Multi-marker quantitative radiomics for mass characterization in dedicated breast CT imaging

Marco Caballo
Department of Medical Imaging, Radboud University Medical Center, PO Box 9101, Nijmegen 6500 HB, The Netherlands

Domenico R. Pangallo
Department of Medical Imaging, Radboud University Medical Center, PO Box 9101, Nijmegen 6500 HB, The Netherlands
Biolab, Department of Electronics and Telecommunication, Politecnico di Torino, Torino 10129, Italy

Wendelien Sanderink
Department of Medical Imaging, Radboud University Medical Center, PO Box 9101, Nijmegen 6500 HB, The Netherlands

Andrew M. Hernandez
Department of Radiology, University of California Davis, Sacramento, CA 95817, USA

Su Hyun Lyu
Department of Biomedical Engineering, University of California Davis, Sacramento, CA 95817, USA

Filippo Molinari
Biolab, Department of Electronics and Telecommunication, Politecnico di Torino, Torino 10129, Italy

John M. Boone
Department of Radiology, University of California Davis, Sacramento, CA 95817, USA
Department of Biomedical Engineering, University of California Davis, Sacramento, CA 95817, USA

Ritse M. Mann
Department of Medical Imaging, Radboud University Medical Center, PO Box 9101, Nijmegen 6500 HB, The Netherlands

Ioannis Sechopoulos[a]
Department of Medical Imaging, Radboud University Medical Center, PO Box 9101, Nijmegen 6500 HB, The Netherlands
Dutch Expert Center for Screening (LRCB), PO Box 6873, Nijmegen 6503 GJ, The Netherlands

Purpose: To develop and evaluate the diagnostic performance of an algorithm for multi-marker radiomic-based classification of breast masses in dedicated breast computed tomography (bCT) images.

Methods: Over 1000 radiomic descriptors aimed at quantifying mass and border heterogeneity, morphology, and margin sharpness were developed and implemented. These included well-established texture and shape feature descriptors, which were supplemented with additional approaches for contour irregularity quantification, spicule and lobe detection, characterization of degree of infiltration, and differences in peritumoral compartments. All descriptors were extracted from a training set of 202 bCT masses (133 benign and 69 malignant), and their individual diagnostic performance was investigated in terms of area under the receiver operating characteristics (ROC) curve (AUC) of single-feature-based linear discriminant analysis (LDA) classifiers. Subsequently, the most relevant descriptors were selected through a multiple-step feature selection process (including stability analysis, statistical significance, evaluation of feature interaction, and dimensionality reduction), and used to develop a final LDA radiomic model for classification of benign and malignant masses, which was then tested on an independent test set of 82 cases (45 benign and 37 malignant).

Results: The majority of the individual radiomic descriptors showed, on the training set, an AUC value deriving from a linear decision boundary higher than 0.65, with the lower limit of the associated 95% confidence interval (C.I.) not overlapping with random chance (AUC = 0.5). The final LDA radiomic model resulted in a test set AUC of 0.90 (95% C.I. 0.80–0.96).

Conclusions: The proposed multi-marker radiomic approach achieved high diagnostic accuracy in bCT mass classification, using a radiomic signature based on different feature types. While future studies with larger datasets are needed to further validate these results, quantitative radiomics applied to bCT shows potential to improve the breast cancer diagnosis pipeline. © *2020 The Authors. Medical Physics* published by Wiley Periodicals LLC on behalf of American Association of Physicists in Medicine [https://doi.org/10.1002/mp.14610]

Key words: breast cancer, breast CT, computer-aided diagnosis, precision medicine, radiomics

1. INTRODUCTION

Radiomics is a growing research field that aims at the extraction of relevant information from medical images through computerized image analysis methods.[1] The motivation behind radiomics is driven by the hypothesis that physiological and pathological conditions (e.g., cancer) imprint different types of information on radiographic images, and that this information can be quantified and used to develop mathematical models for clinical decision support.[1]

Radiomics is being investigated in all areas of oncologic imaging, especially for lung,[2–4] prostate,[5] and breast tumors,[6] since they are the most commonly diagnosed cancers.[7] Especially in the case of breast cancer, several studies have been reported in literature that extract imaging biomarkers and develop diagnostic classification models from different breast imaging modalities.[8–11] These biomarkers can usually be divided into two major categories: texture and shape descriptors. The former are traditionally extracted inside the region of interest containing the prognostic value being investigated (usually breast masses)[12–17] after the boundary of the region is identified through manual or automatic segmentation. These texture biomarkers aim to quantify the brightness, contrast, and heterogeneity of the region. In comparison, shape descriptors are usually calculated on a binarized mask representing the segmented mass boundary, and aim at quantifying morphological aspects of the region in terms of size and contour irregularities.[18]

Although these categories contain the most common radiomic descriptor types reported up to now in literature, some other studies have investigated the use of these and other radiomic biomarkers in different ways, aiming to capture additional characteristics of tumor masses. For example, some investigators have performed the radiomic analysis on the mass periphery, relating the information extracted from the mass margin to the tumor phenotype. This type of analysis was conducted on breast masses acquired with different imaging modalities, such as mammography,[19,20] digital breast tomosynthesis,[21] breast MRI,[22,23] and ultrasound,[24,25] with the aim of quantifying the degree of spiculation through margin sharpness and radial gradient analysis,[19,22–24] the texture and echogenicity along the mass boundary, and the diversity in average intensity values over different margin regions.[21,24]

Therefore, several advancements are being pursued in radiomics, both in the development of new descriptors and algorithms, and in their application to different medical imaging modalities. In the same vein, in this study a multi-marker radiomic algorithm able to capture different tumor characteristics was developed, and applied to diagnose breast masses imaged with dedicated breast computed tomography (bCT).[26] The algorithm includes several radiomic descriptors from different categories, including novel approaches to quantify morphology, degree of infiltration, and texture differences in peritumoral compartments. This multi-marker radiomic analysis, combined with the optimized contrast and resolution characteristics of bCT, may help provide a strong quantification of imaging biomarkers from breast masses, potentially leading to a better characterization of breast tumors.

2. MATERIALS AND METHODS

In this section, the proposed radiomic pipeline is described, including image acquisition protocols, patient dataset, image preprocessing and segmentation, and radiomic feature development and implementation. The developed radiomic features can be divided into three macro-groups (Fig. 1): mass and border texture (Section 2.E), shape and contour (Section 2.F), and margin (Section 2.G) descriptors. Mathematical details and biological motivation are reported for each feature group, along with additional testing on phantom images for each newly developed biomarker. In the last two sections, the methods used to analyze the extracted features, and the developed radiomic-based classification model for breast mass classification into benign and malignant cases, are reported.

2.A. Image acquisition protocol

The image dataset used in this study was acquired with bCT systems from two different institutions. One set of images was acquired using the first- and second-generation bCT prototype systems designed and developed at the University of California (UC), Davis (California, USA) for use in clinical studies under several IRB-approved protocols.[26,27] Both scanner prototypes house a continuous output x-ray tube (Comet, Flamatt, Switzerland), with a nominal 0.4 mm focal spot size, and an 80 kV spectrum with 0.2–0.3 mm Cu filtration. A total of 500 projections were acquired over a 360° scan using a Paxscan 4030CB flat panel detector (Varian Medical Systems, Palo Alto, California, USA) operating at 30 fps (approximately 17 s total scan time) in 2×2 binning mode with dynamic gain. All projection images were reconstructed using a variation of the Feldkamp-filtered backprojection algorithm (with a Shepp-Logan kernel) with an isotropic voxel size of 0.38 mm, and corrected for shading artifacts using a maximum likelihood polynomial fitting approach in the reconstruction space.[28] The tube current was adjusted for each patient scan based on breast size and mammographic density, resulting in a mean glandular dose of approximately 6.0 mGy.

The second set of images was acquired using a clinical bCT prototype installed at Radboud University Medical Center (Nijmegen, the Netherlands). The system, of a similar half cone-beam geometry as those at UC Davis, has an x-ray tube with a tungsten target and aluminum filter, a 0.3 mm nominal focal spot, and a fixed tube voltage set to 49 kV, with the resulting x-ray spectrum having a first half value layer of 1.39 mm Al. The detector is the same 4030CB as that used in the UC Davis systems. The source-to-imager distance is 92.3 cm while the source-to-isocenter distance is 65 cm. The x-ray tube operates in pulsed mode, with a constant 8 ms
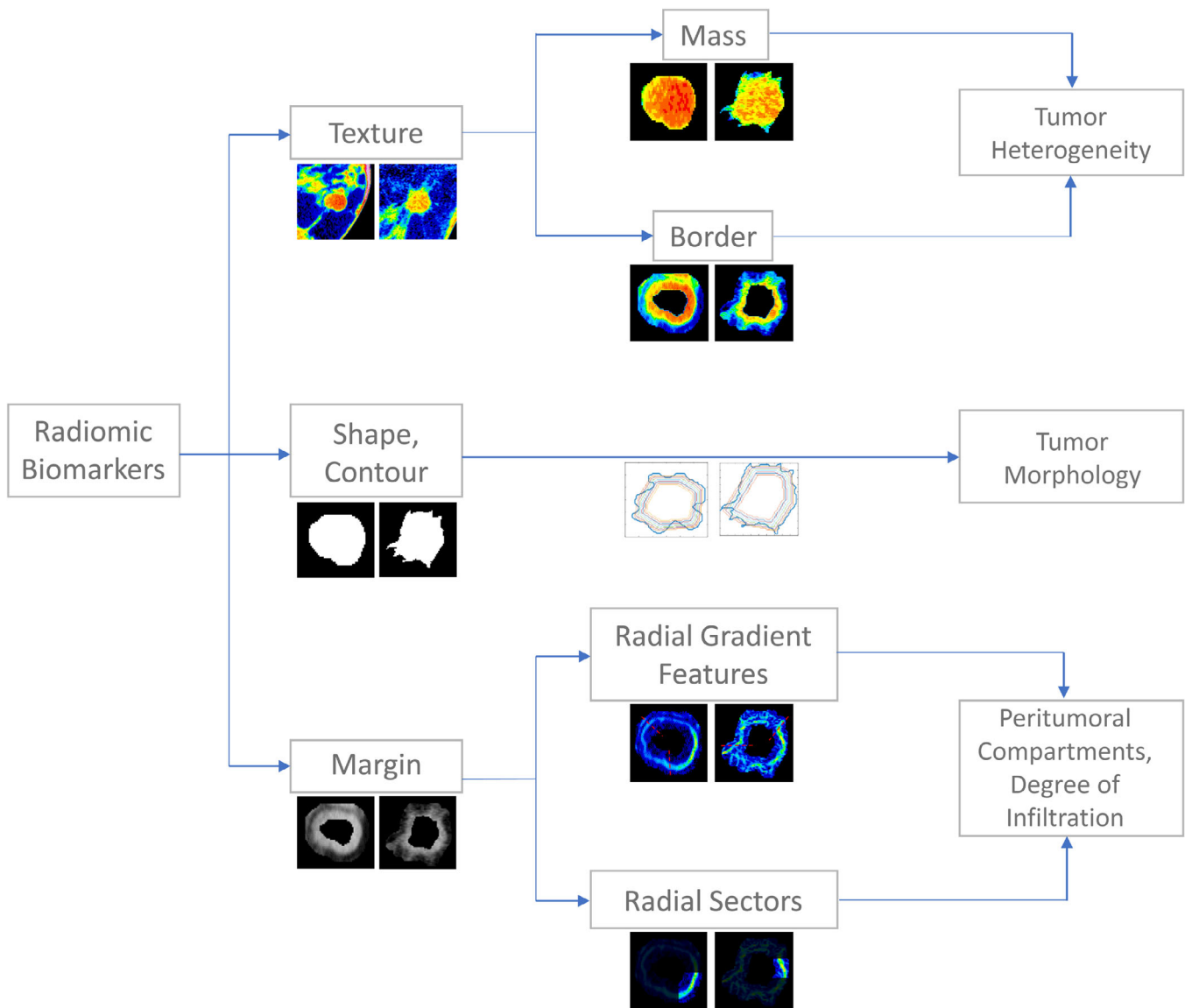
Fig. 1. Scheme of the radiomic feature descriptors described in this study. [Color figure can be viewed at wileyonlinelibrary.com]

pulse; the tube current is automatically set for each patient breast by acquisition of two scout images normal to each other (16 mA, 2 pulses of 8 ms each per projection). According to the signal level in the two scout images, the tube current is set between 12 and 100 mA. A complete bCT scan involves the acquisition of 300 projections over a full 360° revolution of the x-ray tube and detector in 10 s. The images were reconstructed with an isotropic voxel size of 0.273 mm with a filtered backprojection algorithm with Shepp-Logan kernel, and corrected for cupping artifacts with a proprietary correction method. The dose varied for each patient breast, with the average value for a breast of mean size and composition being 8.5 mGy.[29,30]

2.B. Dataset

The complete image dataset consisted of a total of 284 breast masses (178 benign and 106 malignant) from 211 patient scans (age: 35–86 yr old; mean: 57; median: 56). Of these masses, 192 (115 benign and 77 malignant, from 138 patient images) were from the UC Davis dataset, and 92 (63 benign and 29 malignant, from 73 patient images) were from the Radboudumc dataset. All masses were identified and localized on the images by experienced breast radiologists. All cysts were diagnosed through breast ultrasound while the nature of the solid masses was biopsy proven. All images were acquired by trained radiographers and collected as part of ethics board-approved patient trials.

Prior to any analysis, approximately 70% of the masses were assigned to a training set (202 masses, n = 133 benign of which 87 were from the UC Davis dataset and 46 from the Radboudumc dataset, n = 69 malignant of which 55 were from the UC Davis dataset and 14 from the Radboudumc dataset), and approximately 30% to a test set (82 masses, n = 45 benign of which 28 were from the UC Davis dataset and 17 from the Radboudumc dataset, n = 37 malignant of

which 22 were from the UC Davis dataset and 15 from the Radboudumc dataset). The process was performed randomly, after stratifying the cases according to mass type, to approximately respect the case distribution in both training and test set. Masses extracted from the same patient scan were assigned to the same set.

2.C. Image preprocessing

After collecting the patient scans, all images underwent preprocessing to obtain a consistent dataset from the bCT systems of the two institutions, which included the upscaling of the images from UC Davis to the same isotropic voxel size as the images from Radboudumc (0.273 mm), and the compression of the dynamic range to 8 bits for all cases.

Subsequently, each 3D mass image was converted into a set of multiple 2D images, to characterize the mass over multiple image views, as previously performed.[31,32] For each mass, nine square image patches of 128 pixel sides (approximately 35 mm, sufficient to contain all masses used in this studies) were extracted, with the direction of each patch kept parallel to one of the nine symmetry planes of an imaginary cube circumscribing the mass (corresponding to the coronal, sagittal, axial, and six oblique views). Such a nine-view approach was chosen to approximate a 3D object as a stack of different 2D images, allowing to capture each mass radiomic signature from different angles, and providing an augmented dataset for overfitting prevention. Each mass-based radiomic signature can then be obtained by combining the individual signatures from each of the nine views, as described later in Section 2.I.

2.D. Mass segmentation

All extracted mass patches were manually segmented using the polyline toolbox in ImageJ® (LOCI, National Institutes of Health, Bethesda, Maryland, USA) by a medical image analysis scientist with over 3 yr of experience in bCT image analysis and segmentation, under the supervision of a board-certified breast radiologist with experience in bCT. This resulted in binarized segmentation masks where the voxels are labeled as either belonging to the mass or not, which will be used to extract shape and contour features and to localize the texture calculation within the mass and its margins (Sections 2.E–2.G).

As previously reported,[32] a subset of the mass patches (n = 35, extracted from as many different masses) was also manually segmented by three breast radiologists. These additional segmentations were used to assess the stability of the radiomic features over different mass contour delineations, to reduce the bias in the derived radiomic signature (see Section 2.I).

2.E. Texture biomarkers for tumor and border heterogeneity

A total of 327 texture feature descriptors were implemented to quantify imaging biomarkers both inside the mass and within the mass border. The location of the inner part of the mass and its border was extracted from the segmentation mask, with the border defined as the annular region containing five voxels inside and outside the mass boundary along each radial direction. The number of voxels included inside and outside the mass border was set to five (i.e., the total border thickness was set to 10 voxels) to ensure the capture of all border information.

Texture was quantified through descriptors belonging to five major groups: histogram-based,[33] Haralick,[34] run length,[35] structural and pattern,[36–39] and Gabor filters.[40]

All descriptors were previously implemented,[33] and a short description with mathematical details is reported in the online supplemental material (Table S1).

2.F. Shape and contour biomarkers for tumor morphology

The radiomic features described in this section aim to quantify the morphological characteristics of breast masses, which have been shown to be an important biomarker of malignancy.[18] A total of 28 descriptors calculated from the binary segmentation mask are proposed, which include regional features based on geometrical characteristics,[41–45] and advanced metrics based on the mass centroid distance function,[46–48] region boundary descriptors,[46,49] and automatic mapping of mass spiculae and lobes. These latter three major feature groups, which aim at detecting different mass contour shapes [of which some examples are shown in Figs. 2(a) and 2(b)] using diverse scales and mathematical formalisms, are described in the following subsections. All features are reported in Table S2.

2.F.1. Centroid distance features

This set of features is based on the centroid distance function (CDF) of the breast mass. The CDF defines the distance from the mass centroid location (x_C, y_C) for each contour pixel of the mass with coordinates (x_i, y_i):

$$CDF_i = \sqrt{(x_i - x_C)^2 + (y_i - y_C)^2} \qquad (1)$$

This function implicitly encodes the frequency and magnitude of potential lobes, spiculae, and irregularities associated with the mass boundary, showing higher peaks in those regions of the contour that present a larger distance from the mass centroid. In the ideal case of regions with radial symmetry, the CDF will appear as a low-frequency, high magnitude wave for irregular and lobulated masses, with the frequency increasing (and the magnitude decreasing) as the mass becomes more spiculated. For regular, elliptical masses, the CDF will instead show two local minima (in the location of the lowest mass radius, and its specular position), whose amplitude is an indicator of the maximum absolute difference among the mass radii. In the extreme case, the CDF becomes constant for a perfectly round mass. Some examples of extracted CDFs are shown in Fig. 2(c).
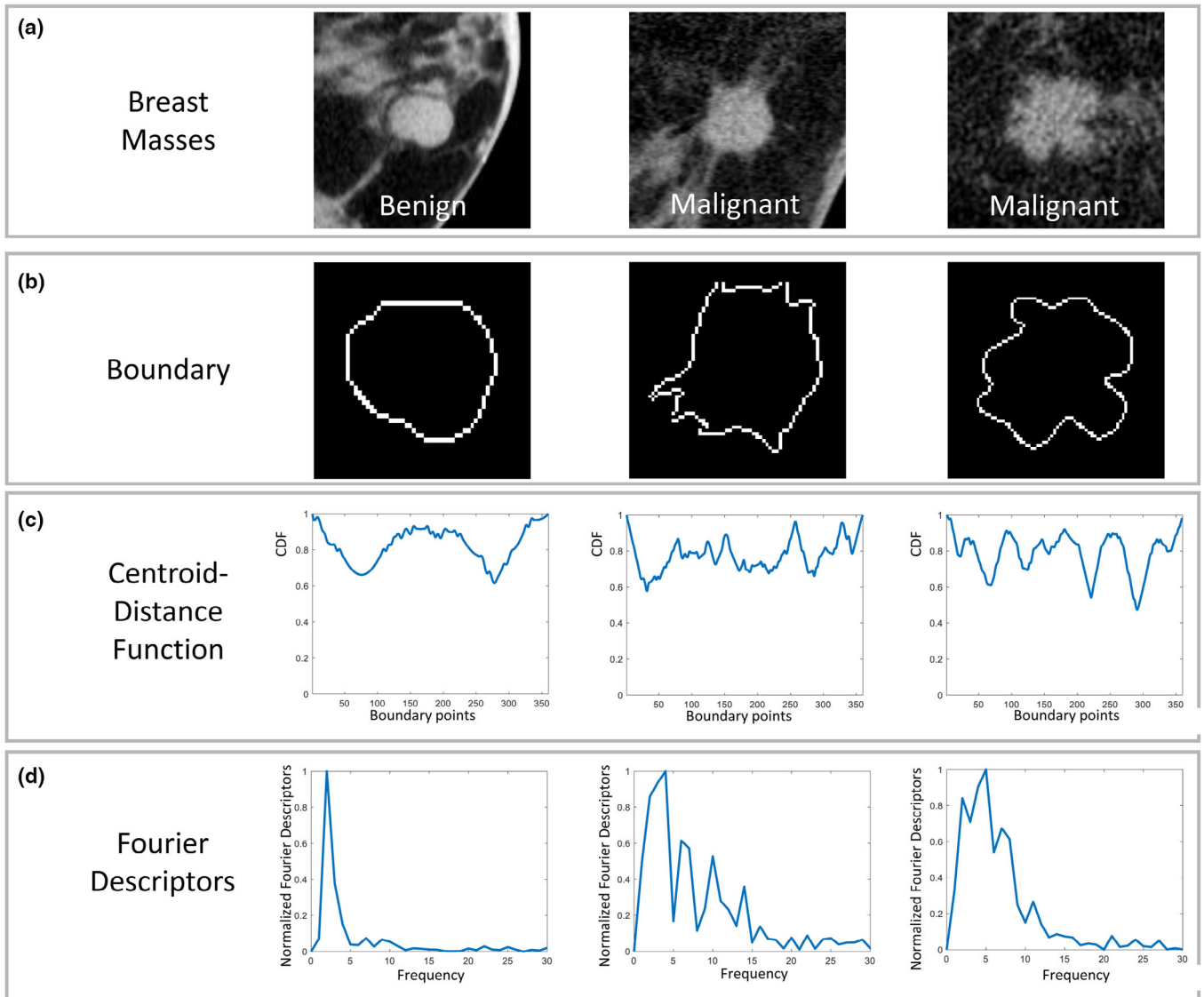
FIG. 2. (a) Examples of breast masses, (b) respective contours of the segmented region, (c) respective centroid distance functions, and (d) power spectrum of the centroid distance function. For benign masses with regular contours, the centroid distance function appears with a much lower frequency content as opposed to malignant masses with irregular shapes. The Fourier descriptors associated with the benign mass show high energy content for the lower frequencies, with the global energy content increasing for malignant masses with a high-frequency associated centroid distance function. [Color figure can be viewed at wileyonlinelibrary.com]

Several well-established descriptors (briefly described in the online supplemental material) were extracted from the CDF to quantify the overall breast mass size in respect to the local contour variations (mean, m_1, standard deviation, M_2), the deviation from a perfect circle (area ratio), the degree of disorder (entropy), and the mass contour roughness.[46,48]

In addition to these descriptors, an additional novel parameter calculated from the mass CDF was developed to quantify different degrees of irregularity in mass contours. To calculate this last feature, the CDF was first normalized to its maximum value, and Fourier-transformed:

$$a_n = FT\{CDF\} \tag{2}$$

Min-max normalization was then performed for scale invariance, and the first Fourier coefficient (corresponding to

the zero frequency) was set to 0 to make the series position-independent. This process resulted in a series of normalized Fourier coefficients corresponding to the CDF spectrum, described as:

$$C_n = \begin{pmatrix} 0 & n=0 \\ 2 \cdot \dfrac{a_n - min(a)}{max(a) - min(a)}; & n=1,2,\ldots,N/2 \end{pmatrix} \tag{3}$$

Finally, the energy content was calculated from this representation, resulting in the proposed radiomic feature descriptor:

$$FD_{energy} = \sum_{n=1}^{N/2} C_n^2 \tag{4}$$

In the case of a high degree of mass irregularity, higher energy content will be distributed across the entire frequency spectrum, resulting in a larger global energy. In the case of regular masses, the energy will be mostly contained at low frequencies. Given that the Fourier descriptors are normalized, this will result in an overall lower energy content. Some examples of these Fourier descriptors extracted from the CDF of real breast masses are shown in Fig. 2(d). In addition, a phantom study showing some results of the analysis of this descriptor is reported in Fig. S1.

2.F.2. Region boundary descriptor

Another descriptor aiming at quantifying the mass contour shape, previously proposed,[46] was implemented, which can distinguish between regular and irregular shapes by analyzing the Fourier transform of the boundary pixel coordinates. Mathematical details are reported, for completeness, in the online supplemental material.

2.F.3. Spicule and lobe map (SLM)

The spicule and lobe map (SLM) aims at analyzing the degree of spiculation and lobe depth by intersecting the original mass contour with its convex enveloping curve. The SLM is designed to discriminate between regular, lobulated, and spiculated shapes. The convex enveloping curve, defined as the smallest region enclosing the mass contour without inflection points, is first generated around the mass, and then iteratively eroded through a morphological circular structuring element whose radius increases with the number of iterations. At each iteration, the number of intersection points between the eroded convex curve and the original mass boundary is detected, and the process is terminated when no further intersections are found. This process extracts information about the mass contour inflections, allowing the ability to map spiculae and lobes in terms of quantity, that is, maximum number of intersections, and size, that is, maximum number of iterations performed prior to the stopping condition (some example on real breast mass contours are shown in Fig. 3).

The maximum number of intersections, $SLM_{intersections}$, and the maximum number of iterations, $SLM_{iterations}$, were used to formulate two new metrics as follows:

$$SLM_P = SLM_{intersections} \cdot SLM_{iterations} \tag{5}$$

$$SLM_R = \frac{SLM_{intersections}}{SLM_{iterations}} \tag{6}$$

Both SLM_P and SLM_R are radiomic descriptors used to map spiculae and lobes of breast masses. Regular masses are expected to show a low number of intersections and a low number of iterations (and therefore a low SLM_P value). In comparison, spiculated and irregular masses should be characterized by a high number of intersections (with the number of iterations being lower for the spiculated cases), and these characteristics are quantified using the SLM_R metric.

The utility of these newly proposed radiomic descriptors in recognizing different realistic mass shapes was evaluated through a phantom study, with a complete description of the process (and related complete findings, Fig. S2 and Table S3) reported in the online supplemental material.

2.G. Margin biomarkers for tumor infiltration degree and peritumoral compartments

The last group of 672 texture-based radiomic features are designed to quantify the degree of infiltration of the mass and its potentially different peritumoral compartments. As opposed to the descriptors presented in Section 2.E, which investigate the texture inside the mass and along its global border, the features proposed here aim at quantifying texture in specific margin regions and orientations, providing complementary information which may strengthen the radiomic signature. These descriptors are divided into two major groups and are described in the following subsections.

2.G.1. Radial gradient features

Radial gradient features are designed to quantify the degree of margin sharpness of breast masses. It is expected that the majority of benign masses present a well-defined margin, indicating an absent or low degree of infiltration (as is the case, i.e., of cysts, non-metastasized lymph nodes, and fibroadenomas). In comparison, malignant masses tend to contain a dense, intricate network of micro-vessels that are often highly concentrated at the tumor periphery. These micro-vessels are used to attract the blood from the existing nearby vessels and therefore nourish the tumor by an increase in blood supply. This usually results in ill-defined boundaries, spiculae, and, consequently, a larger degree of blurring on medical images. To capture these characteristics through image analysis, a group of features evaluating the radial margin gradient distribution was developed.

The boundary of the mass was first identified from the segmentation mask, and the mass margin was extracted as described in Section 2.E. The gradient magnitude (G) of the image (I) was then calculated inside the margin by convolution with a Sobel filter (S).

$$G = \| I * S \| \tag{7}$$

The radial gradient profile was defined for each mass boundary point as the set of pixel values located along the radial direction and covering the entire margin thickness, similar to previous work.[50,51]

From each of the resulting N profiles (where N is the number of the mass boundary pixels), nine features (f) were extracted: mean, standard deviation, maximum, minimum, energy, kurtosis, skewness, entropy, and full-width half maximum (FWHM). Finally, to obtain single measurements related to the entire mass margin, mean ($M_{radial\_gradient}$) and standard deviation ($S_{radial\_gradient}$) for each of the previously mentioned features were calculated and reported as the final radiomic descriptors.
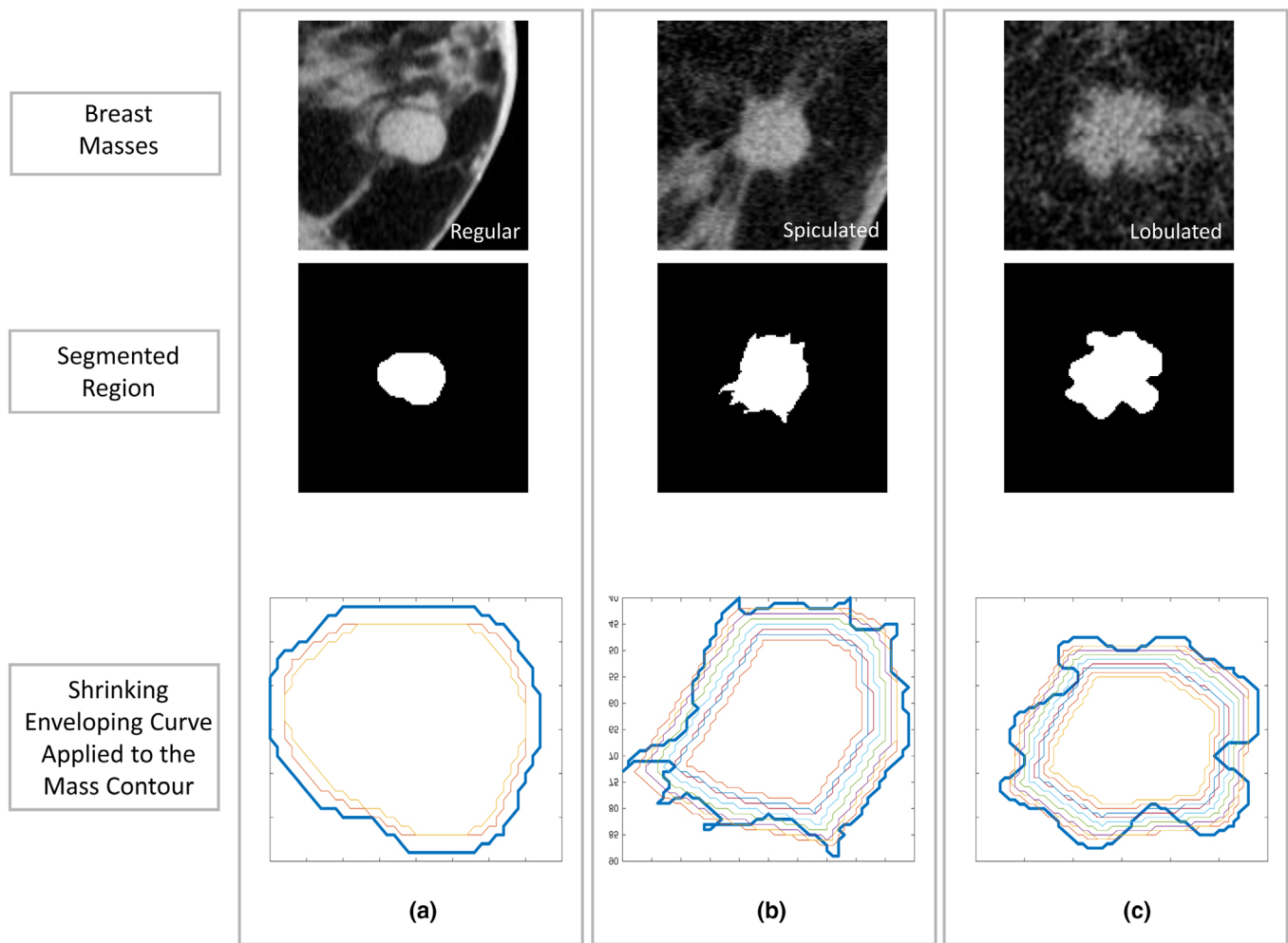
FIG. 3. Examples of the application of the SLM descriptors on real breast mass contours, which evaluate the number of intersection points between the mass boundary and its convex enveloping curve, and the number of shrinking iterations performed on the latter until no further intersections are found. Regular shapes (a) result in few intersections and iterations while both increase as the shape becomes more irregular (b, c). [Color figure can be viewed at wileyonlinelibrary.com]

Intuitively, these features are expected to assume different values according to mass margin sharpness, allowing to detect the overall margin infiltration degree (mean), and possible variations in margin sharpness along the mass contour (standard deviation).

Examples of the application of some of these descriptors on real breast masses are shown in Fig. 4. A phantom study evaluating the effect of different blurring degrees on these features is shown in Fig. S3.

2.G.2. Radial sector features

These final radiomic descriptors evaluate the image texture in discretized regions of the mass margin. A region-based margin analysis is motivated by the fact that some tumors may present multiple phenotypes over different margin locations, and could therefore show diverse texture characteristics across different boundary compartments.[52]

To investigate this effect, the same 327 texture features (f') described in Section 2.E were calculated in 10 different radial sectors (one every 36°) of the mass margin, and the final radiomic descriptors were then represented by their mean ($M_{radial\_sectors}$) and standard deviation ($S_{radial\_sectors}$) across the 10 sectors.

As in the previous subsection, the pair mean–standard deviation was chosen to capture the overall value of each texture feature over the entire margin, and the degree of diversity among the 10 radial sectors evaluated.

Examples of the application of some of these descriptors on real breast masses are shown in Fig. 5.

2.H. Feature evaluation

All previously described, radiomic descriptors were first extracted from the mass images of the training set, and their individual power in discriminating between benign and malignant masses was investigated. For this, each single feature was fed to a linear discriminant analysis (LDA) model,[53] and the performance of the resulting linear decision boundary was evaluated, feature by feature, in terms of area under the
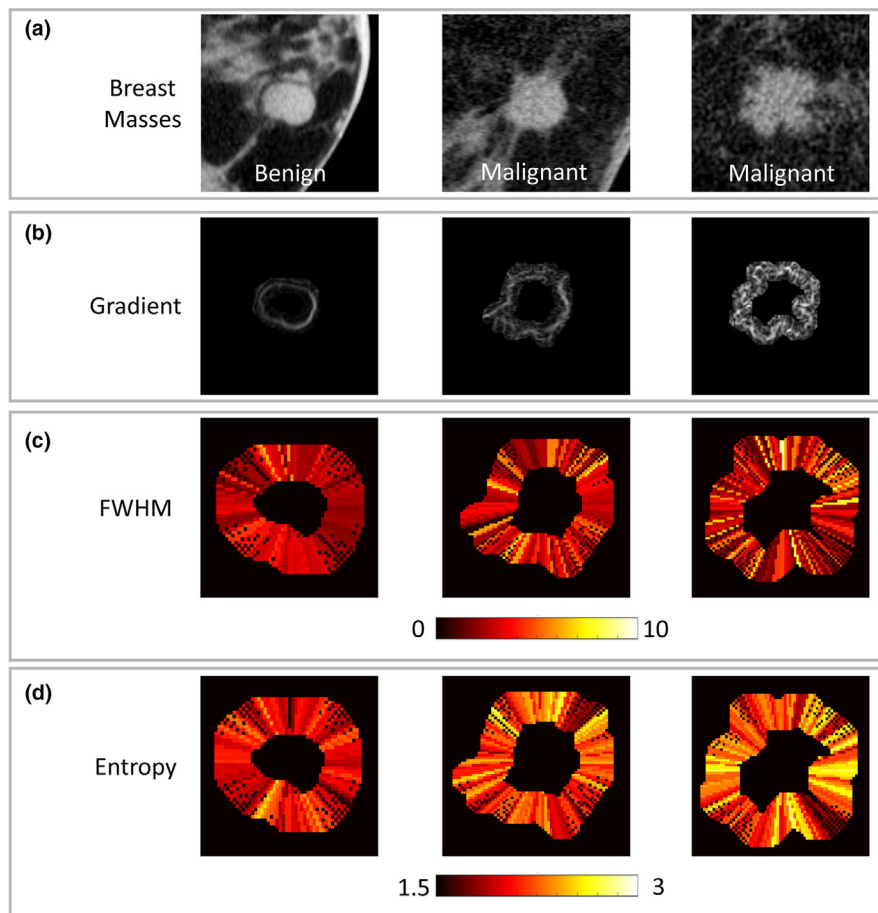
FIG. 4. Examples of the application of some of the radial gradient descriptors on real breast masses. (a) Original breast masses; (b) gradient of the mass margin; heat map of the (c) FWHM and (d) entropy extracted from each radial gradient profile. The benign mass is shown to be characterized by well-defined boundaries, resulting, in this example, in more homogeneous radial gradient features along the margin. In comparison, malignant masses show a higher inhomogeneity in radial gradient features, indicating a blurred and irregular margin. [Color figure can be viewed at wileyonlinelibrary.com]

receiver operating characteristics (ROC) curve (AUC). The possible overlap with random chance (AUC = 0.5) was also assessed, by observing the value of lower limit of the 95% confidence interval (C.I.) for each AUC value, obtained through bootstrapping (1000 bootstraps).[54]

Such an analysis was performed on the training set to understand the overall power of each radiomic feature group, and respective individual features, in characterizing the breast masses without accounting for any potential interaction among different descriptors.

2.I. Radiomic model

To develop a final radiomic-based model for mass classification, another LDA classifier was trained based on multiple radiomic features extracted from the training set masses, after reducing the feature dimensionality through a three-step feature selection pipeline. First, the stability of each of the 1354 features with respect to the mass segmentation was evaluated using the intraclass correlation coefficient (ICC)[55] on the 35 mass patches segmented by the three breast radiologists. Given the high initial feature dimensionality, and the need to

substantially reduce the segmentation bias for a reliable radiomic analysis, the ICC threshold used to consider a feature as stable (and therefore not discarded) was set to higher than 0.9. As an additional stability analysis, systemic differences between the imaging systems used at the two different institutions were investigated, with the aim of discarding those features that are highly dependent on specific imaging system characteristics or acquisition settings. For this, 25 cysts were selected randomly from each of the two datasets (UC Davis and Radboudumc), and the Mann–Whitney U-test was used to test the null hypothesis that the two samples were selected from populations having the same distribution. Features whose distributions demonstrated a statistically significant difference ($P < 0.05$) were eliminated from further analysis. In other words, a feature was considered as robust with respect to image acquisition settings, and therefore suitable for inclusion in subsequent analyses, if the null hypothesis could not be rejected ($P > 0.05$). The p-values were not corrected for multiple comparison, to be more conservative on the number of features to retain, as previously performed.[56] For this analysis, only cysts were chosen, to ensure that any significant difference in feature value found was due
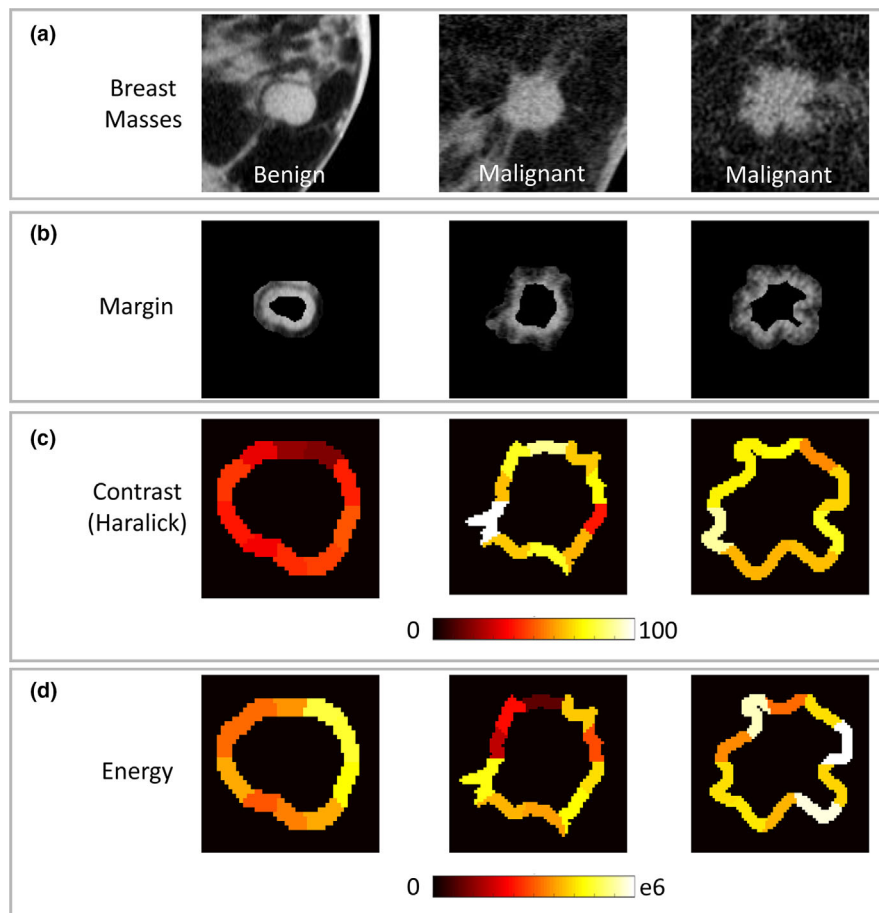
FIG. 5. Examples of the application of some of the radial sector descriptors on real breast masses. (a) Original breast masses; (b) mass margin; heat map of the (c) Contrast (Haralick) and (d) energy extracted from each radial margin sector. Benign masses with homogeneous margins result in homogeneous radial sector features. In comparison, malignant cases show large differences in feature values over the entire margin length. [Color figure can be viewed at wileyonlinelibrary.com]

to imaging system characteristics or acquisition settings, and not due to differences in lesion-type distribution between the two imaging datasets.

Second, the remaining stable features were analyzed statistically using the nonparametric Mann–Whitney *U*-test.[57] Univariate analysis was performed, to test on an individual feature-basis the null hypothesis that samples (benign vs malignant cases from the training set) were selected from populations having the same distribution. The threshold for statistical significance, *P*, was set to 0.05, and adjusted using the Bonferroni correction[58] to account for multiple comparisons. The denominator used for correction was given by the total number of features analyzed, that is, selected after stability to segmentation, and robustness to acquisition settings. As a result of this process, nonstatistically significant features were eliminated. To further reduce the feature space, the ReliefF algorithm[59] was used to select the final most informative descriptors among the stable, statistically significant features, with the number of nearest neighbors in the algorithm, K, set to five. This algorithm is a filter-based approach sensitive to feature interactions, which calculates a relevance score for each feature. This score is based on the identification of

feature value differences between nearest neighbors, and can be applied to rank and select the most informative features. In this study, only the features whose score ranked within the 90th percentile (or higher) of all scores were selected. Finally, to limit the risk of biasing the findings due to potential overfitting, principal component analysis (PCA)[60] was applied on the selected features, and only the first five components were used to train the LDA classifier.

The whole feature analysis and selection method described so far, as well as the training of the LDA classifier, were performed using the masses in the training set, to avoid biasing results toward the test set.

Finally, once the model was trained, radiomic-based classification performance was evaluated on the test set. To provide results on a per-mass level, the LDA-predicted probabilities for the nine views extracted from each mass were averaged into a single score.

3. RESULTS

The phantom studies performed on some of the developed features are described in detail in the online supplemental

material. Briefly, the FD_{energy} [Eq. (4)] increased with irregularity in shape and was invariant to scale, as expected (Fig. S1). The spicule and lobe map (SLM) features resulted in good performance when discriminating between regular, spiculated, and lobulated mass contours (Fig. S2). Lastly, most of the radial gradient features could discriminate between different degrees of blurring in the phantom mass margin (Fig. S3).

Figures 6 shows the overall expression of all radiomic features for all the bCT mass patches of this study.

3.A. Feature evaluation

Overall, over 90% of the features were found to possess informative power (i.e., with an associated AUC 95% C.I. not overlapping with random chance) in recognizing between benign and malignant masses in the training set, when they were fed one-by-one to an LDA classifier. Specifically, 318 of 327 mass texture and 311 of 327 border texture, 26 of 28 shape and contour, and 603 of 654 radial sector features resulted in individual AUC values not overlapping with random chance. For the radial gradient features, instead, only one of 18 descriptor (the average of voxel standard deviations over the N radial profiles) was found to be informative, with an associated AUC value of 0.60.

The average individual AUC values for the mass and border texture were 0.68 (1 standard deviation = 0.03) and 0.67 (1 standard deviation = 0.04), respectively. Noninformative features were minimum voxel value, gray-level 5th percentile, Haralick correlation for all four angles evaluated, and run length gray-level nonuniformity for 0° and 90° (both mass and border), average fractal dimension (mass), and eight Gabor features (border).

The average AUC value for the radial sector features was 0.70 (1 standard deviation = 0.07). Noninformative features were all from first order, Haralick, and run length descriptors.

Of the shape and contour features (average AUC value of 0.75, 1 standard deviation = 0.11), only eccentricity and minimum radius were found noninformative. The remaining descriptors showed, instead, the overall highest discriminant power among all feature groups, with the highest AUC values for individual features being 0.88 (SLM_P), 0.86 (FD_{energy}), and 0.85 (region boundary descriptor).

Table I and Fig. 7 summarize and expand the results reported in this Section.

3.B. Radiomic model

The characteristics of the 82 test set masses (45 benign and 37 malignant) used uniquely for performance evaluation of the final radiomic model are reported in Table II.

Globally, 633 features were found to be stable across multiple segmentations, and 1062 features were found to be robust to the imaging system used. Five hundred and twenty
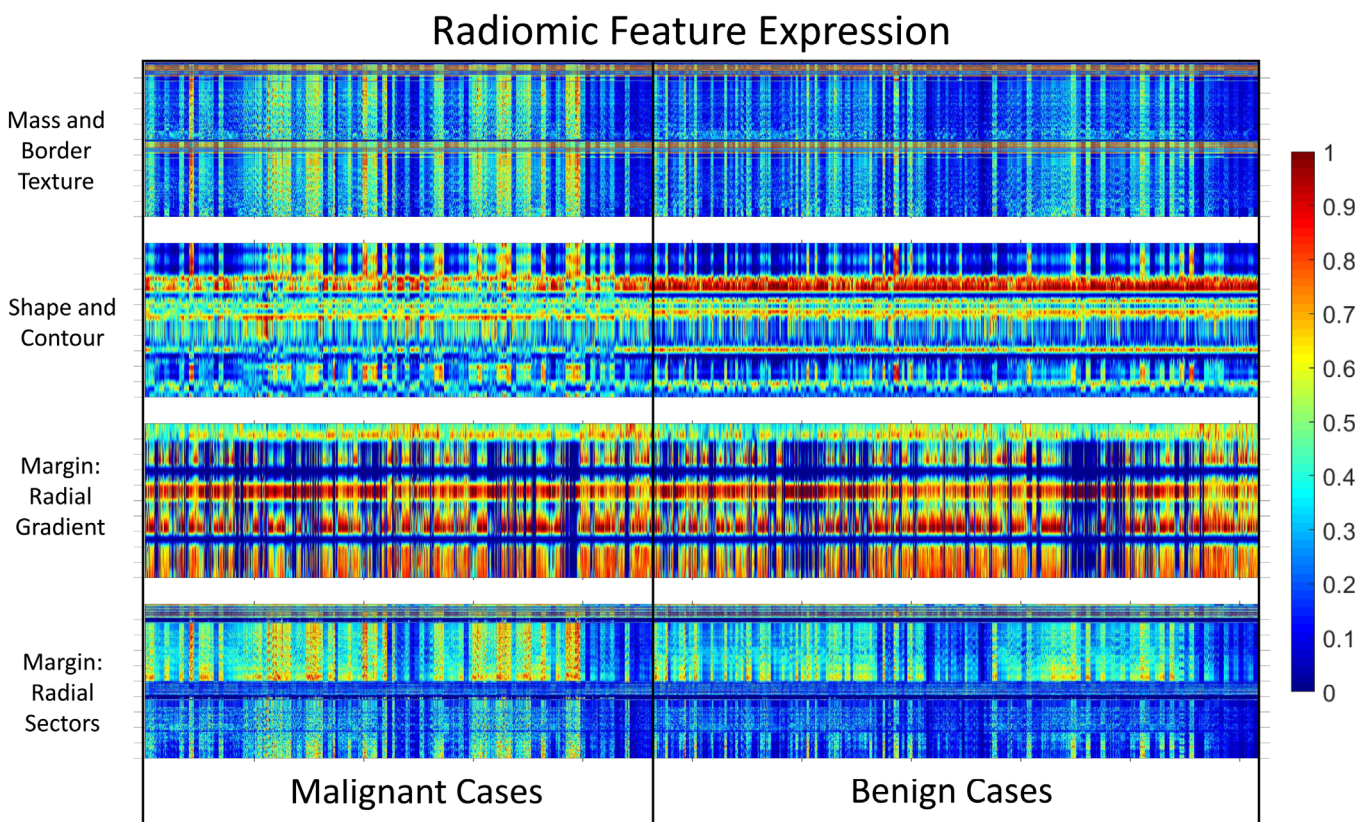
Radiomic Feature Expression

FIG. 6. Radiomic feature expression for the bCT mass dataset. For visualization purposes, each plot was rescaled to a fixed dimension, and features were normalized between 0 and 1. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE I. Results of the feature evaluation process. The table reports the AUC values, per feature group, when each single feature was fed to an LDA model, and the deriving linear decision boundary was used to discriminate benign and malignant masses on the training set.

Feature group	AUC: mean	AUC: standard deviation	AUC: min–max	AUC: median
Mass texture features	0.68	0.03	0.50–0.74	0.69
Mass border features	0.67	0.04	0.50–079	0.68
Shape and contour features	0.75	0.11	0.51–0.88	0.77
Margin – radial gradient features	0.54	0.03	0.50–0.60	0.53
Margin – radial sector features	0.70	0.07	0.45–0.84	0.71

features were stable to both segmentation and system. The results of the first two steps of the feature selection process (stability to segmentation and imaging system, and statistical analysis) are shown in Fig. 8. The percentage of features with high stability to segmentation and robust to imaging system was 46.8% (mass texture), 48.0% (border texture), 25.0% (shape and contour), 61.1% (radial gradients), and 29.4% (radial sectors). After discarding the nonstatistically significant features, these ratios decreased, in absolute values, by 1% (mass texture), 3.5% (shape and contour), 61.1% (radial gradients), and 1.2% (radial sectors) while they remained constant for border texture features.

Overall, of the shape and contour features, simple descriptors such as area, perimeter, axis lengths, and dispersion were stable while complex metrics such as region boundary moments and CDF descriptors were unstable. For mass and border texture, and for radial sector features, there were at least some stable features in all their subgroups. Radial gradient features were, overall, the most stable, but none of the descriptors resulted in statistical significance.

After performing the third feature selection step through the ReliefF algorithm, 36 features were selected (three mass texture, 26 border texture, one shape and contour, and six radial sector features), and their first five components (obtained through PCA) used to develop the radiomic-based LDA model. These features are reported in Table S4.

The final radiomic-based LDA model achieved an AUC on the 82 test set masses of 0.90 (95% C.I. 0.80–0.96). The ROC curve with 95% C.I. is shown in Fig. 9. At a 95% sensitivity (specificity 56%), two of 37 malignant masses (1/6 ductal carcinoma in situ, 1/19 invasive ductal carcinoma) and 20 of 45 benign masses (4/10 fibroadenomas, 16/29 cysts) were misclassified.

4. DISCUSSION

In this work, an algorithm for radiomic-based characterization of breast masses was developed, which includes several descriptors aiming at quantifying imaging biomarkers in terms of mass and border texture, morphology, and margin sharpness and diversity. The radiomic signature deriving from this multi-marker analysis resulted in high performance in recognizing benign and malignant masses in bCT imaging.

Overall, the majority of the implemented and developed radiomic features showed some informative power. Nevertheless, when analyzed on a feature-type basis, specific patterns and differences in performance were observed among the descriptor groups. While texture information, whether calculated inside the lesions or along their border in its entirety,
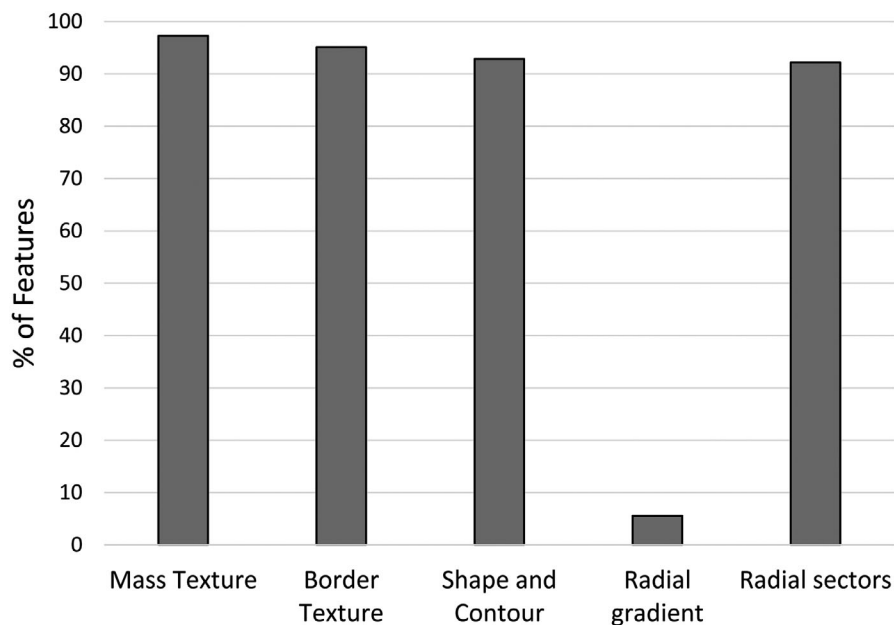

FIG. 7. Percentage of features, per feature group, with informative power (i.e., with an associated lower extreme of the AUC 95% C.I. higher than 0.5), when each single feature was fed to an LDA model, and the deriving linear decision boundary was used to discriminate benign and malignant masses on the training set.

TABLE II. Characteristics of the test set breast masses (N = 82).

Benign masses (n = 45)	
Cyst	29
Fibroadenoma	10
Atypical ductal hyperplasia	1
Blunt duct adenosis	1
Hamartoma	1
Lymph node	2
Fibrocystic change	1
Malignant masses (n = 37)	
Invasive ductal carcinoma	19
Ductal carcinoma *in situ*	6
Invasive lobular carcinoma	2
Invasive mammary carcinoma	3
Adenocarcinoma	1
Combination of tumor types	6

resulted in an overall relevant diagnostic power and good stability, the other feature types showed different characteristics. Specifically, when advanced shape and contour descriptors where analyzed individually, they resulted in the highest discriminant power compared to all other feature types, but also in the lowest stability. This led to their exclusion from the final radiomic model, due to their diagnostic power being strongly dependent on the mass segmentation results. In contrast, radial gradient features showed an opposite pattern, with a high associated stability, but low discriminant power, a result which led to the exclusion of all these descriptors from the final radiomic signature. This finding suggests that analyzing the mass margins in their entirety (or in a finite number of discrete regions) provides more relevant information compared to that obtained from analyzing radial directions,

and it could be due to the finite voxel dimension in the images, and associated partial volume effect, limiting the information extracted from one-dimensional profiles. The introduced radial sector features behaved similar to the texture descriptors calculated on the entire mass border in terms of stability and individual diagnostic power, but they also carried additional, and complementary, information. This is highlighted by the fact that, during the feature reduction step performed through the ReliefF algorithm, some of the radial sector features were ranked among the top descriptors and were, therefore, included in the final radiomic model.

It is interesting to note that the majority of the features included in the final signature were extracted from the mass margin (border texture and radial sector features). This suggests that a margin-based radiomic analysis may lead to a better characterization of breast masses compared to analyzing only the texture information inside the mass. This finding aligns with the BI-RADS[®] [lexicon,61] where the correct evaluation of the tumor margin is one of the strongest biomarkers in the discrimination between benign and malignant breast masses, and is reflected by the margin-based radiomic features showing higher values for malignant lesions, indicating a higher degree of disorder and infiltration (Fig. 6). Further analysis of these features should be performed in future work, to assess their correlation with the underlying tumor biology characteristics observed in pathological samples.

While margin information seems to be the most relevant, the final radiomic signature also included shape and textural descriptors calculated inside the mass boundary, pointing to the benefit of a multi-marker radiomic signature deriving from different feature types. Of course, future studies with larger image datasets are needed to further assess the validity of our results, and of these conclusions. This is especially true regarding feature stability, where additional analyses
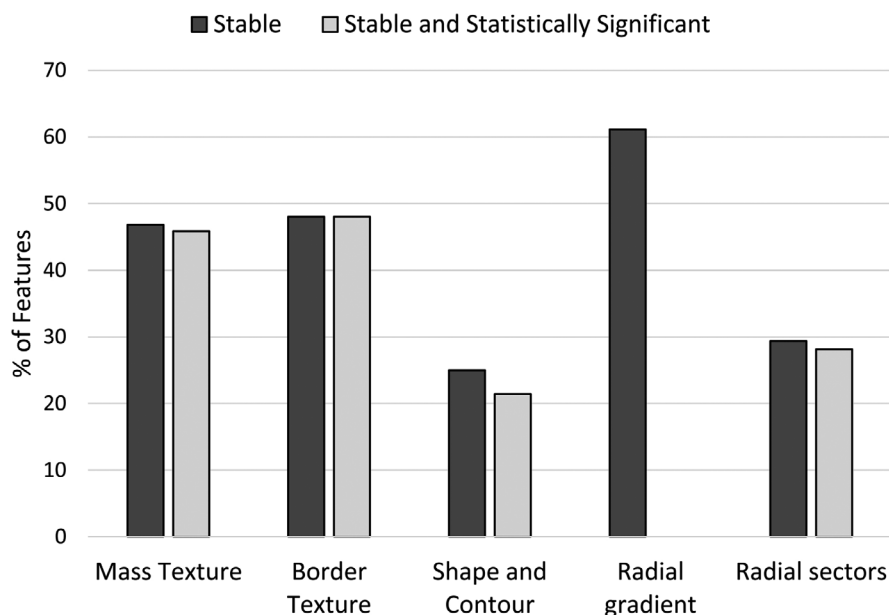

FIG. 8. Percentage of features, per feature group, with high stability to segmentation and robustness to acquisition settings (*stable*), and resulting in statistical significance between benign and malignant masses (*stable and statistically significant*).
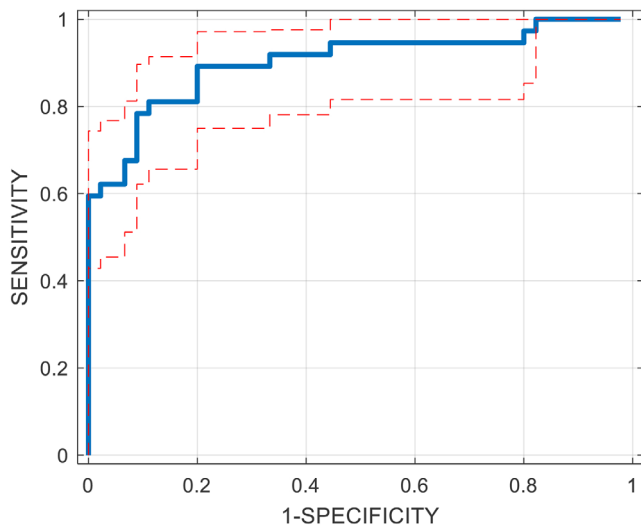
FIG. 9. ROC curve of the final radiomic-based LDA model on the test set masses. Dotted lines indicate the 95% C.I. calculated through bootstrapping (1000 bootstraps). [Color figure can be viewed at wileyonlinelibrary.com]

performed with more images and more segmentation results obtained from different readers (or the same readers multiple times) could provide further insight on which features can be used safely in the proposed radiomic model. Furthermore, results from automated segmentation algorithms could also be assessed, to study how feature stability changes when going from manual to automatic segmentation.

Some of the blocks and parameters used in the proposed radiomic pipeline could also be further investigated in future and with additional images, including the evaluation of different methods for reducing the feature space, different ICC thresholds, and different classification models. While the ICC threshold was already set to a conservative level (0.9), and the choice of the feature reduction method has been shown to have a minimal effect on the results,[62] the classification model selection could, on the contrary, impact the results to a larger extent.[62] In this study, LDA was chosen to demonstrate the radiomic performance in a full feature-driven setting, with the simplest decision boundary possible (i.e., linear). While this allowed for an objective evaluation of the extracted radiomic signature, future work should evaluate the performance variability, and potential improvement, when using different machine learning techniques.

Due to the current limited clinical use of bCT worldwide, resulting in limited data availability, in this study a radiomic model based on mass images acquired with two different systems was developed. While this strengthened our study in terms of generalizability, differences in imaging conditions could affect the value of the extracted radiomic features[63] (although this effect was mitigated by eliminating those features with low robustness with respect to imaging system characteristics and acquisition settings). To estimate the magnitude of this effect, the developed radiomic pipeline (including feature selection and model training) was performed using the UC Davis dataset only (192 masses), and tested on the Radboudumc dataset (92 masses), and vice-versa,

resulting in an AUC value of 0.88 and 0.85, respectively. While this could suggest that the radiomic signature across bCT devices is transferable (if a large training set is used), limited conclusions can, currently, be drawn on the effect of the imaging conditions on the resulting radiomic-based diagnostic accuracy. In fact, this slight difference in AUC values is, probably, mostly due to the difference in size between the two datasets. Therefore, further analyses on the effect of different imaging system and acquisition characteristics on the extracted radiomic signature should be repeated when larger, balanced datasets acquired from the different systems become available (with each dataset possibly following the same distribution of lesion characteristics and types), and with comprehensive phantom studies.

In this study, the radiomic analysis and deriving classification was performed by approximating the 3D breast masses as a set of multiple 2D image patches extracted over different image views. While a fully 3D analysis could provide further insights into tumor characterization, some previous studies have reported high performance with 2D radiomic analyses of tomographic data,[2,5,64] while others also assessed the advantage of 2D over 3D radiomic features,[65] and the potential improvement in performance deriving from their combination.[66] Therefore, a 2D analysis was chosen for two major reasons. First, the advantages of a lower computational complexity eased the validation of the newly proposed radiomic descriptors, since working in a 2D space reduced the complexity in mathematical formulation and implementation, and the potential uncertainties deriving from applying mathematical models (the radiomic descriptors) to a discrete space with limited voxel dimensions (the image). Second, a 2D approach allowed to increase the training set size by extracting multiple patches from each mass, helping increase the robustness of the diagnostic classifier. This can be useful both from the perspective of an augmented dataset (thus preventing overfitting), and in terms of respective feature dimensionality. In fact, working with larger training sets allows for the selection of a higher number of features, which can potentially capture the different tumor characteristics better and, therefore, strengthen the derived final radiomic signature. Finally, it should be noted that the proposed approach, although performed on a 2D basis, did take advantage of the 3D nature of the images, since superimposition of tissue was still avoided, and the final radiomic score for each mass was obtained by combining different signatures extracted over multiple angles. This could provide a stronger characterization compared to simply performing the analysis on a single 2D patch collected from each mass, as often performed in 2D radiomic analyses.[67] To test this latter hypothesis, the final radiomic model was used to reclassify the test set masses based only on the signatures extracted from a single patch (coronal view), resulting in a significantly lower performance (AUC = 0.84 vs 0.90). Nevertheless, a fully 3D radiomic approach should also be investigated in future, to allow for the evaluation, with larger image datasets, of the potential advantage of a fully 3D radiomic signature over the current 2D approach.
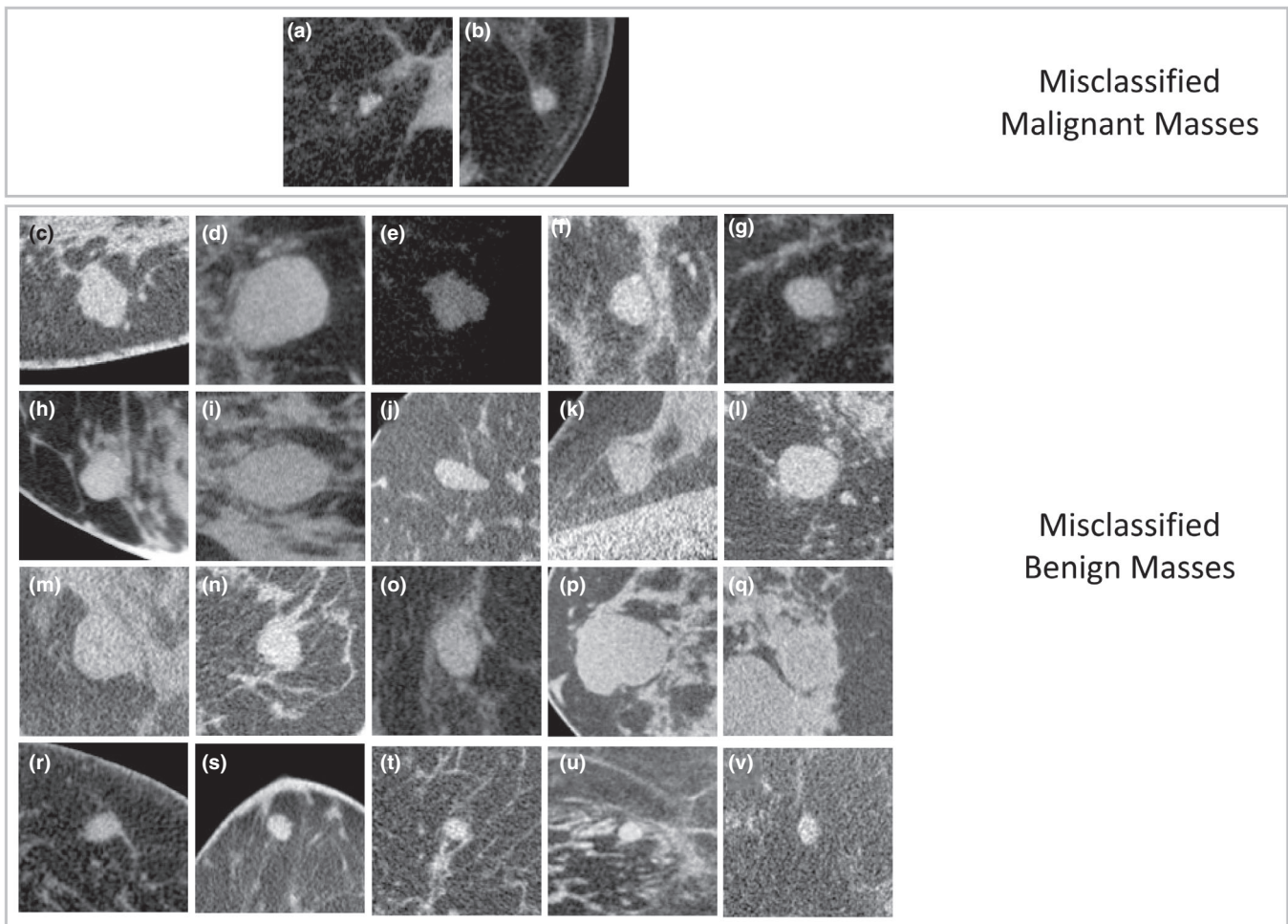
FIG. 10. Misclassified masses at 95% sensitivity operating point.

In this study, images were discretized to 8 bits prior to any analysis. While the results obtained in this work are encouraging, systematic studies on the effect of bin width on the resulting radiomic features are to be performed in future, especially if bCT becomes of widespread use. These studies will allow for the definition of a standardized image discretization methodology in bCT radiomics, helping its potential advancement into the medical imaging realm.

Although promising, the proposed radiomic algorithm can be improved in future work. At a 95% sensitivity (specificity 56%), two malignant and 20 benign masses were misclassified. By isolating these cases, it was observed that the two malignant masses (Fig. 10, panels a–b) presented a very small diameter (8 mm ± 1.2 mm) while the majority (15/20) of benign cases (Fig. 10, panels c–q) presented, instead, an overall larger diameter (16 mm ± 3.4 mm). Therefore, the misclassification is in line with the well-known influence of lesion size on the diagnostic outcome, with a larger size being usually associated with an increased likelihood of malignancy.[68] The five remaining benign misclassified cases had a smaller size (8 mm ± 0.9 mm), and misclassification occurred probably due to their margins, which present some

degree of irregularity (Fig. 10, panels r–v). To improve performance, future plans include increasing the dataset size, the evaluation of additional strategies to merge the 2D view-based radiomic signatures (in addition to averaging), and the 3D implementation of the described features, to investigate the potential advantage of a fully 3D radiomic signature over the current 2D multi-view approach. The use of automatic mass segmentation will also be considered, as well as the comparison, and potential synergy, of the proposed radiomic pipeline with deep learning approaches based on convolutional neural networks.

5. CONCLUSIONS

The proposed radiomic algorithm demonstrated high performance in the classification of benign and malignant masses in bCT imaging, thanks to a multi-marker radiomic signature based on different texture, shape, and margin descriptors. While further research is needed, the proposed approach is a promising application for computer-aided diagnosis of breast cancer, potentially helping improve the diagnostic process through an increase in sensitivity or a reduction of benign biopsies performed.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors John M. Boone and Andrew M. Hernandez have patents (pending and issued) pertaining to breast CT. The author John M. Boone has prior research support, licensing agreements with, and is a shareholder of Izotropic Imaging Corp. of Canada.

a)Author to whom correspondence should be addressed. Electronic mail: Ioannis.Sechopoulos@radboudumc.nl.

REFERENCES

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures. They are data. *Radiology*. 2013;278:563–577.
2. Khorrami M, Khunger M, Zagouras A, et al. Combination of peri- and intratumoral radiomic features on baseline CT scans predicts response to chemotherapy in lung adenocarcinoma. *Radiol Artif Intell*. 2019;1:1–10.
3. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:1–8.
4. Huang Y, Liu Z, He L, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology*. 2016;281:947–957.
5. Varghese B, Chen F, Hwang D, et al. Objective risk stratification of prostate cancer using machine learning and radiomics applied to multiparametric magnetic resonance images. *Sci Rep*. 2019;9:1570–1580.
6. Crivelli P, Ledda RE, Parascandolo N, Fara A, Soro D, Conti M. A new challenge for radiologists: radiomics in breast cancer. *Biomed Res Int*. 2018;7:1–10.
7. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013;49:1374–1403.
8. Drukker K, Giger ML, Joe BN, et al. Combined benefit of quantitative three-compartment breast image analysis and mammography radiomics in the classification of breast masses in a clinical data set. *Radiology*. 2019;290:621–628.
9. Tagliafico AS, Valdora F, Mariscotti G, et al. An exploratory radiomics analysis on digital breast tomosynthesis in women with mammographically negative dense breasts. *Breast*. 2018;40:92–96.
10. Chitalia RD, Kontos D. Role of texture analysis in breast mri as a cancer biomarker: a review. *J Magn Reson Imaging*. 2019;49:927–938.
11. Jahani N, Cohen E, Hsieh MK, et al. Prediction of treatment response to neoadjuvant chemotherapy for breast cancer via early changes in tumor heterogeneity captured by DCE-MRI registration. *Sci Rep*. 2019;9:12114.
12. Xiong Q, Zhou X, Liu Z, et al. Multiparametric MRI-based radiomics analysis for prediction of breast cancers insensitive to neoadjuvant chemotherapy. *Clin Transl Oncol*. 2020;22:50–59.
13. Yang J, Wang T, Yang L, et al. Preoperative prediction of axillary lymph node metastasis in breast cancer using mammography-based radiomics method. *Sci Rep*. 2019;9:4429–4440.
14. Huang S-Y, Franc BL, Harnish RJ, et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *NPJ Breast Cancer*. 2018;4:1–13.
15. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multiparametric breast MRI. *Radiology*. 2019;290:290–297.
16. Lee SE, Han K, Kwak JY, Lee E, Kim E-K. Radiomics of US texture features in differential diagnosis between triple-negative breast cancer and fibroadenoma. *Sci Rep*. 2018;8:13546.
17. Ray S, Prionas ND, Lindfors KK, Boone JM. Analysis of breast CT lesions using computer-aided diagnosis: an application of neural networks on extracted morphologic and texture features. *Proc. of SPIE Med. Im*; 2012;8315:83152E1-E6.
18. Surendiran B, Palaniappan R, Vadivel A. Effect of BIRADS shape descriptors on breast cancer analysis. *Int J Med Eng Inform*. 2015;7:65–79.
19. Li H, Mendel KR, Lan L, Sheth D, Giger ML. Digital mammography in breast cancer: additive value of radiomics of breast parenchyma. *Radiology*. 2019;291:15–20.
20. Sahiner B, Chan HP, Petrick N, Helvie MA, Hadjiiski LM. Improvement of mammographic mass characterization using speculation measures and morphological features. *Med Phys*. 2001;28:1455–1465.
21. Chan H-P, Wu Y-T, Sahiner B, et al. Characterization of masses in digital breast tomosynthesis: comparison of machine learning in projection views and reconstructed slices. *Med Phys*. 2010;37:3576–3585.
22. 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;281:382–392.
23. Gilhuijs KG, Giger ML, Bick U. Computerized analysis of breast lesions in three dimensions using dynamic magnetic resonance imaging. *Med Phys*. 1998;25:1647–1654.
24. Nugroho HA, Triyani Y, Rahmawaty M, Ardiyanto I. Analysis of Margin Sharpness for Breast Nodule Classification on Ultrasound Images. *ICITEE Proc*. 2017;1–5.
25. Lee J-H, Seong Y-K, Chang C-H, et al. Fourier-based shape feature extraction technique for computer-aided b-mode ultrasound diagnosis of breast tumor. *IEEE EMBS Proc*; 2012:6551–6554.
26. Lindfors KK, Boone JM, Nelson TR, Yang K, Kwan ALC, Miller DF. Dedicated breast CT: initial clinical experience. *Radiology*. 2008;246:725–733.
27. Boone JM, Yang K, Burkett GW, et al. An X-ray computed tomography/positron emission tomography system designed specifically for breast imaging. *Technol Cancer Res Treat*. 2010;9:29–44.
28. Ghazi P, Hernandez AM, Abbey C, Yang K, Boone JM. Shading artifact correction in breast CT using an interleaved deep learning segmentation and maximum-likelihood polynomial fitting approach. *Med Phys*. 2019;46:3414–3430.
29. Boone JM, Nelson TR, Lindfors KK, Seibert JA. Dedicated breast CT: radiation dose and image quality evaluation. *Radiology*. 2001;221:657–667.
30. Sechopoulos I, Feng SSJ, D'Orsi CJ. Dosimetric characterization of a dedicated breast computed tomography clinical prototype. *Med Phys*. 2010;37:4110–4120.
31. Setio AAA, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging*. 2016;35:1160–1169.
32. Caballo M, Pangallo DR, Mann RM, Sechopoulos I. Deep learning-based segmentation of breast masses in dedicated breast CT imaging: radiomic feature stability between radiologists and artificial intelligence. *Comput Biol Med*. 2020;118:103629.
33. Caballo M, Teuwen J, Mann RM, Sechopoulos I. Breast parenchyma analysis and classification for breast masses detection using texture feature descriptors and neural networks in dedicated breast CT images. *Proc SPIE Med. Im*. 2019;10950:09500J1.
34. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Trans Syst Man Cybern*. 1973;3:610–621.

35. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process*. 1975;4:172–179.

36. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell*. 2002;24:971–987.

37. Laws KI. Rapid Texture Identification. In: *Proc. SPIE 0238, Image Processing for Missile Guidance*; 1980. https://doi.org/10.1117/12.959169

38. Iwahori Y, Hattori A, Adachi Y, Bhuyan M, Woodham R, Kasugai K. Automatic detection of polyp using Hessian filter and HOG features. *Proc Comput Sci*. 2015;60:730–739.

39. Keller JM, Chen SS, Crownover RM. Texture description and segmentation through fractal geometry. *Comput Vis Graph Image Process*. 1989;45:150–166.

40. Haghighat M, Zonouz S, Abdel-Mottaleb M. CloudID: trustworthy cloud-based and cross-enterprise biometric identification. *Expert Syst Appl*. 2015;42:7905–7916.

41. Liu J, Shi Y. Image Feature Extraction Method Based on Shape Characteristics and Its Application in Medical Image Analysis. In: *Applied Informatics and Communication. ICAIC 2011. Communications in Computer and Information Science*. 2011;224:172-178.

42. Chang C, Liu W, Zhang HJ. Image retrieval based on region shape similarity. *SPIE Proc*. 2001;4315:31–37.

43. Surendiran B, Vadivel A. Mammogram mass classification using various geometric shape and margin features for early detection of breast cancer. *Int J Med Eng Inform*. 2012;4:36–54.

44. Peura M, Iivarinen J. Efficiency of simple shape descriptors. *Proceedings of the Third International Workshop on Visual Form*; 1997:443–451.

45. Kpalma K, Ronsin J. Multiscale contour description for pattern recognition. *Pattern Recogn Lett*. 2006;27:1545–1559.

46. Shen L, Rangayyan RM, Desautels JL. Application of shape analysis to mammographic calcifications. *IEEE Trans Med Imaging*. 1994;13:263–274.

47. Kilday J, Palmieri F, Fox MD. Classifying mammographic lesions using computerized image analysis. *IEEE Trans Med Imaging*. 1993;12:664–669.

48. Gupta L, Srinath MD. Contour sequence moments for the classification of closed planar shapes. *Pattern Recogn*. 1987;20:267–272.

49. Rangayyan RM, El-Faramawy NM, Desautels JE, Alim OA. Measures of acutance and shape for classification of breast tumors. *IEEE Trans Med Imaging*. 1997;16:799–810.

50. Huo Z, Giger ML, Vyborny CJ, et al. Analysis of spiculation in the computerized classification of mammographic masses. *Med Phys*. 1995;22:1569–1579.

51. Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Schmidt RA, Doi K. Automated computerized classification of malignant and benign masses on digitized mammograms. *Acad Radiol*. 1998;5:155–168.

52. Wu J, Cao G, Sun X, et al. Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy. *Radiology*. 2018;288:26–35.

53. Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugen*. 1936;7:179–188.

54. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press; 1994:57.

55. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.

56. Drukker K, Edwards A, Doyle C, Papaioannou J, Kulkarni K, Giger ML. Breast MRI radiomics for the pretreatment prediction of response to neoadjuvant chemotherapy in node-positive breast cancer patients. *J Med Imaging*. 2019;6:034502.

57. Hollander M, Wolfe DA. *Nonparametrical Statistical Methods*. Hoboken, NJ: John Wiley and Sons, Inc.; 1999.

58. Cheverud JM. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity*. 2001;87:52–58.

59. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: introduction and review. *J Biomed Inform*. 2018;85:189–203.

60. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chem Intell Lab Syst*. 1987;2:37–52.

61. D'Orsi CJ, Sickles EA, Mendelson EB, et al. *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology; 2013.

62. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine learning methods for quantitative radiomic biomarkers. *Sci Rep*. 2015;5:13087.

63. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749–762.

64. Cha KH, Hadjiiski L, Chan H-P, et al. bladder cancer treatment response assessment in CT using radiomics with deep-learning. *Sci Rep*. 2017;7:8738.

65. Shen C, Liu Z, Guan M, et al. 2D and 3D CT radiomics features prognostic performance comparison in non-small cell lung cancer. *Transl Oncol*. 2017;10:886–894.

66. Yang L, Yang J, Zhou X, et al. Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur Radiol*. 2019;29:2196–2206.

67. Xu L, Yang P, Yen EA, et al. A multi-organ cancer study of the classification performance using 2D and 3D image features in radiomics analysis. *Phys Med Biol*. 2019;64:215009.

68. Mirza AN, Mirza NQ, Vlastos G, Singletary SE. Prognostic factors in node-negative breast cancer: a review of studies with sample size more than 200 and follow-up more than 5 years. *Ann Surg*. 2002;235:10–26.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Supplementary methods and results.