Quantitative texture analysis: robustness of radiomics across two digital mammography manufacturers' systems

Kayla R. Mendel
Hui Li
Li Lan
Cathleen M. Cahill
Victoria Rael
Hiroyuki Abe
Maryellen L. Giger

Quantitative texture analysis: robustness of radiomics across two digital mammography manufacturers' systems

Kayla R. Mendel,* Hui Li, Li Lan, Cathleen M. Cahill, Victoria Rael, Hiroyuki Abe, and Maryellen L. Giger

University of Chicago, Department of Radiology, Chicago, Illinois, United States

Abstract. The robustness of radiomic texture analysis across different manufacturers of mammography imaging systems is investigated. We quantified feature robustness across mammography manufacturers using a dataset of 111 women who underwent consecutive screening mammography on both general electric and Hologic systems. In each mammogram, a square region of interest (ROI) directly behind the nipple was manually selected. Radiomic features describing parenchymal patterns were automatically extracted on each ROI. Feature comparisons were conducted between manufacturers (and breast densities) using newly developed robustness metrics descriptive of correlation, equivalence, and variability. By examining the distribution of these metric values, we propose the following selection criteria to guide feature evaluation in this dataset: (1) $0.8 < \text{mean of feature ratios} < 1.2$, (2) standard deviation of feature ratios < 0.3 , (3) correlation of features (ρ_F) > 0.5 , and (4) $p < 0.05$. Statistically significant correlation coefficients ranged from 0.13 to 0.68 in comparisons between the two mammographic systems tested. Features describing spatial patterns tended to exhibit high correlation coefficients, while intensity- and directionality-based features had comparatively poor correlation. Our proposed robustness metrics may be used to evaluate other datasets, for which different ranges of metric values may be appropriate. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.5.1.011002](https://doi.org/10.1117/1.JMI.5.1.011002)]

Keywords: radiomics; mammography; robustness; quantitative imaging; breast cancer.

Paper 17140SSR received May 15, 2017; accepted for publication Jul. 10, 2017; published online Sep. 19, 2017.

1 Introduction

Breast cancer is the most frequently diagnosed form of cancer in women. Despite improvements in treatment methods, breast cancer remains the second leading cause of death in women.¹ Breast cancer mortality rates have decreased through improved screening, early detection, risk assessment, and advanced treatment.

Computer-aided diagnosis (CAD) has become a routine part of clinical workflow in mammography screening for early detection of breast cancer. Many clinical CAD systems employ radiomic feature extraction for detection and diagnosis of breast tumors.^{2,3} Radiomic features, such as breast density, have been shown to have a role in predicting risk in asymptomatic women, prior to cancer induction.^{4,5} Research studying risk predictors has also included radiomic texture analysis (RTA), which incorporates textural descriptors of the breast parenchyma.^{6–8}

Within quantitative imaging, the goal of RTA is to extract biologically meaningful information, as opposed to modifications caused by the image acquisition process. These types of modifications may be introduced by sources, including image acquisition parameters, differing manufacturers or models, and differing hospital settings. Because these characteristics do not describe patient biology, efforts to standardize and harmonize imaging data can help remove these outside influences. Small datasets with heterogeneities in acquisition are often combined in the curation of large datasets, which are a key component of “big-data” methods, such as machine learning and deep

learning. Heterogeneities in data caused by nonbiological factors are detrimental to the utility of studies in this area. A first step to achieving data harmonization across imaging datasets is to develop an understanding of how some of the imaging variables affect feature calculations. To this end, this study investigates the variation and robustness of RTA features across two equipment manufacturers. Following this investigation, we intend to use the lessons learned here to develop methods of image correction and feature transformation to harmonize large datasets. This harmonization may enable studies that lead to improved cancer risk assessment, detection, and diagnosis in the clinical arena.

2 Materials and Methods

2.1 Data Acquisition and Database

This retrospective study examined full-field digital mammogram (FFDM) images acquired on a screening population of 111 women with low or average risk of breast cancer. Each case had screening mammograms on a General Electric (GE) system and a Hologic system, separated in time by ~ 1 year (mean = 1.29 years and range = 0.86 to 3.07 years). Of the 111 subjects, 104 were imaged first on a GE system and 7 were imaged first on a Hologic system. No breast procedures were performed on subjects between the two studies. At the date of each subject's GE screening, subject ages ranged from 36 to 88 years (mean = 54.0 years, median = 52 years, and standard deviation = 11 years). A description of the

*Address all correspondence to: Kayla R. Mendel, E-mail: kmendel@uchicago.edu

Table 1 Table reports demographics of the study population. The Hologic and GE imaging dates were separated by ~1 year and breast imaging reporting and data system (BI-RADS) density is not always consistent between imaging exam dates, so the ages and breast density score are reported at the time of the GE exam. Data in parentheses are percentages.

Variable	Study population value
Mean age (year)	54.0
Age (year)	
<40	4 (2.70)
40 to 49	40 (36.04)
50 to 59	38 (34.23)
60 to 69	18 (16.22)
70 to 79	8 (7.21)
>80	3 (2.70)
Breast density score	
A	3 (2.70)
B	34 (30.63)
C	62 (55.86)
D	12 (10.81)

study population is included in Table 1. The FFDM images in this study were reviewed by an expert radiologist, and each case was included in this study only if no detectable abnormalities were observed in the images from both the GE and Hologic sets of mammography equipment. All images were acquired from the University of Chicago Medical Center.

The FFDM images used in this study were collected retrospectively under an institutional review board approved, Health Insurance Portability and Accountability Act compliant protocol. One set of images was acquired on a GE FFDM (Senographe 2000D) at 12-bit quantization with a pixel size of $100 \times 100 \mu\text{m}$. Another set of images was acquired on a Hologic FFDM (Lorad Selenia) at 12-bit quantization with a pixel size of $70 \times 70 \mu\text{m}$. Relevant system characteristics are summarized in Table 2.

Regions of interest (ROIs) of size 256×256 pixels were selected manually from the central breast region directly posterior to the nipple in the craniocaudal projection of the mammographic images. Previous studies have shown that the RTA features extracted from this ROI placement perform superiorly on risk assessment tasks.⁹ This workflow is illustrated diagrammatically in Fig. 1.

2.2 Radiomic Feature Extraction

Features describing mammographic parenchymal texture patterns were extracted autonomously from each ROI. These computer-extracted features were based on algorithmic implementations of mathematical models for texture characteristics and have already been reported extensively in the literature. Texture characteristics were based on intensity, spatial pattern, and directionality within each ROI. Specifically, we

Table 2 Summary of several key differences and similarities between the two systems examined in this paper.

Property	GE Senographe	Hologic Selenia
Pixel size	$100 \mu\text{m} \times 100 \mu\text{m}$	$70 \mu\text{m} \times 70 \mu\text{m}$
Quantization	12-bit	12-bit
Anode material	Rhodium	Tungsten
Filter material	Rhodium	Rhodium
Detector size	24×30.7	24×29
Detector material	Amorphous silicon	Amorphous selenium
Conversion method	Indirect	Direct

computationally implemented calculations of (a) gray-level histogram analysis, (b) fractal dimensionality analysis, including the box-counting method and Minkowski method, (c) Fourier and power spectral analysis, (d) edge frequency analysis, and (e) gray-level co-occurrence matrix features, as detailed elsewhere.^{10–13} We categorized each feature computed by these methods based on the texture quality that it described.

Fig. 1 This flowchart illustrates the image analysis and feature extraction process employed in this study.

These categories include intensity, spatial pattern, and directionality measures.

2.3 Evaluation and Statistical Analysis

A goal of this quantitative imaging study was to develop metrics with which to evaluate the robustness of widely applied radiomic texture features as computed from mammograms across varying manufacturers. To that end, we selected four parameters for the assessment of the robustness of these features. These selection metrics were used to separate radiomic features into categories of robust and nonrobust in terms of consistency across imaging manufacturer. The robustness metrics presented in this study included (a) the mean of feature ratios (MFR) to estimate equivalence, (b) the standard deviation of feature ratios (SDFR) to estimate variability, (c) the Spearman's correlation coefficient (ρ_F) to describe correlation, and (d) the statistical significance of the Spearman's correlation coefficient.

2.3.1 Equivalence: mean of feature ratios

The MFR was computed for each specific texture feature by comparing the ratio of the Hologic feature value with the GE feature value for each case and then computing the mean value of these ratios across all pairs of images in the dataset. This calculation is given in Eq. (1). Highly robust features are expected to produce similar values, regardless of the imaging system employed, so a mean of ratios near unity indicates robustness in this regard

$$\text{MFR} = \frac{1}{N} \sum_{i=1}^N \frac{f_{i,\text{Hologic}}}{f_{i,\text{GE}}}. \quad (1)$$

2.3.2 Variability: standard deviation of feature ratios

The variability between feature ratios was also examined to understand how consistent the feature ratio was across cases. To characterize this variation, the SDFR was computed as described in Eq. (2). The ideal feature would have a constant ratio across manufacturers of one for every case, so a low variability near zero suggests robustness

$$\text{SDFR} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{f_{i,\text{Hologic}}}{f_{i,\text{GE}}} - \text{MFR} \right)^2}. \quad (2)$$

2.3.3 Correlation: feature correlation across imaging systems (ρ_F)

The feature correlation across imaging systems was measured using the Spearman's correlation coefficient (ρ_F). The statistical significance of each correlation was measured by the p -value of ρ_F .

Comparisons between these different robustness metrics were drawn, and ranges of acceptable criteria values as indicators of robustness were selected based on the distribution of features in this dataset. These proposed robustness metrics are presented in Table 3.

2.4 Sensitivity to Region of Interest Placement

This study recognizes the inherent limitation that it is not reasonable to expect the possibility of perfect registration across

Table 3 Summary of the figures of merit used to characterize robustness, possible values that each metric may hold, ideal values indicating "perfect" robustness, and cutoff ranges proposed in this study to indicate robustness.

Name	Description	Possible values	Ideal value	Cutoff decision
MFR	Mean of feature ratios	$\text{MFR} \in \mathbb{R}$	1	$0.8 < \text{MFR} < 1.2$
SDFR	Standard deviation of feature ratios	$\text{SDFR} \geq 0$	0	$\text{SDFR} < 0.3$
ρ_F	Spearman's correlation coefficient	$-1 \leq \rho_F \leq 1$	1	$\rho_F > 0.5$
p	p -value of Spearman's correlation coefficient	$p \geq 0$	0	$p < 0.05$

separate mammograms. We investigated the sensitivity of RTA features to ROI placement in terms of feature robustness by examining nonoverlapping subsamples of each ROI. We selected ROIs of size 128×128 pixels from nonoverlapping positions within each original 256×256 pixel ROI. Radiomic features were computed in each of the sub-ROI positions and correlations between sub-ROIs from each individual case were examined, within each manufacturer image. A strong feature correlation indicated robustness to image registration, while a weak feature correlation indicated poor robustness to image registration and therefore high variability dependent on ROI placement.

2.5 Adjustment for Uniform Pixel Size and Region of Interest Area

Due to the differing pixel sizes between the two systems examined in this study, ROIs with matching pixel dimensions will subsequently correspond to different physical areas. For example, the 256×256 ROIs placed on GE images have an area of 2.56×2.56 cm, while those on Hologic images have an area of 1.79×1.79 cm. To explore the effect that differing the physical area of the ROI has on feature calculations, a region of size 1.79×1.79 cm (179×179 pixels) was cropped from the center of each GE ROI. The cropped ROI was then resized to 256×256 pixels using bicubic interpolation. Pixel interpolation was performed in response to the requirement of several features for a power of two image, which would make the calculation of these features on a 179×179 pixel ROI mathematically impossible. RTA features were then calculated on the resized ROIs, and robustness metrics were calculated to compare with the corresponding Hologic ROIs.

3 Results

The MFR, SDFR, ρ_F , and p -value of ρ_F of each robustness measure were computed. These calculations are summarized in Table 4. Through inspection, empirically based ranges in performance criteria for strong features were selected to differentiate robust from nonrobust features. These criteria included (a) $0.8 < \text{MFR} < 1.2$, (b) $\text{SDFR} < 0.3$, (c) $\rho_F > 0.5$, and (d) $p < 0.05$ (statistically significant correlation).

Features that met the criteria for robustness as proposed in this study included six box-counting fractal dimension features, 20 Minkowski fractal-dimension features, four power law

Table 4 Statistics describing the robustness metrics used in this study describing all features calculated in this study. The large range and standard deviation of each robustness metric suggest a wide range in overall robustness of the features examined in this study.

Robustness metric	Minimum	Maximum	Average	Standard deviation
MFR	−5.4	68.5	18.6	23.7
SDFR	0.	81.1	16.9	22.3
ρ_F	−0.1	0.7	0.3	0.2

features, and four gray-level co-occurrence matrix features. The individual features that met the defined criteria for robustness tended to describe spatial patterns, as opposed to directionality or intensity in each image. A list of these features and corresponding values of the robustness criteria is included in Table 5. The same RTA features were found to meet robustness criteria when a subset of the study population of matched BI-RADS density was examined. These results are presented in Table 6.

As illustrated in Fig. 2, features included in this study displayed a diverse range of ρ_F and MFR. By inspection of this

graph, we determined that spatial pattern features produce comparatively highly correlated and highly similar values across manufacturers compared to intensity or directionality features. Differences among the mean of ratios and ratio of means, as shown in Fig. 3, suggest the existence of cases whose ratios fall far from the mean for several given features. Standard error (SE) of the MFR was calculated using bootstrapping with replacement.¹⁴

Our study found that correlation across nonoverlapping ROIs in a single image were, on average, higher than the correlations across manufacturers. The first, second, and third quartiles of feature correlation computed on nonoverlapping regions within a single image are shown in Fig. 4.

To explore whether the observed correlations were decremented by outliers or whether the observed correlations describe the overall relationship, we trimmed the data points included in correlation calculations for each feature by removing any point that lay in the first or fourth quartile in each of the Hologic and GE system datasets. Correlations were computed on the retained data points, and values of the ρ_F robustness metric were improved in 26 out of 217 features. Each of the 26 features that produced an increased correlation coefficient was defined as nonrobust by our previously described metrics, and the correlation coefficient did not increase by enough to move any of these features into the robust category. An examination of a subset of robust and nonrobust features with illustrative cases of

Table 5 Summary of features found to meet the robustness criteria proposed in this paper. The mean of each feature on the GE and Hologic ROIs is reported separately. The SE reported on the MFR was calculated using replacement bootstrapping (100 samples).

Feature name	μ_{GE}	$\mu_{Hologic}$	MFR $\pm$ SE	SDFR	ρ_F	p-value
Box-counting dimension (all eight points)	2.8	2.9	1.040 $\pm$ 0.002	0.04	0.5	<0.001
Box-counting dimension (first four points)	2.8	2.9	1.016 $\pm$ 0.002	0.03	0.7	<0.001
Box-counting dimension (first six points)	2.7	2.8	1.025 $\pm$ 0.002	0.03	0.6	<0.001
Box-counting dimension (first three points)	3.0	3.0	1.008 $\pm$ 0.002	0.03	0.7	<0.001
Box-counting dimension (last four points)	2.6	2.7	1.037 $\pm$ 0.004	0.05	0.6	<0.001
Box-counting dimension regression y-intercept	−0.8	−0.9	1.157 $\pm$ 0.012	0.17	0.5	<0.001
Global Minkowski fractal dimension*	2.5	2.5	1.008 $\pm$ 0.001	0.02	0.7	<0.001
Minkowski minor axis diameter	9.5	11.2	1.179 $\pm$ 0.008	0.12	0.6	<0.001
Powerlaw beta (fine binning log of average)	2.0	1.9	0.972 $\pm$ 0.006	0.10	0.5	<0.001
Powerlaw beta (fine binning average of log)	1.9	1.9	0.973 $\pm$ 0.007	0.09	0.5	<0.001
Powerlaw beta (coarse binning log of average)	2.4	2.3	0.949 $\pm$ 0.005	0.08	0.6	<0.001
Powerlaw beta (coarse binning average of log)	2.4	2.3	0.952 $\pm$ 0.005	0.08	0.6	<0.001
GLCM correlation	1.0	0.9	0.979 $\pm$ 0.003	0.04	0.6	<0.001
GLCM IMC 1	−0.4	−0.3	0.863 $\pm$ 0.014	0.21	0.6	<0.001
GLCM IMC 2	1.0	0.9	0.980 $\pm$ 0.003	0.04	0.6	<0.001
GLCM maximum correlation coefficient	1.0	0.9	0.978 $\pm$ 0.002	0.04	0.6	<0.001

*In addition to Global Minkowski fractal dimension, 18 angular Minkowski fractal dimension features, each at 20-deg intervals, were found to meet the robustness criteria. These features are omitted from this table for concision.

Table 6 Summary of robustness metrics calculated on subsets of the full data population separated by radiologist-assigned BI-RADS density category. Due to limited quantities of data in these subsets, ρ_F and the corresponding p -value are not reported. The number of cases in each density category is reported, and the left and right breasts of each case were included in the calculation of robustness metrics.

Feature name	BI-RADS density A $n = 3$		BI-RADS density B $n = 34$		BI-RADS density C $n = 62$		BI-RADS density D $n = 12$	
	MFR	SDFR	MFR	SDFR	MFR	SDFR	MFR	SDFR
Box-counting dimension (all eight)	1.01	0.05	1.03	0.04	1.04	0.05	1.03	0.03
Box-counting dimension (first four)	1.01	0.02	1.02	0.03	1.01	0.04	1.03	0.04
Box-counting dimension (first six)	1.01	0.02	1.03	0.03	1.02	0.04	1.03	0.03
Box-counting dimension (first three)	1.00	0.02	1.01	0.03	1.00	0.04	1.03	0.05
Box-counting dimension (last four)	0.99	0.01	1.04	0.05	1.04	0.05	1.02	0.04
Box-counting dimension regression y -intercept	1.02	0.16	1.14	0.17	1.15	0.19	1.11	0.09
Global Minkowski fractal dimension*	1.00	0.01	1.01	0.01	1.01	0.02	1.01	0.02
Minkowski minor axis diameter	1.17	0.16	1.20	0.14	1.16	0.13	1.18	0.09
Powerlaw beta (fine binning log of average)	1.03	0.04	0.96	0.09	0.99	0.12	0.93	0.10
Powerlaw beta (fine binning average of log)	1.04	0.04	0.97	0.10	0.99	0.11	0.93	0.10
Powerlaw beta (coarse binning log of average)	1.00	0.04	0.93	0.08	0.97	0.10	0.93	0.08
Powerlaw beta (coarse binning average of log)	1.00	0.03	0.94	0.08	0.96	0.09	0.93	0.08
GLCM correlation	0.99	0.03	0.98	0.04	0.98	0.05	0.98	0.03
GLCM IMC 1	1.01	0.16	0.86	0.19	0.89	0.27	0.85	0.18
GLCM IMC 2	0.99	0.03	0.98	0.04	0.99	0.05	0.98	0.03
GLCM maximum correlation coefficient	0.98	0.03	0.98	0.04	0.98	0.04	0.98	0.03

*In addition to Global Minkowski Fractal Dimension, 18 angular Minkowski Fractal Dimension features, each at 20° intervals, were found to meet the robustness criteria. These features are omitted from this table for concision.

good feature agreement and poor feature agreement is provided in Fig. 5.

In the investigation of the effect of physical area on feature robustness, we found that, in general, RTA features did not agree better across Hologic and GE images when postprocessing is performed to force agreement in physical dimensions of the ROI and in pixel size. In this comparison, out of 34 features that were robust on the original ROI set, only 25 features were robust on the area-matched ROI set. This indicates, as expected, that added layers of postprocessing, including pixel resizing, impact the robustness of RTA features.

4 Discussion

In this study, we chose to investigate feature robustness across different manufacturers of mammography equipment. We quantified heterogeneities in feature values caused by the variation of this single parameter in the image acquisition process. This examination is intended to represent a first step in motivating a broader look at the quantitative effects that various steps of the image acquisition process have on extracted features and the role that this may play in CAD, evaluation of response to therapy, and other quantitative image analysis.

Our interest was in investigating measures of feature robustness across images acquired under different imaging parameters and with different imaging equipment. As such, we chose to use descriptors of equivalence, variability, and correlation in judging the agreement in feature values across acquisition on different manufacturers. Ideally, robust features would have perfect equivalence across manufacturers. However, we accept that, due to many factors, some of which are listed in Table 2, this is highly unlikely. Therefore, we expanded our criteria to assess the correlation between feature values as produced by the two different manufacturers. We justified this because, if features are not equivalent, they still may be used to perform comparisons by applying a transformation to features on one set of the data.

Several factors affect the values of features examined in this study. As explored in this paper, the registration of the ROI plays a substantial role in the computation of feature values. This was illustrated by the wide range of correlation coefficients for features calculated in adjacent but not overlapping ROIs in a single image. The effect of image acquisition parameters was removed by comparing correlation coefficients in a single image. Therefore, any deviation in feature value within a single image at different points demonstrates sensitivity of that feature

Fig. 2 This scatter plot shows the correlation and mean of ratios of various features examined in this study. The color of each point shows the categorical basis of that feature. Solid black lines show the ranges of the robustness selection criteria employed in this study. Most features with high correlation and mean of ratios near a value of one tend to be spatial pattern features, as opposed to intensity- or directionality-based features. Graph (a) is scaled to show all features included in this study and (b) is the same plot reproduced with a narrower scale to show in more detail the robustness of features included.

to ROI placement. For features that are highly sensitive to ROI placement, any differences in feature placement across images from different manufacturers would produce changes leading to induction of heterogeneity in the dataset. Li et al.⁹ have shown that changing the breast region in which an ROI is placed can

significantly decrease the utility of texture features in the task of assessing risk.⁹ In future studies, we will study the effect of much smaller variations in ROI placement.

Feature values are also likely sensitive to characteristics of the mammography equipment and postprocessing. While pixel

Fig. 3 This plot shows the mean of ratios and ratio of means of each feature investigated in this study. Features with a low number of outliers have similar values for the ratio of means and mean of ratios, while features with data deviating from a uniform pattern tended to have greater values for the mean of ratios compared with the ratio of means. This observation motivated the use of the mean of ratios, as opposed to the ratio of means, in the statistical analysis involved in this study.

Fig. 4 The box plot shows the effect of spatial registration on radiomic features extracted from each GE and Hologic FFDM image. The blue boxes indicate the first and third quartiles of feature correlations when 64×64 pixel sub-ROIs of nonoverlapping location in a single image are compared. The red horizontal lines represent the median of feature correlations.

Fig. 5 This figure illustrates example ROIs for the case of good and poor agreements across robust and nonrobust features. Robust features are those that met the robustness criteria proposed in this paper on a population level. Good and poor agreements were defined on the patient level, looking at the agreement of a single feature value on two images of a single patient.

size is accounted for in the feature computation algorithm, pixel size may affect the values of different features. This is because larger pixels do not show finer detail and some features that compute feature on a finer scale may lose this information.

A limitation in this study was the use of “for presentation” as opposed to raw “for processing” images. The images used in this

study were processed by manufacturer-defined algorithms prior to feature extraction, and the algorithms used were not held constant over cases because of the clinical workflow. We intend to focus future work on studying the robustness of features on a single manufacturer over iterations of processing algorithms. This may prove useful for CAD systems and guiding the

evolution of employed features as processing algorithms evolve. While the systems used in this study are no longer state of the art, they are sufficient for presenting new metrics for use in assessing robustness. We plan to extend this work to examine other manufacturers and additional models of mammography equipment. Additionally, this study was limited by the small size of data included. We are continuing efforts to collect additional data, which may improve the generalizability of the results presented in this paper.

It is important to also note that scans on GE and on Hologic machines were separated in time by about 1 year, as each was part of normal screening for the respective patient. It has been well-documented that parenchymal patterns, including density, change over a woman's lifespan.¹⁵ As we did not obtain scans on separate sets of mammography equipment on the same date, this factor could not be eliminated from our study. However, it would be interesting to explore the quantitative effect of age on the feature values explored in this study.

5 Conclusion

The field of radiomics depends on large datasets to draw robust conclusions about the detection, diagnosis, and therapy response of various diseases. One method by which large datasets might be produced is by combining smaller datasets acquired under different parameters, such as with different imaging equipment from different manufacturers. In doing so, there is the innate assumption that the differences between radiomic features are due to the patient biology and not differences in image acquisition. Therefore, this study aimed to draw conclusions as to which texture-based radiomic features are robust to the merging of datasets acquired on different mammography equipment. In this study, we proposed a set of robustness metrics including (a) the MFR to estimate equivalence, (b) the SDFR to estimate variability, (c) the Spearman's correlation coefficient (ρ_F) to describe correlation, and (d) the statistical significance of the Spearman's correlation coefficient to describe the robustness of radiomic features. These metrics were applied to FFDM texture features derived from cases imaged on two digital mammography systems from different manufacturers and used to differentiate robust from nonrobust features in this regard.

By characterizing the robustness of RTA features across heterogeneous image sets, we hope to lay the groundwork for future efforts to develop methods of standardization and harmonization of data. This stands to produce more homogeneous datasets and further improve big data studies and the implementation of machine learning in quantitative imaging.

Disclosures

M.L.G. is a stockholder in R2 Technology/Hologic and a cofounder and shareholder in Quantitative Insights. M.L.G. and H.L. receive royalties from Hologic, GE Medical Systems, MEDIAN Technologies, Riverain Medical, Mitsubishi, and Toshiba. It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interest that would reasonably appear to be directly and significantly affected by the research activities.

Acknowledgments

This work was supported, in part, by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (NIH) under Grant No. T32 EB002103

and the National Cancer Institute of the NIH under Grant Nos. NIH QIN U01 CA195564 and U01 189240.

References

1. A. Jemal et al., "Cancer statistics, 2008," *CA Cancer J. Clin.* **58**, 71–96 (2008).
2. R. M. Rangayyan, F. J. Ayres, and J. E. Desautels, "A review of computer-aided diagnosis of breast cancer: the detection of subtle signs," *J. Franklin Inst.* **344**(3–4), 312–348 (2007).
3. K. Ganesan et al., "Computer-aided breast cancer detection using mammograms: a review," *IEEE Rev. Biomed. Eng.* **6**, 77–98 (2013).
4. J. Brisson, C. Diorio, and B. Masse, "Wolfe's parenchymal pattern and percentage of the breast with mammographic densities: redundant or complementary classifications?" *Cancer Epidemiol. Biomarkers Prev.* **12**, 728–732 (2003).
5. H. Li et al., "Pilot study demonstrating potential association between breast cancer image-based risk phenotypes and genomic biomarkers," *Med. Phys.* **41**(3), 031917 (2014).
6. A. Gastounioti, E. F. Conant, and D. Kontos, "Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment," *Breast Cancer Res.* **18**(1), 91 (2016).
7. Z. Huo et al., "Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection," *Med. Phys.* **27**(1), 4–12 (2000).
8. P. Taylor et al., "Measuring image texture to separate 'difficult' from 'easy' mammograms," *Br. J. Radiol.* **67**(797), 456–463 (1994).
9. H. Li et al., "Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location," *Med. Phys.* **31**(3), 549–555 (2004).
10. Z. Huo and D. E. Wolverton, "Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: feature selection," *Med. Phys.* **27**, 4–12 (2000).
11. H. Li, M. L. Giger, and O. I. Olopade, "Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment," *Acad. Radiol.* **14**, 513–521 (2007).
12. H. Li, M. L. Giger, and O. I. Olopade, "Power spectral analysis of mammographic parenchymal patterns for breast cancer risk assessment," *J. Digital Imaging* **21**, 145–152 (2008).
13. H. Li et al., "Comparative analysis of image-based phenotypes of mammographic density and parenchymal patterns in distinguishing between BRCA1/2 cases, unilateral cancer cases, and controls," *J. Med. Imaging* **1**(3), 031009 (2014).
14. B. Efron, "Better bootstrap confidence intervals," *J. Am. Stat. Assoc.* **82**(397), 171–185 (1987).
15. K. Ghosh et al., "Association between mammographic density and age-related lobular involution of the breast," *J. Clin. Oncol.* **28**(13), 2207–2212 (2010).

Kayla R. Mendel is a graduate student in medical physics at the University of Chicago. Her primary research interests include quantitative texture analysis for risk assessment in medical imaging. Her research focuses on the applications of texture descriptors to improve early detection and risk assessment of breast cancer in women, and improvements in screening image analysis.

Hui Li has been working on quantitative imaging analysis on medical images for over a decade. His research interests include breast cancer risk assessment, diagnosis, prognosis, response to therapy, understanding the relationship between radiomics and genomics, and their future roles in precision medicine with both conventional and deep learning approaches.

Li Lan has been working on breast image analysis research at the University of Chicago since 1996. Her research interests include developing user friendly workstations/software packages, database management, and data analysis.

Cathleen M. Cahill: Biography is not available.

Victoria Rael graduated from the University of Chicago in 2017, where she studied biology and biochemistry. She is interested in pursuing graduate school research in cancer biology.

Hiroyuki Abe is a professor of the Department of Radiology at the University of Chicago Medicine. He is a highly experienced breast imager with a strong research track record. His clinical work includes diagnostic interpretation of mammograms, ultrasounds, and MRIs while performing various types of image-guided procedures. He is actively working with medical physicists and clinical colleagues in the translation of methods of acquisition and analysis of breast MRI, ultrasound, and mammographic images.

Maryellen L. Giger is a professor of radiology and medical physics at the University of Chicago. She is a member of the NAE and is a fellow of SPIE, AAPM, AIMBE, and IEEE. She works in the areas of computer-aided diagnosis, quantitative image analysis, radiomics, and imaging-genomics, focusing on novel methods for characterizing breast cancer on mammography, breast CT, ultrasound, and MRI. She has published over 200 peer-reviewed papers.