao C, Wu J, Chen L, Hong N. A Triple-Classification Radiomics Model for the Differentiation of Primary Chordoma, Giant Cell Tumor, and Metastatic Tumor of Sacrum Based on T2-Weighted and Contrast-Enhanced T1-Weighted MRI. *J Magn Reson Imaging*. 2019;49(3). doi:10.1002/jmri.26238

170. Wang Z-J, Wang M-Q, Duan F, et al. Transcatheter arterial chemoembolization followed by immediate radiofrequency ablation for large solitary hepatocellular carcinomas. *World J Gastroenterol*. 2013;19(26):4192-4199. doi:10.3748/wjg.v19.i26.4192

171. Filograna L, Lenkowicz J, Cellini F, et al. Identification of the most significant magnetic resonance imaging (MRI) radiomic features in oncological patients with vertebral bone marrow metastatic disease: a feasibility study. *Radiol Medica*. 2019;124(1). doi:10.1007/s11547-018-0935-y

172. Lang N, Zhang Y, Zhang E, et al. Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI. *Magn Reson Imaging*. 2019;64. doi:10.1016/j.mri.2019.02.013

173. Peh WCG. CT-guided percutaneous biopsy of spinal lesions. *Biomed Imaging Interv J*. 2006;2(3). doi:10.2349/biij.2.3.e25

174. Fiz F, Viganò L, Gennaro N, et al. Radiomics of liver metastases: A systematic review. *Cancers (Basel)*. 2020;12(10). doi:10.3390/cancers12102881

175. Stanzione A, Gambardella M, Cuocolo R, Ponsiglione A, Romeo V, Imbriaco M. Prostate MRI radiomics: A systematic review and radiomic quality score assessment. *Eur J Radiol*. 2020;129. doi:10.1016/j.ejrad.2020.109095

176. Chetan MR, Gleeson F V. Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. *Eur Radiol*. 2021;31(2). doi:10.1007/s00330-020-07141-9

177. Kendrick J, Francis R, Hassan GM, et al. Radiomics for Identification and Prediction in Metastatic Prostate Cancer: A Review of Studies. *Front Oncol*. 2021;11. doi:10.3389/fonc.2021.771787

178. McCormack VA, Joffe M, van den Berg E, et al. Breast cancer receptor status and stage at diagnosis in over 1,200 consecutive public hospital patients in Soweto, South Africa: a case series. *Breast Cancer Res*. 2013;15(5):R84. doi:10.1186/bcr3478

179. Veronesi U, Paganelli G, Viale G, et al. A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer. *N Engl J Med*. 2003;349(6):546-553. doi:10.1056/NEJMoa012782

180. Dafni U, Grimani I, Xyrafas A, Eleftheraki AG, Fountzilas G. Fifteen-year trends in metastatic breast cancer survival in Greece. *Breast Cancer Res Treat*. 2010;119(3). doi:10.1007/s10549-009-0630-8

181. Mauri D, Polyzos NP, Salanti G, Pavlidis N, Ioannidis JPA. Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. *J Natl Cancer Inst*. 2008;100(24). doi:10.1093/jnci/djn414

182. Daniels B, Kiely BE, Lord SJ, et al. Trastuzumab for metastatic breast cancer: Real world outcomes from an Australian whole-of-population cohort (2001–2016). *Breast*. 2018;38. doi:10.1016/j.breast.2017.11.007

183. Uhr JW, Pantel K. Controversies in clinical cancer dormancy. *Proc Natl Acad Sci U S A*. 2011;108(30). doi:10.1073/pnas.1106613108

184. Sosa MS, Bragado P, Aguirre-Ghiso JA. Mechanisms of disseminated cancer cell dormancy: An awakening field. *Nat Rev Cancer*. 2014;14(9). doi:10.1038/nrc3793

185. Koscielny S, Tubiana M, Lê MG, et al. Breast cancer: Relationship between the size of the primary tumour and the probability of metastatic dissemination. *Br J Cancer*. 1984;49(6). doi:10.1038/bjc.1984.112

186. Santucci D, Faiella E, Cordelli E, et al. The impact of tumor edema on T2-weighted 3T-MRI invasive breast cancer histological characterization: A pilot radiomics study. *Cancers (Basel)*. 2021;13(18). doi:10.3390/cancers13184635

187. Sopik V, Narod SA. The relationship between tumour size, nodal status and distant metastases: on the origins of breast cancer. *Breast Cancer Res Treat*. 2018;170(3). doi:10.1007/s10549-018-4796-9

188. Kim JY, Kim JJ, Hwangbo L, Kang T, Park H. Diffusion-weighted imaging of invasive breast cancer: Relationship to distant metastasis-free survival. *Radiology*. 2019;291(2). doi:10.1148/radiol.2019181706

189. Xiao W, Zheng S, Yang A, et al. Breast cancer subtypes and the risk of distant metastasis at initial diagnosis: A population-based study. *Cancer Manag Res*. 2018;10. doi:10.2147/CMAR.S176763

190. Cheng X, Xia L, Sun S. A pre-operative MRI-based brain metastasis risk-prediction model for triple-negative breast cancer. *Gland Surg*. 2021;10(9). doi:10.21037/gs-21-537

8 Appendix: published works

1. Calabrese, A.; **Santucci, D**.; Gravina, M.; Faiella, E.; Cordelli, E.; Soda, P.; Iannello, G.; Sansone, C.; Zobel, B.B.; Catalano, C.; de Felice, C. 3T-MRI Artificial Intelligence in Patients with Invasive Breast Cancer to Predict Distant Metastasis Status: A Pilot Study. Cancers 2023, 15, 36. https://doi.org/10.3390/cancers15010036
2. Celli V, Guerreri M, Pernazza A, Cuccu I, Palaia I, Tomao F, Di Donato V, Pricolo P, Ercolani G, Ciulla S, Colombo N, Leopizzi M, Di Maio V, Faiella E, **Santucci D**, Soda P, Cordelli E, Perniola G, Gui B, Rizzo S, Della Rocca C, Petralia G, Catalano C, Manganaro L. MRI- and Histologic-Molecular-Based Radio-Genomics Nomogram for Preoperative Assessment of Risk Classes in Endometrial Cancer. Cancers (Basel). 2022 Nov 29;14(23):5881. doi: 10.3390/cancers14235881. PMID: 36497362; PMCID: PMC9739755.
3. D. Vertulli, **D. Santucci**, F. Esperto, B. Beomonte Zobel, R.F. Grasso, E. Faiella. , Impact of adipose tissue distribution on prostate cancer recurrence after radical prostatectomy. Actas Urol Esp (Engl Ed). 2022 Oct. DOI: 10.1016/j.acuroe.2022.06.008
4. Faiella, E.; Calabrese, A.; **Santucci, D**.; Corti, R.; Cionfoli, N.; Pusceddu, C.; de Felice, C.; Bozzini, G.; Mazzoleni, F.; Muraca, R.M.; Moramarco, L.P.; Venturini, M.; Quaretti, P. Green Tattoo Pre-Operative Renal Embolization for Robotic-Assisted and Laparoscopic Partial Nephrectomy: A Practical Proof of a New Technique. *J. Clin. Med*. 2022, *11*, 6816. https://doi.org/10.3390/jcm11226816
5. Faiella E, Calabrese A, **Santucci D**, de Felice C, Pusceddu C, Fior D, Fontana F, Piacentino F, Moramarco LP, Muraca RM, Venturini M. Combined Trans-Arterial Embolization and Ablation for the Treatment of Large (>3 cm) Liver Metastases: Review of the Literature. Journal of Clinical Medicine. 2022; 11(19):5576. https://doi.org/10.3390/jcm11195576
6. **Santucci D**, Faiella E, Gravina M, Cordelli E, de Felice C, Beomonte Zobel B, Iannello G, Sansone C, Soda P. CNN-Based Approaches with Different Tumor Bounding Options

for Lymph Node Status Prediction in Breast DCE-MRI. Cancers. 2022; 14(19):4574. https://doi.org/10.3390/cancers14194574

7. Pusceddu C, De Francesco D, Ballicu N, **Santucci D**, Marsico S, Venturini M, Fior D, Moramarco LP, Faiella E. Safety and Feasibility of Steerable Radiofrequency Ablation in Combination with Cementoplasty for the Treatment of Large Extraspinal Bone Metastases. Curr Oncol. 2022 Aug 20;29(8):5891-5900. doi: 10.3390/curroncol29080465. PMID: 36005203; PMCID: PMC9406475.
8. Faiella E, **Santucci D**, Vertulli D, Esperto F, Messina L, Castiello G, Papalia R, Flammia G, Scarpa RM, Fiore M, Trodella LE, Ramella S, Grasso RF, Beomonte Zobel B. The role of multiparametric mri in the diagnosis of local recurrence after radical prostatectomy and before salvage radiotherapy. Actas Urol Esp (Engl Ed). 2022 Sep;46(7):397-406. English, Spanish. doi: 10.1016/j.acuroe.2021.12.011. Epub 2022 Jun 28. PMID: 35778338.
9. Faiella E, Vertulli D, Esperto F, Cordelli E, Soda P, Muraca RM, Moramarco LP, Grasso RF, Beomonte Zobel B, **Santucci D**. Quantib Prostate Compared to an Expert Radiologist for the Diagnosis of Prostate Cancer on mpMRI: A Single-Center Preliminary Study. Tomography. 2022 Aug 13;8(4):2010-2019. doi: 10.3390/tomography8040168. PMID: 36006066; PMCID: PMC9415513.
10. Pusceddu C, De Francesco D, Ballicu N, **Santucci D**, Marsico S, Venturini M, Fior D, Moramarco LP, Faiella E. Safety and Feasibility of Steerable Radiofrequency Ablation in Combination with Cementoplasty for the Treatment of Large Extraspinal Bone Metastases. Curr Oncol. 2022 Aug 20;29(8):5891-5900. doi: 10.3390/curroncol29080465. PMID: 36005203; PMCID: PMC9406475.
11. Faiella E, Castiello G**, Santucci D**, Pacella G, Bernetti C, Muley Villamu M, Antonelli Incalzi R, Beomonte Zobel B, Quattrocchi CC, Grasso RF. Analysis of Risk Factors of Soft Tissue Bleeding in COVID-19 Patients: A Point of View After Two Years of Pandemic. J Clin Med Res. 2022 May;14(5):188-195. doi: 10.14740/jocmr4708. Epub 2022 May 31. PMID: 35720229; PMCID: PMC9187356.
12. Faiella, E.; **Santucci, D**.; Calabrese, A.; Russo, F.; Vadalà, G.; Zobel, B.B.; Soda, P.; Iannello, G.; de Felice, C.; Denaro, V. Artificial Intelligence in Bone Metastases: An

MRI and CT Imaging Review. Int. J. Environ. Res. Public Health 8 Feb 2022, 19, 1880. https://doi.org/10.3390/ijerph19031880

13. Stefanucci R, **Santucci D**, Rossi S, Sammarra M, Faiella E, Cordelli E, Altomare V, Grasso RF, Zobel BB. A case of secretory carcinoma in a patient with a story of controlateral medullary carcinoma. Cancer Diagnosis & Prognosis. Jan 2022 2(1); 87-92. DOI: 10.21873/cdp.10081
14. **Santucci D,** Faiella E, Cordelli E, Calabrese A, Landi R, de Felice C, Beomonte Zobel B, Grasso RF, Iannello G, Soda P. The Impact of Tumor Edema on T2-Weighted 3T-MRI Invasive Breast Cancer Histological Characterization: A Pilot Radiomics Study. Cancers (Basel). 2021 Sep 15;13(18):4635. doi: 10.3390/cancers13184635. PMID: 34572862; PMCID: PMC8464682.
15. **Santucci D**, Faiella E, Calabrese A, Beomonte Zobel B, Ascione A, Cerbelli B, Iannello G, Soda P, de Felice C. On the Additional Information Provided by 3T-MRI ADC in Predicting Tumor Cellularity and Microscopic Behavior. Cancers (Basel). 2021 Oct 15;13(20):5167. doi: 10.3390/cancers13205167. PMID: 34680316; PMCID: PMC8534264.
16. **Santucci D**, Faiella E, Cordelli E, Sicilia R, de Felice C, Zobel BB, Iannello G, Soda P. 3T MRI-Radiomic Approach to Predict for Lymph Node Status in Breast Cancer Patients. Cancers (Basel). 2021 May 6;13(9):2228. doi: 10.3390/cancers13092228. PMID: 34066451; PMCID: PMC8124168
17. Calabrese A, **Santucci D**, Landi R, Beomonte Zobel B, Faiella E, de Felice C. Radiomics MRI for lymph node status prediction in breast cancer patients: the state of art. J Cancer Res Clin Oncol. 2021 Jun;147(6):1587-1597. doi: 10.1007/s00432-021-03606-6. Epub 2021 Mar 23. PMID: 33758997.
18. Cordelli, E., Sicilia, R., **Santucci, D**., De Felice, C., Quattrocchi, C.C., Zobel, B.B., Iannello, G., Soda, P. Radiomics-Based Non-Invasive Lymph Node Metastases Prediction in Breast Cancer, *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, Rochester, MN, USA, 2020, pp. 486-491, doi: 10.1109/CBMS49503.2020.00098.

19. Faiella E, **Santucci D**, Bernetti C, Schena E, Pacella G, Beomonte Zobel B, Grasso RF (2020) Combined trans-arterial embolisation and microwave ablation for the treatment of large unresectable hepatic metastases (>3 cm in maximal diameter), International Journal of Hyperthermia, 37:1, 1395-1403, DOI: 10.1080/02656736.2020.1849823
20. Scaglione M, Galluzzo M, **Santucci D**, Trinci M, Messina L, Laccetti E, Faiella E, Beomonte Zobel B. Small bowel obstruction and intestinal ischemia: emphasizing the role of MDCT in the management decision process. Abdom Radiol (NY). 2020 Oct 14. doi: 10.1007/s00261-020-02800-3. Epub ahead of print. PMID: 33057806 (IF: 2.249)