

3D DCE-MRI Radiomic Analysis for Malignant Lesion Prediction in Breast Cancer Patients

Abstract

Rationale and Objectives: To develop and validate a radiomic model, with radiomic features extracted from breast Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) from a 1.5T scanner, for predicting the malignancy of masses with enhancement. Images were acquired using an 8-channel breast coil in the axial plane. The rationale behind this study is to show the feasibility of a radiomics-powered model that could be integrated into the clinical practice by exploiting only standard-of-care DCE-MRI with the goal of reducing the required image pre-processing (i.e., normalization and quantitative imaging map generation).

Materials and Methods: 107 radiomic features were extracted from a manually annotated dataset of 111 patients, which was split into discovery and test sets. A feature calibration and pre-processing step was performed to find only robust non-redundant features. An in-depth discovery analysis was performed to define a predictive model: for this purpose, a Support Vector Machine (SVM) was trained in a nested 5-fold cross-validation scheme, by exploiting several unsupervised feature selection methods. The predictive model performance was evaluated in terms of Area Under the Receiver Operating Characteristic (AUROC), specificity, sensitivity, PPV and NPV. The test was performed on unseen held-out data.

Results: The model combining Unsupervised Discriminative Feature Selection (UDFS) and SVMs on average achieved the best performance on the blinded test set: AUROC = 0.725 ± 0.091 , sensitivity = 0.709 ± 0.176 , specificity = 0.741 ± 0.114 , PPV = 0.72 ± 0.093 , and NPV = 0.75 ± 0.114 .

Conclusion: In this study, we built a radiomic predictive model based on breast DCE-MRI, using only the strongest enhancement phase, with promising results in terms of accuracy and specificity in the differentiation of malignant from benign breast lesions.

Key Words: Breast cancer; dynamic contrast-enhanced magnetic resonance imaging; machine learning; radiomics; unsupervised feature selection, support vector machines

Abbreviations List:

3DS	3D Shape
ADH	Atypical Ductal Hyperplasia
AUROC	Area Under the Receiver Operating Characteristic
CV	Cross-Validation
DCE-MRI	Dynamic Contrast-Enhanced Magnetic Resonance Imaging
DGUFs	Dependence Guided Unsupervised Feature Selection
FS	Feature Selection
GLCM	Gray Level Co-occurrence Matrix features
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
ICC	Intraclass Correlation Coefficient
LRLGLE	Long Run Low Gray Level Emphasis
NGTDM	Neighboring Gray Tone Difference Matrix
NIfTI	Neuroimaging Informatics Technology Initiative
NPV	Negative Predictive Value

PPV	Positive Predictive Value
SOR	Standard-of-Reference
SRE	Short Run Emphasis
SVM	Support Vector Machine
TE	Echo Time
TR	Repetition Time
UDFS	Unsupervised Discriminative Feature Selection
UFSOL	Unsupervised Feature Selection with Ordinal Locality

Introduction

Breast cancer currently represents the most common non-skin cancer in women and men, accounting for 11.7% of all new cancer diagnoses in 2020 [1]. Because of its incidence and clinical impact, early and accurate cancer detection and characterization is of utmost importance. Diagnosis of early invasive breast cancer relies on clinical evaluation, radiological imaging and image-guided biopsy. Breast MRI is commonly used as a screening and problem-solving tool [2] and for the local staging of breast cancer [3]. Its high sensitivity for the detection of breast lesions makes MRI a valuable screening tool, especially in patients at high risk for developing breast cancer. However, several studies have shown limited specificity when it comes to lesion characterization leading to high recall rates and the need for invasive and costly biopsies [4]. Radiomics studies in breast imaging have shown promising results for lesion characterization, prediction of nodal metastases, tumor subtype, response predictions and prognostication [5, 6].

Among imaging techniques, Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) provides both morphological (e.g., size, margins, shape) [7] and hemodynamic information, assessing tumor vascularity and defining enhancement kinetic curves, with a reported sensitivity higher than 90% for breast cancer diagnosis [8]. However, the specificity of DCE-MRI is still substantially low (about 72%) and it is often necessary to continue the diagnostic work-up with a biopsy [8].

In breast imaging, radiomics was applied to several imaging modalities, including MRI, mammography, ultrasound, and digital breast tomosynthesis [9]. In the recent review by [10], 63% of the reviewed radiomics studies were based on MRI, thus demonstrating the high relevance for the scientific community. Particularly, the application of radiomics in breast DCE-MRI was assessed by addressing various issues in the evaluation of breast lesions from the diagnosis (e.g., characterization of breast lesions, prediction of breast cancer histological types and correlation with receptor status) to the prognosis (e.g., lymph node metastases, tumor response to neoadjuvant systemic therapy, recurrence risks) [11].

Radiomics extracts a large amount of quantitative imaging features from medical images, conveying more information than the visual and qualitative patterns observable by the radiologists' naked eye [12]. In particular, radiomic biomarkers – which has attracted the attention of the scientific community, especially in the case of oncological imaging – can be associated with clinical outcomes [13]. With this regard, quantitative imaging biomarkers have shown great potential in prediction, prognosis or treatment response assessment [14, 15]. To increase the robustness of these biomarkers, systematic studies on radiomic feature reliability were presented [16, 17]. As a matter of fact, initiatives and roadmaps were conducted in the

last years to support the lack of reproducibility and validation of radiomics studies [18–20]. Finally, to facilitate the clinical translation of the developed predictive and prognostic models, biological meaning and validation have been recently investigated [21, 22].

The development of robust biomarkers could accelerate their incorporation into personalized medicine [23]. Thus, standardization initiatives have been carried out by the scientific community to deal with the lack of reproducibility and validation of radiomics studies [20]. An accurate and careful preliminary analysis of the robustness of the radiomic features is mandatory to define clinically relevant biomarkers able to provide generalization abilities on external datasets.

The primary objective of this study is to develop and validate a radiomic model capable of predicting malignant breast masses by using 3D radiomic features extracted from DCE-MRI, while the rationale is to show the feasibility of a radiomics-powered model that could be integrated into the clinical practice by exploiting only standard-of-care DCE-MRI with the goal of reducing the required image pre-processing (i.e., normalization and quantitative imaging map generation).

Material and Methods

Patients and imaging data selection

Retrospective data collection was approved by the local Ethics Committee. The requirement for evidence of informed consent was waived because of the retrospective nature of our study.

A total of 194 Breast DCE-MRI exams, performed from November 2019 to October 2020 at the *****BLINDED***** were recruited for this study. The exclusion criteria applied during population enrolment are illustrated in Figure 1.

The whole dataset resulting consisted of 111 lesions (size range: 5-60 mm; mean size \pm SD: 16.67 ± 11.18 mm), depicted at DCE-MRI as masses with enhancement in 111 patients (110 women and 1 man; age range: 23–72 years; mean age \pm SD: 47.83 ± 9.16 years), and classified by a breast radiologist (with more than 30-year experience) according to BI-RADS criteria [24].

In patients with multi-focal lesions only the largest one was considered. As standard-of-reference (SOR), we included a definitive histological diagnosis for all breast lesions classified as BI-RADS 4-5 and a definitive histological diagnosis or a complete follow-up at 24 months for BI-RADS 3 lesions.

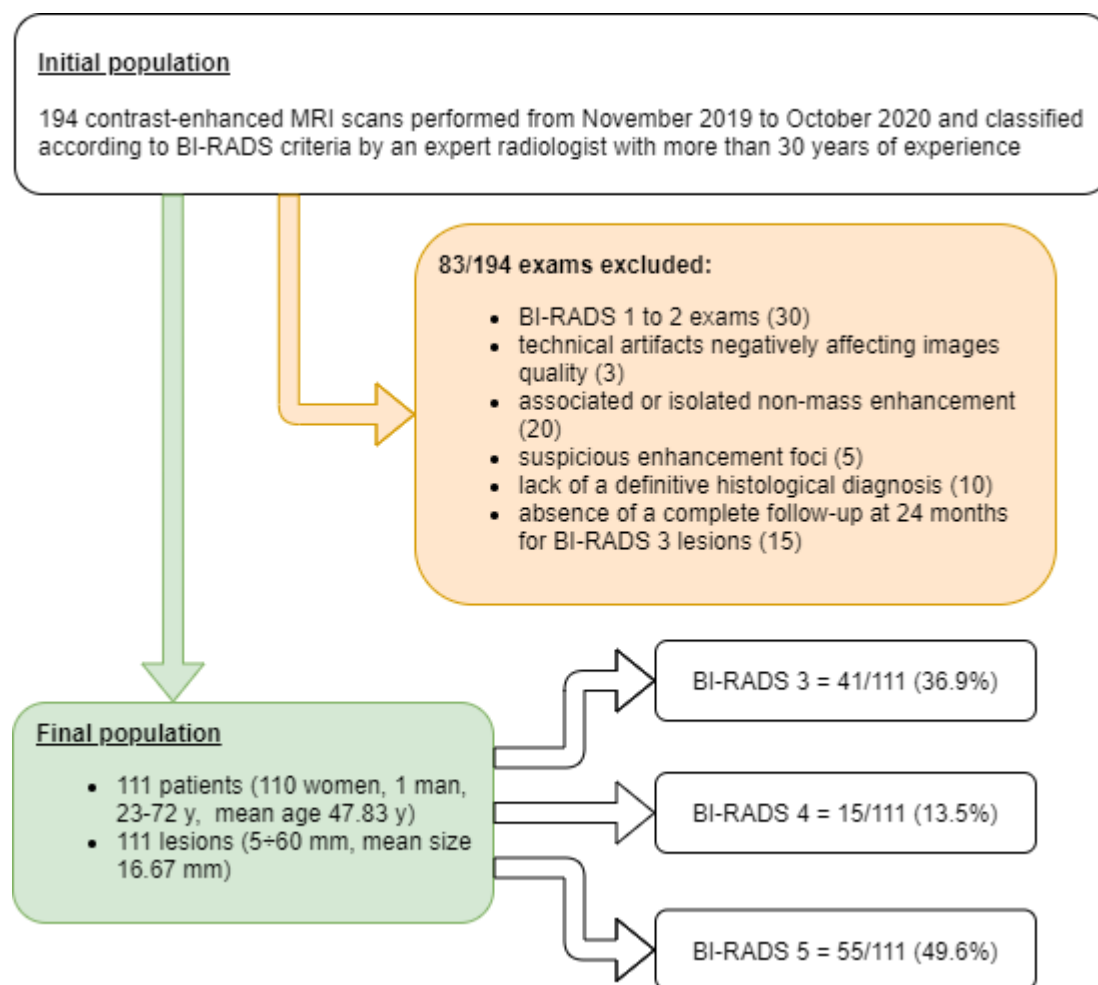


Figure 1. Diagram illustrating the criteria considered in this study for inclusion/exclusion.

MR images were acquired using a 1.5T scanner (Signa HDxt; GE Healthcare, Barrington, IL, USA) using an 8-channel breast coil and a standard protocol including T2w FSE with and without fat saturation, Diffusion-Weighted Imaging (DWI) and DCE-MRI (T1-weighted three-dimensional spoiled gradient echo) sequences in the axial plane. The second one was analyzed for radiomics purposes and imaging parameters are described in Table 1.

The whole dataset of 111 breast DCE-MRI sequences was divided into two groups, *via* an 80%-20% split hold-out approach, obtaining a:

- *discovery set*: used to define the best predictive model, in terms of feature selection (FS) method and classifier;
- *test set*: used in the test phase on unseen data.

The discovery set was used, by means of a nested *k*-fold cross-validation, to find the best predictive models, while the test set was adopted to validate the final model. This partitioning was performed 50 times by stratified sampling.

Table 1. T1-weighted three-dimensional spoiled gradient echo sequence parameters.

DCE-MRI Characteristic	Value
sequence type	DCE-MRI
series description	Ax VIBRANT mphase
scanner model	GE Signa HDxt (1.5T)
repetition time (TR)	37.720–56.920 ms
echo time (TE)	17.640–26.800 ms
flip angle	10°
matrix size	512X512 pixels
slice thickness	2–3 mm
spacing between slices	1–1.5 mm
pixel spacing	0.6875–0.7422 mm

Image processing

The overall workflow of this study is depicted in Figure 2. Each step in the image processing and analysis pipeline is described in the following sections.

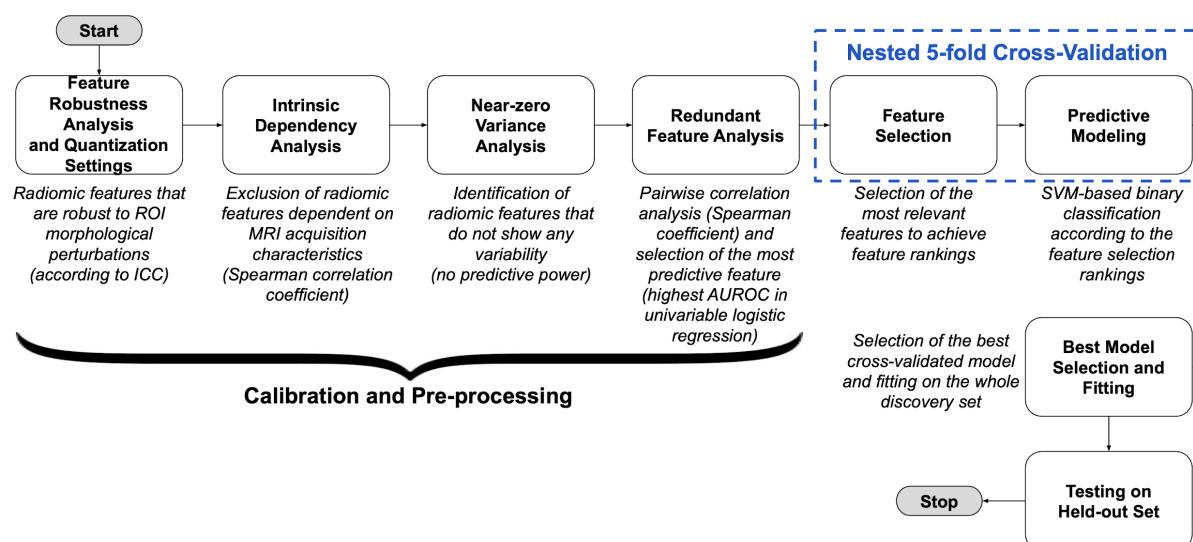


Figure 2. Overall workflow of the developed radiomic pipeline for breast cancer classification on MRI.

Lesion segmentation

For each mass with enhancement included in our study, the tumor volume of interest was determined by manual slice-by-slice segmentation on DCE-MRI images and performed by an breast radiologist, with more than 5-year experience in breast MRI imaging, in consensus with

a consultant breast radiologist (with more than 30-year experience in breast imaging). All the breast masses were manually segmented using a MatLab-coded custom tool [25]. The ROIs were delineated on the whole tumor on the DCE-MRI images with the strongest enhancement phase [26]. In particular, among the phases provided by the VIBRANT sequences (6 or 7, across different patients), the observer selected the one where the breast mass was more evident than in the background: on average, phase 3.66 ± 1.13 was chosen for segmenting the analyzed lesions.

The acquisition of DCE-MRI involves the administration of a contrast medium, which better depicts morphological/physiological characteristics of the tissues, where the examination includes various acquisitions in well-defined time intervals. Considering a specific position within the acquired volume, each voxel has a Time Intensity Curve, $TIC(t)$, reflecting the signal intensity variations due to the absorption/release of the contrast medium. The time course of the $TIC(t)$ curves can help clinicians to infer the type of lesion (e.g., benign *versus* malignant).

In Zhang *et al.* [27], only one time-point of the DCE-MRI series was chosen. The strongest enhancement phase can better reflect the tumor heterogeneity and invasiveness by relying upon the subtracted DCE-MRI images according to the literature [27, 28]. Therefore, our choice to analyze a single phase is consistent with the literature. In particular, relying upon the enhancement curves (with the goal of increasing the reproducibility), only the strongest enhancement phase [26, 27], was selected and analyzed.

Lesion segmentation was performed without including peritumoral tissues. The corresponding data were stored in the Neuroimaging Informatics Technology Initiative (NIfTI) format [9]. Figure 3 shows two examples of breast lesion segmentation.

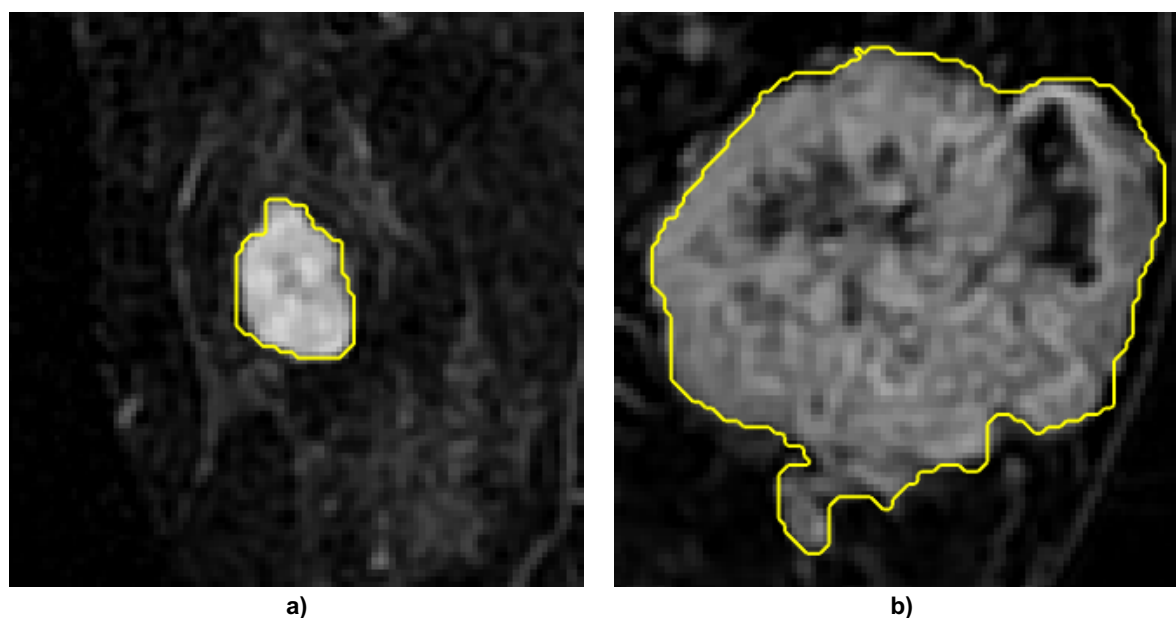


Figure 3. Two segmentation examples of breast lesions (axial view): **a)** benign lesion; **b)** malignant lesion. To improve the graphical representation, only a crop of the original 512×512 DCE-MRI image is shown. The images are displayed with the same zoom factor.

a) a 42-year old woman with a palpable lump in her left breast. MRI showed the presence, in the inferior outer quadrant, of a mass measuring lesion (20.5 mm in diameter), slightly hyperintense

in T2-Weighted (T2W) sequences, with strong post-contrast enhancement, with high wash-in rate (>200%) and type II enhancement curve. The lesion was proven to be a fibroadenoma.

- b) a 65-year old man with Klinefelter syndrome and gynecomastia, performed a breast MR local staging of a known breast cancer. Four lesions were found in the upper outer quadrant of the right breast, with inhomogeneous enhancement and a type III enhancement curve. The lesion was proven to be an invasive ductal carcinoma (ER: 100%; PgR: 5%, KI67: 80%. HER-2: 2+) and the patient underwent mastectomy.

Radiomic Feature Extraction

The features were extracted (from the 3D ROIs delineated in the previous step) using PyRadiomics, an open-source Python package developed for the standardization of radiomic feature extraction [29]. We used PyRadiomics version 2.0 and Python 3.7.5. Along with shape-based features, six feature classes were extracted: (1) first-order intensity histogram statistics, (2) Gray Level Co-occurrence Matrix features (GLCM) [30][31], (3) Gray Level Run Length Matrix (GLRLM) [32], (4) Gray Level Size Zone Matrix (GLSZM) [33], (5) Gray Level Dependence Matrix (GLDM) [34], and (6) Neighboring Gray Tone Difference Matrix (NGTDM) [35].

A well-established, practical rule of thumb states that at least 10 samples (i.e., patients) are needed for each feature in a model based on binary classifiers [12]. Indeed, due to the small sample size, we preferred to use only original features to avoid the processing of additional features extracted using convolutional image filters on the input medical images (e.g., Laplacian of Gaussian, logarithmic, exponential, gradient, wavelets). Moreover, clear guidelines for filtered versions of the images are not yet available, with the release of the Image Biomarker Standardisation Initiative (IBSI) [36] Chapter 2 still being under preparation (<https://theibsi.github.io/ibsi2/>).

Calibration and Pre-processing

Calibration and pre-processing steps were performed to identify a subset of features that are independent from the MRI acquisition parameters, informative, and non-redundant, by following the guidelines outlined in [13].

Supplementary Section S1 ‘Calibration and Pre-processing Details’ provides an in-depth description of the implemented pre-processing steps, while Section S2 ‘Radiomic Features’ describes the final set of analyzed features.

Feature Selection and Predictive Modeling

The aim of FS is to reduce the data dimensionality by selecting only a subset of features to create a model [37]. According to Supplementary Section S3 ‘Feature Selection Methods’, we considered these methods [38–40]:

- Unsupervised Discriminative Feature Selection (UDFS) [41];
- Dependence Guided Unsupervised Feature Selection (DGUFS) [42];
- Unsupervised Feature Selection with Ordinal Locality (UFSOL) [43].

The predictive modeling was performed by a Support Vector Machine (SVM) [44, 45] trained and tested using a nested 5-fold cross-validation (CV) procedure (Figure 4).

This discovery phase was aimed to find the FS methods with the best performance in combination with the SVM-based predictive modeling. The discovery set consisting of 89 patients was used and 50 repetitions of the nested 5-fold CV training were performed to obtain average model results (see Supplementary Section S4 ‘Nested k-fold Cross-Validation’).

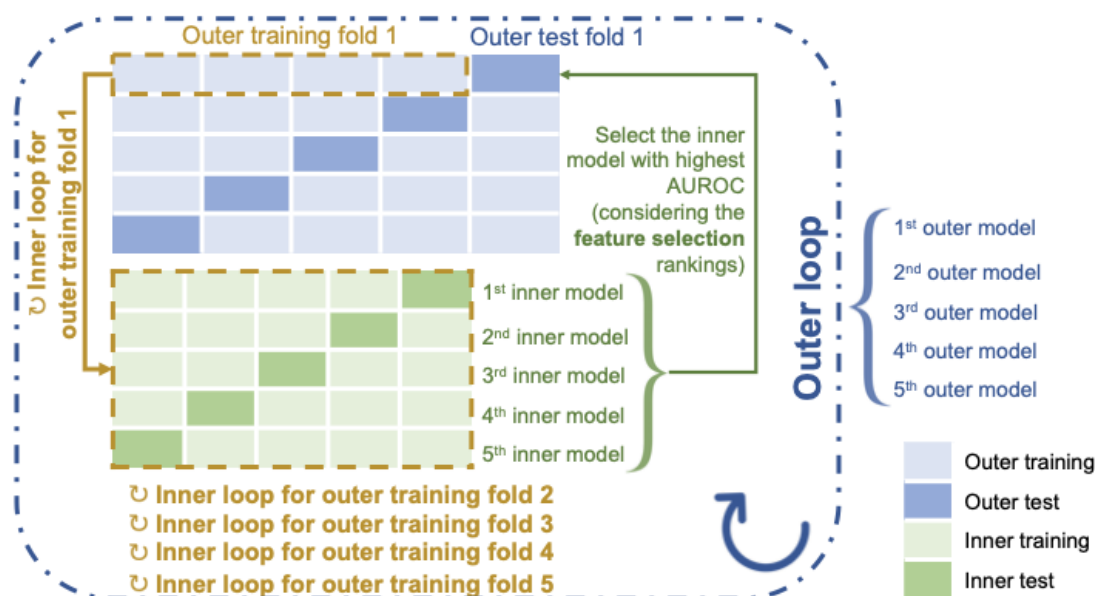


Figure 4. Nested k -fold cross-validation scheme (with $k=5$).

The used evaluation metrics were the AUROC, sensitivity and specificity, along with Positive Predictive Value (PPV) and Negative Predictive Value (NPV) to better investigate true positive and true negative results, respectively [46].

Relevant Feature Analysis and Testing on the Held-out Set

After the discovery phase, the most relevant features were selected according to the best three FS methods. With the goal of assessing the radiomic signature performance in the clinical practice, the predictive model was evaluated on a held-out test set (composed of unseen data after the hold-out splitting). In particular, the best three models identified in the discovery phase were retrained on the entire discovery set and tested on the test set. Similar to the discovery phase, 50 repetitions of the hold-out with stratified sampling were considered. See Supplementary Section S5 ‘Relevant Features and Radiomic Signatures’.

Statistical and Computational Analysis

All the statistical and computational analyses were performed using the MatLab[®] R2019b (64-bit version) environment (The MathWorks, Natick, MA, USA).

For feature robustness analyses, the two-way random-effects model (or mixed-effects), consistency, single rater/measurement, ICC(3,1) was used [47] (for the definition, see

Supplementary Section S1 ‘Calibration and Pre-processing Details’). The intrinsic dependency analysis made use of the Spearman correlation coefficient ($p < 0.0001$ as a cutoff), while the redundant feature analysis used the Spearman correlation coefficient for pairwise feature comparison ($\rho > 0.90$). No multiple-comparison correction was used to keep a reasonable number of features.

For distribution comparisons, the non-parametric Wilcoxon signed-rank test on paired samples was used, using a significance level of 0.05).

To realize the feature selection, the MatLab-coded ‘Feature Selection Library v6’ (<https://arxiv.org/abs/1607.01327>) was used [22–24]. The SVM models were based on the Statistics and Machine Learning Toolbox provided by MatLab.

Results

Experiments were aimed to quantify capabilities of the predictive models (FS method + classifier) in breast lesions characterization tasks. This section reports details about *i*) the models discovered and tested, *ii*) the radiomics signature obtained by each model, *iii*) the classification performance.

Study Population

Malignant lesions had a significantly (Mann-Whitney U test) higher volume at baseline compared to benign lesions (median (interquartile range): 3.532 (8.8034) cm³ vs. 0.692 (1.910) cm³; $p = 0.007$).

According to the SOR, 103/111 masses with enhancement were histologically characterized resulting in 57/111 benign (51.35%), 2/111 (1.80%) high-risk (with uncertain malignant potential (i.e., atypical ductal hyperplasia, ADH), and 52/111 malignant (46.85%) breast lesions. Predicting the upgrade from ADH to malignant lesions is still an open question in DCE-MRI radiomic studies [48]. As a matter of fact, In literature studies with long-term follow-up, atypical hyperplasia has been shown to have a 4× relative risk factor for future breast cancer [49]. Therefore, with the goal of obtaining a highly sensitive predictive model able to discriminate the patients into two classes for recall visits and diagnosis, we made a conservative choice for clinical purposes: lesions of uncertain malignant potential (i.e., ADH) were included in the class of malignant lesions since they are considered as high-risk lesions.

Furthermore, 8 benign breast lesions, with imaging characteristics suggesting the diagnosis of fibroadenoma, were stable in a follow-up of at least two years.

The final diagnosis of these lesions is described in Tables 2 and 3. Molecular subtypes were determined according to the St. Gallen International Expert Consensus 2013 [50].

Table 2. Histological features of the included malignant and high-risk breast lesions.

Histological Diagnosis	N	%
Invasive ductal carcinoma	29	53.7

Invasive lobular carcinoma	12	22.2
In situ ductal carcinoma with central necrosis	2	3.7
Atypical ductal hyperplasia	2	3.7
Low-grade intraductal carcinoma	5	9.3
Mucinous carcinoma	1	1.9
Tubular carcinoma	1	1.9
Malignant phylloides	1	1.9
Ductal carcinoma with mucinous aspects	1	1.9
Total	54	100
Tumor grade		
G1	5	9.2
G2	15	27.8
G3	34	63
Molecular subtype		
Luminal A	22	40.7
Luminal B	29	53.7
Triple negative	3	5.6

Table 3. Final diagnosis of the included benign breast lesions as by SOR.

Diagnosis	N	%
Fibroadenoma	21	36.8
Adenosis	12	21.1
Usual ductal hyperplasia	5	8.8
Intraductal papilloma	4	7
Apocrine metaplasia	4	7
Corpuscular cyst	3	5.3
Mastitis	2	3.5
Intramammary lymph node	2	3.5

Benign phylloides	1	1.8
Fibrosis	1	1.8
Sclerosing adenosis	1	1.8
Columnar cell hyperplasia	1	1.8
Total	57	100

DCE-MRI radiomics predicts lesion malignancy

Following the steps in Figure 2 (more details are provided in Section S1 ‘Calibration and Pre-processing Details’), the highest number of robust features was obtained using 16 bins in the gray level quantization (Supplementary Table S1). Therefore, out of the initial 107 features the number of highly robust features was 84. The intrinsic dependence analysis between these radiomic features and MRI acquisition parameters – based on a Spearman correlation analysis – showed 72 interdependent features. Afterwards, the near-zero variance analysis did not exclude any feature, while 19 non-redundant features were found. These steps are summarized in Supplementary Table S2.

These remaining 19 features were considered as the set to be fed to the FS methods to rank them and, successively, train/test the SVM. Supplementary Sections S3 ‘Feature Selection Methods’ provides the list of all FS methods, while Table S3 reports the final set of 19 features selected after the calibration and pre-processing phase.

Among the investigated FS methods, the ones obtaining better performance were: UDFS, DGUFS, and UFSOL. Table 4 provides the classification metrics obtained in the discovery phase by the best 3 predictive models (which used an SVM and the 3 FS methods previously described).

Table 4. Classification metrics obtained in the discovery phase by the predictive models exploiting the best 3 FS methods and the SVM. For each of these metrics, the mean value and standard deviation over all 50 repetitions were calculated. Boldface numbers denote the highest value for each metric.

Predictive Model	AUROC	Sensitivity	Specificity	PPV	NPV
<i>UDFS + SVM</i>	0.557±0.07	0.622±0.17	0.493±0.184	0.552±0.107	0.452±0.145
<i>DGUFS + SVM</i>	0.647±0.058	0.722±0.089	0.574±0.148	0.652±0.059	0.573±0.153
<i>UFSOL + SVM</i>	0.597±0.053	0.778±0.094	0.426±0.107	0.567±0.062	0.599±0.124

After the choice of the best FS and fitted model the relevant features were used to retrain the SVM model and perform hold-out testing. The final signatures for each of the three best FS methods – which were used in the external test – were composed of the relevant radiomic features listed in Figure 5. Table 5 provides the classification metrics obtained by the best

three predictive models in the hold-out testing phase cohort (see Supplementary Material S5 'Relevant Features and Radiomic Signatures').

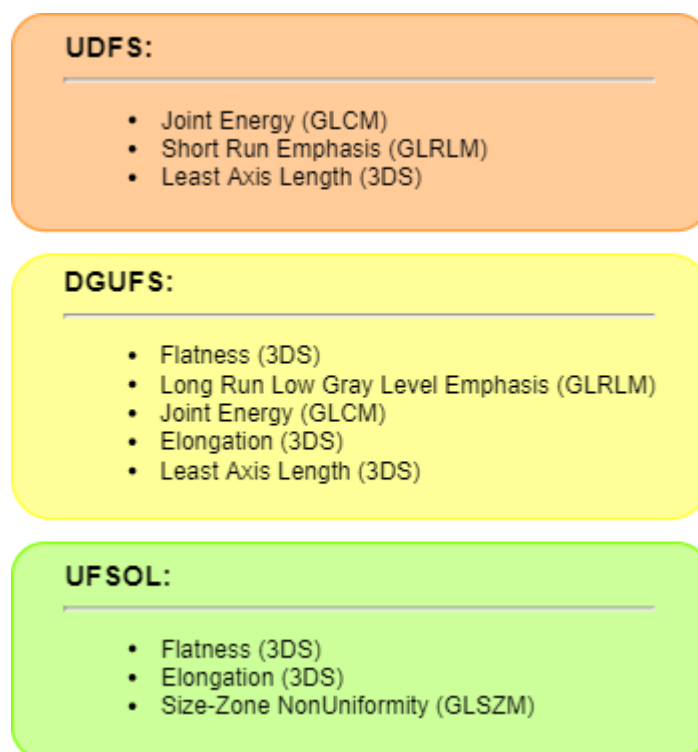


Figure 5. Final radiomic signatures for each of the three best FS methods, in the external test phase.

Table 5. Classification metrics obtained in the hold-out testing phase by the predictive models exploiting the 3 FS methods and SVM. For each of these metrics, the mean value and standard deviation over all 50 repetitions were calculated. Boldface numbers denote the highest value for each metric.

Predictive Model	AUROC	Sensitivity	Specificity	PPV	NPV
<i>UDFS + SVM</i>	0.725±0.091	0.709±0.176	0.741±0.114	0.72±0.093	0.75±0.114
<i>DGUFS + SVM</i>	0.723±0.086	0.705±0.173	0.741±0.11	0.719±0.088	0.747±0.11
<i>UFSOL + SVM</i>	0.708±0.089	0.796±0.128	0.62±0.144	0.669±0.098	0.777±0.125

Discussion

Various breast cancer predictive radiomic models were built-up by using different quantitative radiomic features extracted from MRI sequences, showing promising results, with the goal of predicting the lesion malignancy in a non-invasive way [11]. Nevertheless, different breast MRI protocols, as well as a wide spectrum of lesion segmentation and feature extraction methods, have been proposed so far. DCE-MRI is the most common MRI technique used to characterize breast lesions relying upon both morphologic and hemodynamic features [51–53]. Although circumscribed masses are generally suggestive of benign lesions and non-circumscribed masses are suspicious for carcinoma, margin analysis is highly dependent on the spatial

resolution, thus low spatial resolution scans could affect margin evaluation, in particular for small masses. Furthermore, although benign lesions generally follow persistent curves, as well as many malignant lesions follow “wash-out” curves, a “plateau” curve can be observed with both benign and malignant lesions; therefore, there is often an overlap between the kinetic curves that depict malignant and benign lesions [24].

In this study, we developed a predictive model based on 3D radiomic features extracted from DCE-MRI sequences (using only the strongest enhancement phase), focusing on breast masses with enhancement, aiming at predicting breast lesion malignancy. After careful pre-processing steps (including feature robustness and redundancy analyses), we used the remaining features to build radiomic signatures, investigating three feature-ranking methods (UDFS, DGUFS and UFSOL) and one classifier (SVM). Good performance for both ranking methods and classifiers used in our work were reported in previous studies [54, 55]. Three radiomic signatures were defined, with the best performance in terms of AUROC and specificity achieved by the FS method UDFS coupled with the SVM classifier (0.725 ± 0.091 and 0.741 ± 0.114 , respectively), while the highest sensitivity and NPV values was obtained by UFSOL+SVM (0.796 ± 0.128 and 0.777 ± 0.125 , respectively). All MRI series involved in our study were acquired at the same institution with the same imaging protocol, with the result of reducing the variability in image acquisitions. We analyzed a homogeneous dataset that allowed us to carefully process and assess the extracted radiomic features, thus increasing the result reliability, as well as the model generalization on external patient cohorts. Despite the limited number of clinical cases, an in-depth and careful analysis was proposed to build radiomic models that predict malignant lesions in patients with breast cancer. Our study was conducted in accordance with [56], where the generalization abilities of radiomic models in multi-centric MRI datasets have been recently addressed. It is worth noting that the use of breast lesion classification based on the ACR BI-RADS lexicon also facilitated the sample standardization of the enrolled cases [8].

Considering that we analyzed only the strongest DCE-MRI phase, thus enabling an efficient and less protocol-dependent approach compared to DCE pharmacokinetic modeling, our results are consistent with recent literature studies. In particular, Whitney *et al.* [57] achieved an average AUROC of 0.846 (95% confidence interval: 0.808–0.875) by using FS from all the features and exploiting pharmacokinetic modeling [58]. Nevertheless, the multiparametric MRI-based radiomic model developed by Zhang *et al.* [59] demonstrated higher diagnostic ability for differentiating benign and malignant breast lesions (AUROC=0.921), increasing the discriminating power of radiomic features extracted from DCE pharmacokinetic parameter maps alone (AUROC=0.836), as well as the IsoSVM Classification model proposed by Parekh *et al.* [60], which also included radiomic feature mapping of T2w, T1w, DWI, and ADC map images. These two works (i.e., [59],[60]) analyzed different types of MR sequences, thus introducing further processing and dependence on the MRI acquisition protocols. Unlike these literature works, where multiple sequences were used, our main objective was to develop a radiomics-powered model to be integrated into the clinical practice by exploiting standard-of-care MRI and reducing the required image pre-processing (i.e., normalization and quantitative imaging map generation). Therefore, considering standard-of-care imaging only, the dependency on the scanner/protocols is minimized and the full potential of the strongest enhancement phase DCE-MRI (as suggested by Teruel *et al.* [28] for clinical/pathological response prediction) was drawn by relying upon robust radiomics analyses. By doing so, our model might be suitably deployed onto other institutions without requiring specific pre-

processing. As a matter of fact, our radiomic model achieved an AUROC of 0.725 ± 0.091 comparable to the results presented in Zhang *et al.* [59] for the T1-weighted imaging input (AUROC=0.730) in the differentiation of benign and malignant breast lesions using SVMs. It is worth noting that a direct comparison between the two contributions is not fair since the analyzed datasets are different; moreover, we used a conservative training/testing strategy aimed at future multicenter studies (i.e., a nested cross-validation as recently shown in [56]) and several feature selection methods to avoid over-optimistic results despite the relatively limited sample size.

The combination of radiomic features extracted from different types of sequences could increase the predictive power of the model developed for the problem at hand. Our work primarily aims at clinical feasibility and, for this reason, we analyzed only the subtracted images (obtained from two DCE-MRI time-points), which currently represent the routine examination in breast cancer patients [61, 62].

Other recent studies demonstrated that peritumoral tissue inclusion during segmentation led to higher accuracy compared to tumor alone [63]. Our analysis focused on extracting radiomic features from a single post-contrast phase. Furthermore, providing the classifier with supplementary information derived from texture features of the first and later post-contrast phases in a dynamic manner, could potentially improve the performance of lesion characterization, as determined in other studies [64]. Analyzing features selected in the three optimal radiomic signatures built-up in our study, 3DS features (*least axis length*, *flatness* and *elongation*) were prevalent and occurred in all signatures, *joint energy* was the only GLCM feature selected by 2 out of 3 radiomic models, whereas *Short Run Emphasis* (SRE) and *Long Run Low Gray Level Emphasis* (LRLGLE) were the two GLRLM features selected by UDFS and DGUFS, respectively. Therefore, *Size-Zone NonUniformity*, a GLSZM feature, was employed only in the UFSOL-derived signature.

According to previous research, our results confirm that morphological features, reflecting information about the whole lesion shape, may be used for differential diagnosis in radiomic models [65–67]. On the contrary, in the best performing radiomic models built-up in other studies, such as in [59], shape features accounted for a small proportion of the features in all models, and they were not selected in some models.

Joint Energy, a measure of homogeneous patterns in the image, is a GLCM feature selected by two of our optimal radiomic models. Early work utilized parameters calculated from GLCM to discriminate between benign and malignant lesions [68, 69]. The benign or malignant nature of a lesion can be also inferred *via* its homogeneous or heterogeneous enhancement appearance, and prior studies have shown that such GLCM-derived features can be used to characterize breast lesions with high diagnostic accuracy [69–71].

In our study, among GLRLM features, LRLGLE and SRE were included in two radiomic signatures. By evaluating the robustness of radiomic features in MRI, Cattell *et al.* [72] showed that GLRLM features were found to have moderate robustness (0.5–0.9). Moreover, in the study conducted by Gibbs *et al.* [68], both these GLRLM features, LRLGLE and SRE, demonstrated significant differences in the differentiation between benign and malignant lesions.

According to a very recent meta-analysis conducted by Zhu *et al.* [73], the characterization of breast masses on DCE-MRI alone showed high sensitivity and AUROC (0.95 and 0.92, respectively) whereas specificity remained lower (0.71). Radiomic models could be capable of overcoming this issue: two of our radiomic signatures showed slightly better performance in terms of specificity, thus further efforts are needed to optimize these results.

Considering also the recent increased evidence of deposition of gadolinium in the brain [74], the development of contrast-free examinations, even of comparable accuracy, appears to be highly attractive. The combination of breast multiparametric MRI radiomics with unenhanced breast MRI protocols is attractive, as it may develop into a tool to decrease user-dependence of interpretation. In particular, quantitative DWI showed a higher specificity to differentiate between benign and malignant breast lesions compared to DCE-MRI [75, 76].

The limitations of our study are listed in what follows:

- the analyzed dataset was collected from a single center, and an additional external validation of the proposed model is required;
- the lesion ROIs were manually annotated by a single observer, and then checked by an experienced reader, helping to further increase the reliability of the delineation process;
- this was a patient-driven rather than a lesion-driven study: in patients with multi-focal breast lesions only the largest lesion was considered. We assumed that, enrolling multiple lesions of the same patient as separate lesions, a practice often used to increase the available samples, would have introduced potentially a redundancy of data (belonging to the same person) into the dataset, from which the predictive model must then be built, affecting the capabilities of the model.

As future directions, we are planning to extend our single-center study to a larger patient cohort and also to non-mass enhancement breast lesions. Moreover, a multicentric study would allow us to assess the generalization abilities of the developed predictive model. Lastly, prospective studies are required to explore how these radiomics-powered predictive models could be deployed into the clinical practice for the characterization of breast masses in DCE-MRI. Complementing studies on BI-RADS 4 and 5 characterization [77, 78], analyzing BI-RADS 3 lesions by an *ad-hoc* method would be clinically relevant to avoid unnecessary biopsies and resulting in an optimization of the current patient management.

Acknowledgments

BLINDED

References

1. International Agency for Research on Cancer. (2020) The Global Cancer Observatory. World Health Organization.
2. Brennan S, Liberman L, Dershaw DD, Morris E (2010) Breast MRI screening of women with a personal history of breast cancer. *AJR Am J Roentgenol* 195:510–516
3. Girometti R, Tomkova L, Cereser L, Zuiani C (2018) Breast cancer staging: Combined digital breast tomosynthesis and automated breast ultrasound versus magnetic resonance imaging. *Eur J Radiol* 107:188–195

4. Tagliafico AS, Piana M, Schenone D, et al (2020) Overview of radiomics in breast cancer diagnosis and prognostication. *Breast* 49:74–80
5. Ye D-M, Wang H-T, Yu T (2020) The Application of Radiomics in Breast MRI: A Review. *Technol Cancer Res Treat* 19:1533033820916191
6. Bickelhaupt S, Paech D, Kickingereeder P, et al (2017) Prediction of malignancy by a radiomic signature from contrast agent-free diffusion MRI in suspicious breast lesions found on screening mammography. *J Magn Reson Imaging* 46:604–616
7. Chen W, Giger ML, Bick U (2006) A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. *Acad Radiol* 13:63–72
8. Ghazala S (2016) Characterization of Suspicious Breast Lesions with Dynamic Contrast enhanced MRI In comparison to Conventional mammography and Ultrasonography. *JCPCR* 4.: <https://doi.org/10.15406/jcpcr.2016.04.00121>
9. Rogers W, Thulasi Seetha S, Refaee TAG, et al (2020) Radiomics: from qualitative to quantitative imaging. *Br J Radiol* 93:20190948
10. Conti A, Duggento A, Indovina I, et al (2020) Radiomics in breast cancer classification and prediction. *Semin Cancer Biol*. <https://doi.org/10.1016/j.semcancer.2020.04.002>
11. Orlando A, Dimarco M, Cannella R, Bartolotta TV (2020) Breast dynamic contrast-enhanced-magnetic resonance imaging and radiomics: State of art. *Artificial Intelligence in Medical Imaging* 1:6–18
12. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 278:563–577
13. Papanikolaou N, Matos C, Koh DM (2020) How to develop a meaningful radiomic signature for clinical use in oncologic patients. *Cancer Imaging* 20:33
14. Yankeelov TE, Mankoff DA, Schwartz LH, et al (2016) Quantitative Imaging in Cancer Clinical Trials. *Clin Cancer Res* 22:284–290
15. Rundo L, Militello C, Vitabile S, et al (2019) A survey on nature-inspired medical image analysis: A step further in biomedical data integration. *Fund Inform* 171:345–365
16. Granzier RWY, Verbakel NMH, Ibrahim A, et al (2020) MRI-based radiomics in breast cancer: feature robustness with respect to inter-observer segmentation variability. *Sci Rep* 10:14163
17. Escudero Sanchez L, Rundo L, Gill AB, et al (2021) Robustness of radiomic features in CT images with different slice thickness, comparing liver tumour and muscle. *Sci Rep* 11:8262
18. Raunig DL, McShane LM, Pennello G, et al (2015) Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res* 24:27–67
19. O'Connor JPB, Aboagye EO, Adams JE, et al (2017) Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 14:169–186
20. Zwanenburg A, Vallières M, Abdalah MA, et al (2020) The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput

Image-based Phenotyping. *Radiology* 295:328–338

21. Tomaszewski MR, Gillies RJ (2021) The Biological Meaning of Radiomic Features. *Radiology* 299:E256
22. Martin-Gonzalez P, Crispin-Ortuzar M, Rundo L, et al (2020) Integrative radiogenomics for virtual biopsy and treatment monitoring in ovarian cancer. *Insights Imaging* 11:94
23. Lambin P, Leijenaar RTH, Deist TM, et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762
24. Morris EA, Comstock CE, Lee CH, et al. (2013) ACR BI-RADS Magnetic Resonance Imaging
25. Militello C, Rundo L, Dimarco M, et al (2021) Semi-automated and interactive segmentation of contrast-enhancing masses on breast DCE-MRI using spatial fuzzy clustering. *Biomed Signal Process Control* (Accepted, In press)
26. Liu J, Sun D, Chen L, et al (2019) Radiomics Analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging for the Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. *Front Oncol* 9:980
27. Zhang Y, Chan S, Park VY, et al (2020) Automatic Detection and Segmentation of Breast Cancer on MRI Using Mask R-CNN Trained on Non-Fat-Sat Images and Tested on Fat-Sat Images. *Acad Radiol*. <https://doi.org/10.1016/j.acra.2020.12.001>
28. Teruel JR, Heldahl MG, Goa PE, et al (2014) Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *NMR Biomed* 27:887–896
29. van Griethuysen JJM, Fedorov A, Parmar C, et al (2017) Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 77:e104–e107
30. Haralick RM, Shanmugam K, Dinstein I 'hak (1973) Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics SMC-3*:610–621
31. Haralick RM (1979) Statistical and structural approaches to texture. *Proceedings of the IEEE* 67:786–804
32. Galloway MM (1975) Texture analysis using gray level run lengths. *Computer Graphics and Image Processing* 4:172–179
33. Thibault G, Angulo J, Meyer F (2014) Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Trans Biomed Eng* 61:630–637
34. Sun C, Wee WG (1983) Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing* 23:341–352
35. Amadasun M, King R (1989) Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 19:1264–1274
36. Zwanenburg A, Abdalah MA, Apte A, et al (2018) PO-0981: Results from the Image Biomarker Standardisation Initiative. *Radiotherapy and Oncology* 127:S543–S544
37. Hastie, T., Tibshirani, R., & Friedman, J. (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer

38. Roffo G, Melzi S, Cristani M (2015) Infinite Feature Selection. In: 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, pp 4202–4210
39. Roffo G, Melzi S, Castellani U, Vinciarelli A (2017) Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach. 2017 IEEE International Conference on Computer Vision (ICCV)
40. Roffo G, Melzi S (2017) Ranking to Learn: New Frontiers in Mining Complex Patterns 19–35
41. Yang Y, Shen HT, Ma Z, et al L2,1-Norm Regularized Discriminative Feature Selection for Unsupervised Learning. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI-11). AAAI Press, pp 1589–1594
42. Guo J, Zhu W (2018) Dependence Guided Unsupervised Feature Selection. In: Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). pp 2232–2239
43. Guo J, Guo Y, Kong X, He R (2017) Unsupervised feature selection with ordinal locality. 2017 IEEE International Conference on Multimedia and Expo (ICME)
44. Boser BE, Guyon IM, Vapnik VN (1992) Support Vector Machines. Springer Reference
45. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory - COLT '92. ACM Press, New York, New York, USA, pp 144–152
46. Altman DG, Bland JM (1994) Statistics Notes: Diagnostic tests 2: predictive values. *BMJ* 309:102–102
47. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86:420–428
48. Lo Gullo R, Vincenti K, Rossi Saccarelli C, et al (2021) Diagnostic value of radiomics and machine learning with dynamic contrast-enhanced magnetic resonance imaging for patients with atypical ductal hyperplasia in predicting malignant upgrade. *Breast Cancer Res Treat* 187:535–545
49. Hartmann LC, Degnim AC, Santen RJ, et al (2015) Atypical hyperplasia of the breast--risk assessment and management options. *N Engl J Med* 372:78–89
50. Goldhirsch A, Winer EP, Coates AS, et al (2013) Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* 24:2206–2223
51. Wedegärtner U, Bick U, Wörtler K, et al (2001) Differentiation between benign and malignant findings on MR-mammography: usefulness of morphological criteria. *Eur Radiol* 11:1645–1650
52. Baum F, Fischer U, Vosshenrich R, Grabbe E (2002) Classification of hypervascularized lesions in CE MR imaging of the breast. *Eur Radiol* 12:1087–1092
53. Szabó BK, Aspelin P, Wiberg MK, Boné B (2003) Dynamic MR imaging of the breast. Analysis of kinetic and morphologic diagnostic criteria. *Acta radiol* 44:379–386
54. Tian L, Zhang D, Bao S, et al (2021) Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas. *Clin Radiol* 76:158.e19–

158.e25

55. Xu W, Hao D, Hou F, et al (2020) Soft Tissue Sarcoma: Preoperative MRI-Based Radiomics and Machine Learning May Be Accurate Predictors of Histopathologic Grade. *AJR Am J Roentgenol* 215:963–969
56. Doran SJ, Kumar S, Orton M, et al (2021) “Real-world” radiomics from multi-vendor MRI: an original retrospective study on the prediction of nodal status and disease survival in breast cancer, as an exemplar to promote discussion of the wider issues. *Cancer Imaging* 21:37
57. Whitney HM, Taylor NS, Drukker K, et al (2019) Additive Benefit of Radiomics Over Size Alone in the Distinction Between Benign Lesions and Luminal A Cancers on a Large Clinical Breast MRI Dataset. *Acad Radiol* 26:202–209
58. Chen W, Giger ML, Lan L, Bick U (2004) Computerized interpretation of breast MRI: investigation of enhancement-variance dynamics. *Med Phys* 31:1076–1082
59. Zhang Q, Peng Y, Liu W, et al (2020) Radiomics Based on Multimodal MRI for the Differential Diagnosis of Benign and Malignant Breast Lesions. *J Magn Reson Imaging* 52:596–607
60. Parekh VS, Jacobs MA (2017) Integrated radiomic framework for breast cancer and tumor biology using advanced machine learning and multiparametric MRI. *NPJ Breast Cancer* 3:43
61. Kuhl C (2007) The Current Status of Breast MR Imaging Part I. Choice of Technique, Image Interpretation, Diagnostic Accuracy, and Transfer to Clinical Practice. *Radiology* 244:356–378
62. Yeh ED (2011) Breast magnetic resonance imaging: current clinical indications. *Obstet Gynecol Clin North Am* 38:159–77, ix
63. Zhou J, Zhang Y, Chang K-T, et al (2020) Diagnosis of Benign and Malignant Breast Lesions on DCE-MRI by Using Radiomics and Deep Learning With Consideration of Peritumor Tissue. *J Magn Reson Imaging* 51:798–809
64. Nagarajan MB, Huber MB, Schlossbauer T, et al (2013) Classification of Small Lesions in Breast MRI: Evaluating The Role of Dynamically Extracted Texture Features Through Feature Selection. *J Med Biol Eng* 33.: <https://doi.org/10.5405/jmbe.1183>
65. Tzacheva AA, Najarian K, Brockway JP (2003) Breast cancer detection in gadolinium-enhanced MR images by static region descriptors and neural networks. *J Magn Reson Imaging* 17:337–342
66. Gilhuijs KGA, Giger ML, Bick U (1998) Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Medical Physics* 25:1647–1654
67. Liney GP, Sreenivas M, Gibbs P, et al (2006) Breast lesion analysis of shape technique: semiautomated vs. manual morphological description. *J Magn Reson Imaging* 23:493–498
68. Gibbs P, Onishi N, Sadinski M, et al (2019) Characterization of Sub-1 cm Breast Lesions Using Radiomics Analysis. *J Magn Reson Imaging* 50:1468–1477
69. Gibbs P, Turnbull LW (2003) Textural analysis of contrast-enhanced MR images of the breast. *Magn Reson Med* 50:92–98

70. Chen W, Giger ML, Li H, et al (2007) Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magn Reson Med* 58:562–571
71. Nie K, Chen J-H, Yu HJ, et al (2008) Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. *Acad Radiol* 15:1513–1525
72. Cattell R, Chen S, Huang C (2019) Robustness of radiomic features in magnetic resonance imaging: review and a phantom study. *Visual Computing for Industry, Biomedicine, and Art* 2
73. Zhu C-R, Chen K-Y, Li P, et al (2020) Accuracy of multiparametric MRI in distinguishing the breast malignant lesions from benign lesions: a meta-analysis. *Acta radiol* 284185120963900
74. Guo BJ, Yang ZL, Zhang LJ (2018) Gadolinium Deposition in Brain: Current Scientific Evidence and Future Perspectives. *Front Mol Neurosci* 11:335
75. Chen X, Li W-L, Zhang Y-L, et al (2010) Meta-analysis of quantitative diffusion-weighted MR imaging in the differential diagnosis of breast lesions. *BMC Cancer* 10:693
76. Baxter GC, Graves MJ, Gilbert FJ, Patterson AJ (2019) A Meta-analysis of the Diagnostic Performance of Diffusion MRI for Breast Lesion Characterization. *Radiology* 291:632–641
77. Luo W-Q, Huang Q-X, Huang X-W, et al (2019) Predicting Breast Cancer in Breast Imaging Reporting and Data System (BI-RADS) Ultrasound Category 4 or 5 Lesions: A Nomogram Combining Radiomics and BI-RADS. *Sci Rep* 9:11921
78. Hao W, Gong J, Wang S, et al (2020) Application of MRI Radiomics-Based Machine Learning Model to Improve Contralateral BI-RADS 4 Lesion Assessment. *Front Oncol* 10:531476

3D DCE-MRI Radiomic Analysis for Malignant Lesion Prediction in Breast Cancer Patients

Acknowledgements

This study has received funding by the GeSeTon project (Italian MISE Grant No. 489 of 21/02/2018).

This study has also been partially supported by The Mark Foundation for Cancer Research and Cancer Research UK Cambridge Centre [C9685/A25177] and by the Royal Society for the International Exchanges 2020 Cost Share with the Italian CNR (project No. IEC/R2/202313).

Additional support was also provided by the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.