10.1148/radiol.2019181113

Appendix E1

Section 1. Preparation of Radiomic Features

A. Preparation of Radiomic Features Extracted from Breast Tumor

Our computerized breast lesion analysis methods for mammography involve several steps. The first step is breast lesion segmentation. The localized breast lesion is automatically segmented from the background parenchyma using a radial gradient index (RGI) segmentation method (36) to generate the initial breast lesion contour which is relatively close to the actual margin. Then an active contour method (20) based on level sets is applied on the initial segmentation to generate the final segmentation. Next, a total of 32 radiomic features are automatically extracted. These descriptors can be divided into six categories. Spiculation features include a radial-edge gradient analysis to derive two spiculation related features, the "Normalized Radial gradient," and "FWHM" (full width at half maximum) (1,37). The spiculation measure is calculated on four neighborhood regions, including grown region, margin, encompassing region, and surrounding periphery. Radial gradient index (21), margin sharpness measure (21), density related mass features (21), size, shape features (38) and textural features based on gray level co-occurrence (GLCM) are also calculated (22,39,40). Table E1 gives the name, brief description, and category of all the radiomic features extracted from breast mass lesion.

B. Preparation of Radiomic Features Extracted from Parenchyma Stroma

The computerized parenchymal texture analysis method involves several steps. The first step is regions-of-interest (ROIs) selection. ROIs of 256 by 256 pixels are manually selected from the central breast region immediately behind the nipple of craniocaudal (CC) projection of mammographic view, since that region usually includes the dense parts of the breast. ROIs are selected such that regions along the skin line, which contains subcutaneous fat, are not included. Our prior studies demonstrated the benefit of the ROI behind the nipple in assessing risk of cancer, and its improved performance as compared with ROIs throughout the breast (24). We believe, use of the entire breast area is ideal for evaluating the percent breast density. However, the results from our study best assess the texture of dense regions, which usually occurs behind the nipple as opposed to percent dense. The detailed explanation regarding ROI extraction has been reported elsewhere (24). Then, quantitative radiomic texture feature extractions are performed on each ROI to characterize the mammographic parenchymal patterns. A total of 45 texture features are calculated from each individual ROIs to characterize the local composition, contrast, homogeneity, and coarseness of breast parenchyma (39–43). These features are based on *(a)* gray-level histogram analysis to characterize the local tissue composition, *(b)* fractal analysis, including box-counting and Minkowski methods, *(c)* edge-frequency analysis, *(d)* Fourier analysis, including root mean square variation, first moment of power spectrum, and power law beta from power spectral analysis, *(e)* textural analysis using neighborhood gray-tone difference matrix and gray-level co-occurrence matrix to characterize the spatial relationship among gray levels. Table E2 gives the name, brief description, and category of all radiomic features extracted from the breast parenchyma stroma.

References

36. Kupinski MA, Giger ML. Automated seeded lesion segmentation on digital mammograms. IEEE Trans Med Imaging 1998;17(4):510–517 .

37 Giger ML, Vyborny CJ, Schmidt RA. Computerized characterization of mammographic masses: analysis of spiculation. Cancer Lett 1994;77(2-3):201–211 .

38. Beutel J, ed. Handbook of medical imaging. Bellingham, Wash: SPIE Press, 2000.

39. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Trans Syst Man Cybern 1973;SMC-3(6):610–621 .

40. Sonka M, Hlavac V, Boyle R. Image processing, analysis, and machine vision. 3rd ed. Toronto: Thompson Learning, 2008.

41. Li H, Giger ML, Olopade OI, Chinander MR. Power spectral analysis of mammographic parenchymal patterns for breast cancer risk assessment. J Digit Imaging 2008;21(2):145–152 .

42. Li H, Giger ML, Olopade OI, Lan L. Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment. Acad Radiol 2007;14(5):513–521 .

43. Li H, Giger ML, Olopade OI, Margolis A, Lan L, Chinander MR. Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms. Acad Radiol 2005;12(7):863–873 .

Table E1: List of names, descriptions, and categories of all 32 radiomic features extracted from the breast lesion.

Category	Name	Description
Spiculation features	FWHM (grown region)	Measure of the shape of the normalized cumulated edge-gradient distribution within grown lesion region
	Normalized radial gradient (grown region)	Quantitative measure of the spiculation of a mass lesion within grow lesion region
	FWHM (margin)	Measure of the shape of the normalized cumulated edge-gradient distribution along lesion margin
	Normalized radial gradient (margin)	Quantitative measure of the spiculation of a mass lesion along its margin
	FWHM (encompassing region)	Measure of the shape of the normalized cumulated edge-gradient distribution within encompassing region
	Normalized radial gradient (encompassing region)	Quantitative measure of the spiculation of a mass lesion within encompassing region
	FWHM (surrounding periphery)	Measure of the shape of the normalized cumulated edge-gradient distribution within surrounding periphery
	Normalized radial gradient (surrounding periphery)	Quantitative measure of the marginal spiculation of a mass lesion within surrounding periphery
Radial gradient index	Average radial gradient	Average radial gradient within the encompassing region
Margin sharpness	Margin sharpness	Quantitative measure of the lesion margin
Density features	Average gray value	Average the gray-level values of each pixel within the grown region of a mass
	Contrast1	Difference between the average gray level of the grown region and the 20% threshold of the surrounding fatty area
	Contrast2	Difference between 95% threshold of the grown region and the 20% threshold of the surrounding fatty area
	Standard deviation of gray level	Standard deviation of gray value within a mass region to quantify the heterogeneity of the lesion
Size	Effective diameter	Diameter of a circle with the same area as the lesion
Shape	Circularity	Ratio of area of grown potential mass within effective circle over area of potential mass

	Irregularity	Deviation of the perimeter of the mass lesion from the perimeter of a circle with equivalent area
	Compactness	Ratio of area of grown region over perimeter of grown region squared
Texture features based on GLCM	Contrast (GLCM)	Measure of local image variations
	Correlation (GLCM)	Measure of image linearity
	Difference entropy (GLCM)	Measure of the randomness of the difference of neighboring pixels' gray-levels
	Difference variance (GLCM)	Measure of variations of difference of gray-levels between pixel-pairs
	Energy (Angular second moment)	Measure of image homogeneity
	Entropy	Measure of the randomness of the gray-levels
	Inverse difference moment	Measure of the image homogeneity
	Information measure of correlation 1 (IMC1)	Measure of nonlinear gray-level dependence
	Information measure of correlation 2 (IMC2)	Measure of nonlinear gray-level dependence
	Maximum correlation coefficient (MCC)	Measure of nonlinear gray-level dependence
	Sum average	Measure of the overall image brightness
	Sum entropy	Measure of the randomness of the sum of gray-levels of neighboring pixels
	Sum variance	Measure of the spread in the sum of the gray-levels of pixel-pairs distribution
	Sum of squares (variance)	Measure of the spread in the gray-level distribution

Note.—FWHM = Full Width at Half Maximum, GLCM: Gray-Level Co-occurrence Matrix.

Table E2: List of names, descriptions, and categories of all 45 radiomic features extracted from the parenchyma stroma.

Category	Name	Description
Features based on gray-level and gray-level histogram	Average	Average gray value within region of interest
	95% threshold	gray level threshold yielding 95% of the area under the histogram of the region
	5% threshold	gray level threshold yielding 5% of the area under the histogram of the region
	Balance1	Ratio of (95% threshold-Average) to (Average-5% threshold)
	70% threshold	gray level threshold yielding 70% of the area under the histogram of the region
	30% threshold	gray level threshold yielding 30% of the area under the histogram of the region
	Balance2	Ratio of (70% threshold-Average) to (Average-30% threshold)
	Skewness	The denseness measure used to characterize local tissue composition
Texture features based on NGTDM	Coarseness	Measure of image coarseness calculated from neighborhood gray-tone difference matrix
	Contrast	Contrast measure calculated from neighborhood gray-tone difference matrix
Fractal dimension based on box-counting method	D_{BC} [1–6]	Fractal dimension estimated based on box-counting method
Fractal dimension based on Minkowski method	D_M	Fractal dimension estimated based on Minkowski method
Power law beta	Beta [1–8]	Exponent beta estimated based on power law spectrum analysis
Feature based on edge frequency	MeanEdgeGradient	Average of edge gradient
	MaxEdgeGradient	Maximum edge gradient
	MinEdgeGradient	Minimum edge gradient

	StdDevEdgeGradient	Standard deviation of edge gradient
Features based on Fourier transform analysis	RMS	Root-mean-square variation (RMS) based on Fourier transform analysis
	FMP	First moment of power spectrum (FMP) based on Fourier transform analysis
Texture features based on GLCM	Contrast (GLCM)	Measure of local image variations
	Correlation (GLCM)	Measure of image linearity
	Difference entropy (GLCM)	Measure of the randomness of the difference of neighboring pixels' gray-levels
	Difference variance (GLCM)	Measure of variations of difference of gray-levels between pixel-pairs
	Energy (Angular second moment)	Measure of image homogeneity
	Entropy	Measure of the randomness of the gray-levels
	Inverse difference moment	Measure of the image homogeneity
	Information measure of correlation 1 (IMC1)	Measure of nonlinear gray-level dependence
	Information measure of correlation 2 (IMC2)	Measure of nonlinear gray-level dependence
	Maximum correlation coefficient (MCC)	Measure of nonlinear gray-level dependence
	Sum average	Measure of the overall image brightness
	Sum entropy	Measure of the randomness of the sum of gray-levels of neighboring pixels
	Sum variance	Measure of the spread in the sum of the gray-levels of pixel-pairs distribution
	Sum of squares (Variance)	Measure of the spread in the gray-level distribution

Note.—NGTDM = Neighborhood Gray-Tone Difference Matrix, GLCM = Gray-Level Co-occurrence Matrix, RMS = Root Mean Square, FMP = First Moment of Power Spectrum.

Table E3: Classification performance in the task of distinguishing between malignant and benign lesions using tumor features and combined tumor and parenchyma texture features across different categories of patients.

Cases	Number of tumors	Benign	Malignant	AUC (tumor features)	AUC (tumor & parenchyma features)
All	182	76	106	0.79 ± 0.03	0.84 ± 0.03
BI-RADS score					
BI-RADS 4	99	69	30	0.77 ± 0.05	0.81 ± 0.05
BI-RADS 5	83	7	76	0.76 ± 0.14	0.81 ± 0.10
Age					
<50	62	36	26	0.87 ± 0.05	0.88 ± 0.05
≥50	120	40	80	0.74 ± 0.05	0.83 ± 0.04
Breast density rating					
Fatty (A+B)	99	41	58	0.79 ± 0.05	0.86 ± 0.04
Dense (C+D)	83	35	48	0.78 ± 0.05	0.84 ± 0.04

Note.—BI-RADS = Breast Imaging Reporting and Data System, AUC = Area Under the Curve.