

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Parenchymal texture analysis in digital mammography: robust texture feature identification and equivalence across devices

Brad M. Keller
Andrew Oustimov
Yan Wang
Jinbo Chen
Raymond J. Acciavatti
Yuanjie Zheng
Shonket Ray
James C. Gee
Andrew D. A. Maidment
Despina Kontos

Parenchymal texture analysis in digital mammography: robust texture feature identification and equivalence across devices

Brad M. Keller,^{a,†} Andrew Oustimov,^{a,†} Yan Wang,^a Jinbo Chen,^b Raymond J. Acciavatti,^a Yuanjie Zheng,^a Shonket Ray,^a James C. Gee,^a Andrew D. A. Maidment,^a and Despina Kontos^{a,*}

^aUniversity of Pennsylvania, Perelman School of Medicine, Department of Radiology, 3600 Market Street, Suite 360, Philadelphia, Pennsylvania 19104, United States

^bUniversity of Pennsylvania, Perelman School of Medicine, Department of Biostatistics and Epidemiology, 3600 Market Street, Suite 360, Philadelphia, Pennsylvania 19104, United States

Abstract. An analytical framework is presented for evaluating the equivalence of parenchymal texture features across different full-field digital mammography (FFDM) systems using a physical breast phantom. Phantom images (FOR PROCESSING) are acquired from three FFDM systems using their automated exposure control setting. A panel of texture features, including gray-level histogram, co-occurrence, run length, and structural descriptors, are extracted. To identify features that are robust across imaging systems, a series of equivalence tests are performed on the feature distributions, in which the extent of their intersystem variation is compared to their intrasystem variation via the Hodges–Lehmann test statistic. Overall, histogram and structural features tend to be most robust across all systems, and certain features, such as edge enhancement, tend to be more robust to intergenerational differences between detectors of a single vendor than to intervender differences. Texture features extracted from larger regions of interest (i.e., >63 pixels²) and with a larger offset length (i.e., >7 pixels), when applicable, also appear to be more robust across imaging systems. This framework and observations from our experiments may benefit applications utilizing mammographic texture analysis on images acquired in multivendor settings, such as in multicenter studies of computer-aided detection and breast cancer risk assessment. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.2.2.024501]

Keywords: digital mammography; parenchymal pattern; robust texture features.

Paper 14107PR received Aug. 13, 2014; accepted for publication Mar. 13, 2015; published online Apr. 3, 2015.

1 Introduction

Breast cancer is the most commonly diagnosed cancer in American women after skin cancer, with about one in eight expected to develop invasive breast cancer over the course of their lifetime.¹ With the advent of early cancer detection through screening and treatment, the five-year survival rate for women diagnosed with breast cancer in western countries is currently up to 89%.² Approximately 95% of screening centers in the United States utilize full-field digital mammography (FFDM) systems as their primary screening tool,³ which rapidly replaced screen-film mammography over the last decade due to advantages of FFDM systems, such as improved image contrast and dynamic range.⁴ Mammographic images have extensively been used in the literature to generate image-based features for examining the potential risk for breast cancer.^{5–20} For example, mammographic percent density (PD%)^{5,6,15} has been repeatedly shown to be a strong risk factor for breast cancer. In addition, it has also been suggested in more recent studies that parenchymal texture features,^{7,8} which are measures of the local distribution of the parenchymal tissue, may also provide complementary information regarding an individual woman's risk for breast cancer.

For studies analyzing parenchymal texture features for breast cancer risk assessment, one additional factor that needs to be accounted for is the mammography system used for image acquisition. With the wide-scale implementation of screening programs utilizing FFDM systems, large clinical data sets typically include images acquired from several commercial FFDM systems built with different detector technologies and image acquisition parameters. Differences between x-ray systems introduce inherent differences to the acquired images, which subsequently affect extracted mammographic features, such as breast density²¹ and, as a result, higher-order measures of breast parenchymal texture. This effect can be more pronounced for more granular type of features that directly depend on the variation of image intensity values, such as texture features, compared to more global image measures, such as breast PD%. For example, Manduca et al.⁸ noted that accounting for the parameters used to acquire images from different mammography systems is important, specifically with regards to affecting the numerical values of texture features. This could, in turn, bias any subsequent investigation of associations between such features and any related outcomes of interest, such as the risk of developing breast cancer.

Motivated by this problem, we propose a statistical framework to evaluate the robustness of texture features across FFDM

*Address all correspondence to: Despina Kontos, E-mail: Despina.Kontos@uphs.upenn.edu

†Co-first authors

systems from different vendors, focusing on the effect of different x-ray systems on image intensity and texture. We use a physical breast phantom (Gammex 169 “Rachel,” Madison, Wisconsin), which enables the imaging of a consistent (i.e., ground truth) parenchymal tissue pattern across different systems, allowing us to directly evaluate any added system-induced effects on image intensity and texture. To demonstrate proof-of-concept and be applicable to the clinical setting, we compare three FFDM systems (Senographe 2000D and Senographe DS, General Electric Medical Systems, Milwaukee, Wisconsin; and Selenia Dimensions, Hologic Inc., Bedford, Massachusetts) operated at the clinical automated exposure control setting and design a series of experiments to evaluate whether the distributions of a panel of different texture features are robust and statistically equivalent across these different imaging systems. While our framework is presented here using three specific imaging systems from two vendors and a selected set of texture features previously used for breast cancer risk assessment, our method can, in principal, generalize across a broader range of image acquisition systems and features. As such, it could have value for studies utilizing mammographic feature analysis in a multivendor setting, such as studies focusing on breast cancer risk assessment, breast tissue characterization, and computer-aided diagnosis.

2 Materials and Methods

2.1 Breast Phantom Imaging

An anthropomorphic breast phantom²² (Gammex 169 “Rachel,” Madison, Wisconsin) was imaged on three FFDM systems. The first two systems, the Senographe 2000D and DS, utilize equal-sized, indirect flat panel detectors with a common spatial resolution of 0.1 mm/pixel, although the GE DS has a smaller x-ray tube relative to GE 2000D. In contrast to these systems, the Selenia Dimensions utilizes an amorphous selenium direct flat panel detector with a higher spatial resolution of 0.07 mm/pixel. Last, the three units have different automated exposure control systems, resulting in different dose performances.

To account for system noise effects and better assess inter-system differences, digital mammograms of the Rachel phantom were acquired five times on each FFDM unit, with no physical movement of the phantom during the image acquisition process. The raw (FOR PROCESSING) images stored at the common vendor-specified 14-bit gray-level depth were used for subsequent analysis. The GE systems utilized an Rh/Rh target/filter combination and the Hologic system utilized a W/Rh combination. Acquisition parameters were kVp = 29 and mAs = 71 for GE 2000D, kVp = 29 and mAs = 90 for the GE DS system, and kVp = 31 and mAs = 160 for the Hologic Selenia Dimensions, when using the automated exposure control setting for each model digital mammography machine.

2.2 General Statistical Framework for Robust Feature Identification

For the purposes of our study, we define a robust texture feature to be a feature for which the distributions of the corresponding feature images from a given pair of systems are statistically equivalent. Specifically, given a texture feature, the process to identify whether it is robust can be divided into three main steps (Fig. 1). First, a series of phantom images are acquired from all mammography systems and the breast region is segmented.

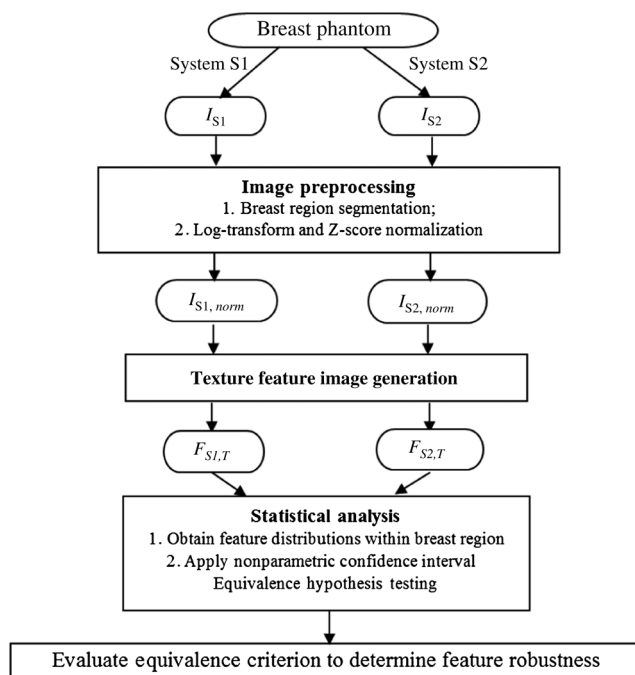


Fig. 1 Workflow of our statistical framework for texture feature equivalence assessment. I_S , digital mammography image acquired from system S ; $I_{S, norm}$, I_S after image intensity normalization; $F_{S, T}$, texture feature image for system S , texture feature T .

Image intensity values are first log-transformed to account for the relationship between raw gray-level values and exposure, then inverted so that radio-opaque foreground objects (i.e., the breast) are of a higher pixel intensity than the background (i.e., air), and, finally, are normalized by a z-score transformation²³ within the breast region. Briefly, z-score normalization is a linear transformation of image pixels, or more generally any distribution, so that the average value of the distribution of image pixels is made to be zero and the standard deviation of the distribution of values is one. In practice, z-score normalization has the effect of aligning different histograms of different images without altering the shape of their distribution.²⁴

For each log-transformed, z-score normalized image, texture features are generated using automated computerized analysis^{25,26} from each of the five images acquired on each digital mammography system, generating a set of five pixel-value distributions per each texture-parameterization-system combination. Next, these distributions are compared across acquisition systems via a series of nonparametric equivalence tests,²⁷ testing the hypothesis that the median of the differences in texture feature values across acquisition systems is within a prespecified equivalence range.²⁷ Details of the procedure are provided in Sec. 2.6. All statistical analysis in this study was performed with SAS 9.4 statistical computing software.

2.3 Image Preprocessing

As the region occupied by the breast phantom covers only a fraction of the total image field [Fig. 2(a)], we perform a preprocessing step to (1) remove the phantom bounding box from any subsequent analysis and (2) segment the breast region [Figs. 2(b) and 2(c)]. The bounding box is cropped based on the detection of the box boundary [shown in Fig. 2(a)]. Given the differences in the edge response and spatial resolution

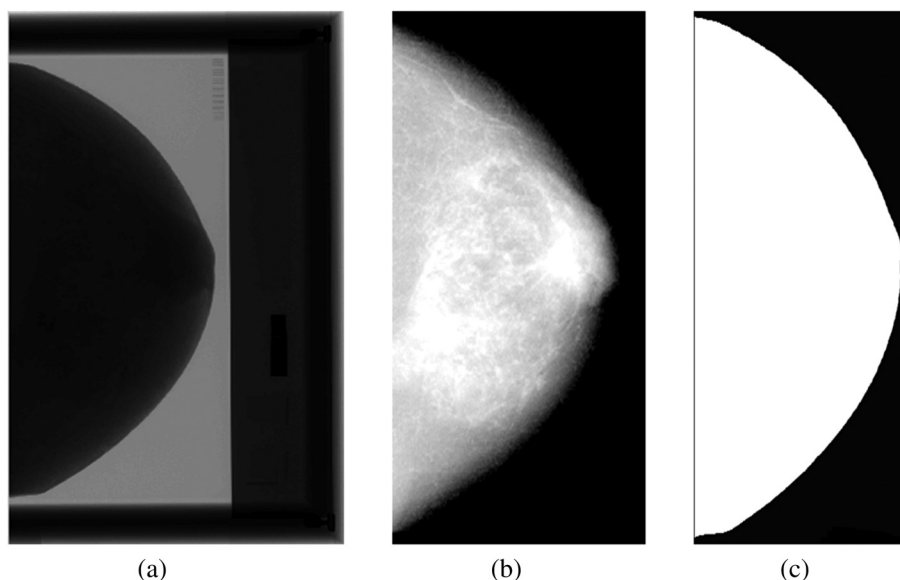


Fig. 2 Raw digital mammogram of the Rachel phantom acquired on GE 2000D: (a) the original raw phantom image, (b) exclusion of the breast phantom bounding box, shown after the inverse-log transformation for visualization, and (c) the binary mask of the breast region.

between the three systems, the breast region is segmented by using intensity thresholding in each phantom image, which is then used for the analysis of all images acquired by that system.

2.4 Texture Feature Generation

We used an automated image analysis pipeline²⁵ that can generate a corresponding texture feature image for each texture descriptor. Specifically, we use a lattice-based strategy, where at each lattice point [shown in blue in Fig. 3(a)], a texture feature value is calculated within the local square region [shown in red in Fig. 3(a)] surrounding the lattice point. Both the lattice size L (i.e., the distance between neighborhood lattice points) and the local window size W are parameters that can be optimized for texture feature generation. In our study, the lattice size and

the local window size are set to be equal. Five different window sizes W , equal to 15, 31, 63, 127, and 255 pixels², were evaluated to capture parenchymal texture patterns at different granularities.

The texture features used in our study can be broadly categorized into four groups: (1) gray-level histogram features;²⁸ (2) co-occurrence features;^{29,30} (3) run-length features;^{31,32} and (4) structural features.^{20,33–35} All features and their corresponding shorthand notations are summarized in Table 1 (detailed mathematical notations are also included in the Appendix). These features were selected in our study for two main reasons. First, most of these features have been widely used for mammographic image analysis and shown to have value specifically for breast tissue characterization in breast cancer risk assessment.^{17,34,36} Second, features capturing the local tissue architecture, such as the local binary pattern (LBP)³³ and the edge enhancing index,^{35,37} are relatively more recent texture features. Traditionally, these features have been used in computer vision,^{33,37,38} while recently they have also been applied for mammographic pattern analysis and breast cancer risk assessment.³⁵

2.5 Parameters in Texture Feature Generation

These texture features are computed using certain parameters, summarized into two main categories: (1) parameters that are involved in the numerical implementation and can be decided depending on numerical precision, computational cost, and related factors and (2) parameters introduced in the mathematical computation directly related to inherent properties captured by the feature (e.g., co-occurrence pixel distance, etc.).

In general, there is no single best choice for parameters within the first category, due to trade-offs between precision and efficiency. Features in groups 1 to 3 in Table 1 all depend upon such a parameter, specifically the number of histogram bins. In this study, this parameter was set to a fixed value of 128 based on previous studies.²⁸ For features with parameters in the second category, we describe in detail our approach for the co-occurrence, LBP, and edge enhancing index features.

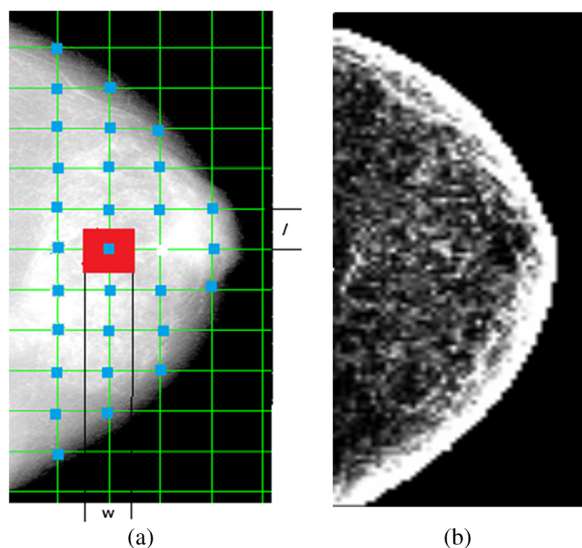


Fig. 3 (a) The lattice scheme for texture image generation and (b) a gray-level histogram (STD) image.

Table 1 A list of all texture features analyzed in our study.

Feature class	Feature name	Notation
Group 1 Gray-level histogram	Max	MAX
	Min	MIN
	Mean	MEAN
	Sum	SUM
	Entropy	ETP
	Kurtosis	KTS
	Sigma	STD
	Skewness	SKEW
	5th percentile	5TH
	5th mean	5THM
	95th percentile	95TH
	95th mean	95THM
	Group 2 Co-occurrence	Cluster shade
Energy		ENG
Entropy		CETP
Inertia		INT
Correlation		COR
Haralick correlation		HCOR
Inverse difference moment		IDM
Group 3 Run length	Grey-level nonuniformity	GLN
	Run-length nonuniformity	RLN
	Run percentage	RP
	High gray-level run emphasis	HGLRE
	Long run emphasis	LRE
	Low gray-level run emphasis	LGLRE
	Short run emphasis	SRE
Group 4 Structural features	Local binary pattern	LBP
	Edge enhancing index	EEI
	Fractal dimension	FD

Co-occurrence features describe the spatial relationship between pixels and are based on the gray-level co-occurrence matrix (GLCM).²⁹ Given an image f , an N by N GLCM matrix $G = g(i, j)$ is defined as

$$g(i, j) = |\{(x_1, y_1), (x_2, y_2) | f(x_1, y_1) = i, f(x_2, y_2) = j, x_2 - x_1 = l |\cos \theta|, y_2 - y_1 = l |\sin \theta|\}|, \quad (1)$$

where N is the number of gray levels used to describe image f , l is the pixel offset length, and θ is the offset angle. N is commonly chosen as a power of 2, typically between 16 and 256. In our experiments, N was chosen equal to 128 to balance computational precision with efficiency. The combination of parameters l and θ determines the search direction and the neighborhood size of each pixel pair. Features are typically calculated using the same offset length l along four directions (0, 45, 90, and 135 deg) and then averaged, based on the assumptions that these features are orientation invariant^{39,40} and that one single direction might not give sufficient texture information.³⁰ With respect to the offset length l , in several studies, it is chosen by default to be equal to 1 pixel, however, this is not always justified based on its actual physical dimensions.^{7,41} Here we theorize that the proper selection of l is important for the robustness of the corresponding texture features, as it may be associated with the blurring effect, and therefore, artificially introduced texture, in FFDM systems using an indirect-conversion detector. To evaluate such effects, we varied the offset length l from 1 to 13 pixels (i.e., 0.1 mm up to 1.3 mm) by using 2-pixel increments.

LBP^{33,38} capture the relationship between central and neighboring pixels. Given an image f , the LBP $_{P,R}$ value at pixel (x_c, y_c) is defined as

$$\text{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, \\ (x_p, y_p) = \left[x_c + R \cos\left(\frac{2\pi p}{P}\right), y_c - R \sin\left(\frac{2\pi p}{P}\right) \right], \\ g_c = f(x_c, y_c), \quad g_p = f(x_p, y_p), \quad p = 0, \dots, P-1, \\ s(x) = 1, \quad x \geq 0; \quad s(x) = 0, \quad x < 0. \quad (2)$$

The two parameters, neighborhood size R and number of neighborhood pixels P , define the LBP.¹⁸ Previous studies have shown that the performance of LBP features in pattern detection varies with the change of parameters R and P . The common choices of R in previous studies in computer vision include 1, 2, and 3 pixels.³³ In our study, three pairs of (R, P) values were used, namely (1, 8), (2, 20), and (3, 36) pixels.

Finally, the edge enhancing index is based on edge enhancing diffusion,⁴² which preserves and enhances edge structures within an image. The flow-like structures within the breast tissue suggest that this type of feature could be an appropriate descriptor for characterizing the structural properties of the parenchymal pattern.

Given an image f , the edge enhancing index is defined as

$$E_\sigma(f) = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \eta} \right)^2. \quad (3)$$

Here, λ_1 and λ_2 with $\lambda_1 > \lambda_2$ are eigenvalues of the diffusion tensor matrix D defined as

$$D = \frac{1}{(\partial_x f_\sigma)^2 + (\partial_y f_\sigma)^2} \\ \times \begin{bmatrix} c_1 (\partial_x f_\sigma)^2 + c_2 (\partial_y f_\sigma)^2 & (c_2 - c_1) \partial_x f_\sigma \partial_y f_\sigma \\ (c_2 - c_1) \partial_x f_\sigma \partial_y f_\sigma & c_1 (\partial_y f_\sigma)^2 + c_2 (\partial_x f_\sigma)^2 \end{bmatrix} \\ f_\sigma = f * G_\sigma, \quad c_2 = e^{-[(\partial_x f_\sigma)^2 + (\partial_y f_\sigma)^2]/k^2}, \quad c_1 = \frac{1}{5} c_2$$

When $\lambda_1 \approx \lambda_2$, $E_\sigma \approx 0$, and when $\lambda_1 \gg \lambda_2$, $E_\sigma \approx 1$ and η is an empirical normalizing factor set equal to 10 for this study. The parameter evaluated in our study is the Gaussian kernel size σ . Gaussian smoothing (i.e., $f * G_\sigma$) is used as a common image preprocessing step to remove image noise before image analysis and the kernel size determines the degree of smoothing (i.e., the extent of image details that are preserved). In our study, G_σ was varied from 1 to 15 pixels.

2.6 Statistical Analysis

To assess the robustness of texture features across any two FFDM systems (i.e., pair-wise comparison), the distributions of texture feature values were compared via a series of nonparametric equivalence tests, testing the hypothesis that the median of the differences, Δ , in texture feature values across the two acquisition systems is within an equivalence range equal to the maximum intrasystem variation of the two systems being compared pair-wise. For each system, the intrasystem variation was determined by the maximum 95% confidence interval (CI) resulting from the $\binom{5}{2}$ intrasystem comparisons. The test statistic Δ is computed as follows: given observations from two independent acquisition systems, x_i for $i = 1, \dots, n$ and y_j for $j = 1, \dots, m$, respectively, a difference Z_{ij} is calculated between each pair (x_i, y_j) . There are $m * n$ such differences. The test statistic Δ is equal to $\text{median}\{Z_{ij}\}$ and is equivalent to the Hodges–Lehmann statistic.⁴³

Intra- and intersystem comparisons for each different window size utilized different numbers of observations. Comparisons of values derived from the larger window sizes (e.g., 255 and 127 pixels²) resulted in smaller sample sizes. This is because as the lattice size increases, larger amounts of the image (in intensity space) are represented by each pixel in feature space, thus resulting in a smaller number of feature space pixels needed to represent the entire image. These smaller sample size comparisons were carried out via a nonparametric equivalence test procedure in which data-derived confidence intervals were utilized.²⁷ The derivation of these confidence intervals can be briefly described as follows. First, an empirical distribution for the independent difference Z_{ij} is computed as $\hat{F}(z) = (1/mn) \sum_{ij} C(z - Z_{ij}) = (1/mn)S(z)$, where $c(z) = 1, 1/2, 0$ for $z > 0, z = 0, z < 0$, respectively. General U-statistic theory implies the asymptotic standard normality of the statistic, $T(\Delta) = [\hat{F}(\Delta) - (1/2)] / \left\{ \sqrt{\hat{V}[F(\Delta)]} \right\} = [S(\Delta) - (mn/2)] / \left\{ \sqrt{\hat{V}[S(\Delta)]} \right\}$ for any Δ with $F(\Delta) = (1/2)$. Confidence interval bounds are then defined as

$$CL_L = \sup\{\Delta | T(\Delta^*) \leq (-u_\alpha) \forall \Delta^* < \Delta\}, \quad (4)$$

$$CL_U = \inf\{\Delta | T(\Delta^*) \geq (+u_\alpha) \forall \Delta^* > \Delta\}, \quad (5)$$

where u_α denotes the $(1 - \alpha)$ percentile of the standard normal distribution. Computing the distribution of the independent differences directly from the data and utilizing them to build the CIs for the test statistic allows us to assess distributional differences without making any assumptions regarding their shape. Finally, comparisons for the smaller window sizes (i.e., 63, 31, and 15 pixels²) resulted in much larger sample sizes rendering the empirical procedure computationally intractable,

thus, asymptotically derived CIs for the Hodges–Lehmann statistic⁴³ were utilized.

Intersystem robustness was assessed according to the following four category scale: (1) features for which the intermachine CI was entirely contained within the specified equivalence bounds were deemed very robust (VR), implying that the variation between the systems being compared is not significantly different from the intrasystem variation; (2) cases in which the intermachine CIs were not fully contained within the equivalence bounds, but still crossed $\Delta = 0$ [indicating that the test statistic is not significantly (at $\alpha = 0.05$) different from zero, thus implying no significant difference in texture feature distributions across systems] and for which the intermachine test statistic was contained in the intrasystem CI, were deemed robust (R); (3) cases for which only one of the criteria described in (1) or (2) were met were deemed less robust (LR); and (4) cases that did not meet any of the criteria described in (1), (2), and (3) were deemed not robust (NR).

3 Results

Results of the equivalence assessment comparing the distributions of all texture features between each pair of systems are provided, where the degree of feature equivalence between the GE 2000D-GE DS, Hologic Selenia Dimensions-GE 2000D, and Hologic Selenia Dimensions-GE DS comparisons are shown in Tables 2, 3, and 4, respectively.

In general, gray-level histograms and structural features appear to be more robust than co-occurrence and run-length texture features, while smaller window sizes tend to result in fewer robust texture features. Larger window sizes (i.e., ≥ 63 pixels²) appear to provide generally more robust texture features. The majority of structural texture features found to be VR appear in the LBP subcategories of all three tables. The effect of increased window size is most evident in the co-occurrence and run-length features of all three comparisons, but is also apparent in the gray-level histogram features of the GE 2000D-GE DS comparison, which exhibits VR features for window sizes of 127 and 255 pixels². The fractal dimension feature also appears to benefit from increased window size as the majority of the VR features for this section are for the 63 and larger window sizes. The edge enhancing index features are least robust across the Selenia-GE 2000D systems, marginally more robust for the Selenia-GE DS systems, and most robust for the GE 2000D-GE DS comparison. Last, few co-occurrence features were found to be robust when computed at an offset length of 1. Figure 4 provides texture maps and histograms of the inverse difference moment features on the three systems computed at offset length 1 and window size 63, illustrating this finding.

Tables 5–7 show the effect of the offset length parameter, l , used in the computation of the GLCM for the co-occurrence texture features, when a representative window size is fixed at 127 pixels² for the three pair-wise system comparisons. Overall, co-occurrence features tend to be more robust across the Selenia-GE 2000D comparison, where VR designations appear for the majority of features and for all offset lengths. While cluster shade and Haralick correlation are robust across all comparisons and for all offset lengths, the correlation feature is least robust across the Selenia-GE DS comparison where the feature is found to be NR for all offset lengths except 1. Increasing the offset length appears to benefit robustness among the entropy, energy, inertia, and inverse difference moment features, and is most apparent across the Selenia-GE DS and

Table 2 Feature robustness between the GE 2000D and GE DS systems. VR, very robust; R, robust; LR, less robust; NR, not robust.

Feature group	Feature	Parameter	Window size (<i>W</i> , pixel)					
			15	31	63	127	255	
Gray-level histogram	5TH		R	R	R	R	R	
	5THM		NR	R	R	R	VR	
	95TH		R	R	R	VR	VR	
	95THM		R	R	R	VR	VR	
	ETP		NR	NR	R	VR	VR	
	KTS		NR	NR	R	VR	VR	
	MAX		NR	R	R	R	R	
	MEAN		R	R	R	VR	VR	
	MIN		NR	LR	R	R	VR	
	STD		NR	NR	R	R	R	
	SKEW		LR	R	R	VR	VR	
	SUM		R	R	R	VR	VR	
Co-occurrence	CSD	Offset length <i>l</i> = 1 pixel	NR	R	R	R	VR	
	COR		NR	NR	NR	R	R	
	ENG		NR	NR	NR	NR	R	
	ETP		NR	NR	NR	NR	NR	
	HCOR		NR	NR	R	R	VR	
	INT		NR	NR	NR	NR	NR	
	IDM		NR	NR	NR	NR	NR	
	GLN		LR	LR	R	R	R	
Run length	HGLRE		R	NR	R	VR	VR	
	LRE		NR	NR	NR	NR	LR	
	LGLRE		NR	NR	NR	VR	VR	
	RLN		NR	NR	NR	NR	NR	
	RP		NR	NR	NR	NR	LR	
	SRE		NR	NR	NR	NR	LR	
	Structural	LBP	$R = 1, P = 8$	VR	VR	VR	R	R
			$R = 2, P = 20$	VR	R	VR	R	R
$R = 3, P = 36$			VR	R	R	R	VR	
EEI		$R = 1$	NR	LR	R	R	R	
		$R = 5$	R	R	R	VR	NR	
		$R = 10$	R	R	R	R	R	
		$R = 15$	LR	R	R	R	VR	
FD		NR	NR	NR	VR	NR		

Table 3 Feature robustness between the Hologic Selenia Dimensions and GE 2000D systems. VR, very robust; R, robust; LR, less robust; NR, not robust.

Feature group	Feature	Parameter	Window size (<i>W</i> , pixel)					
			15	31	63	127	255	
Gray-level histogram	5TH		VR	VR	VR	VR	R	
	5THM		R	VR	VR	VR	VR	
	95TH		R	R	R	R	R	
	95THM		R	R	R	R	R	
	ETP		NR	NR	NR	R	R	
	KTS		NR	NR	R	R	R	
	MAX		R	R	R	LR	R	
	MEAN		R	R	R	R	R	
	MIN		NR	R	VR	VR	R	
	STD		NR	NR	NR	LR	R	
	SKEW		NR	R	R	VR	VR	
	SUM		R	R	R	R	R	
Co-occurrence	CSD	Offset length <i>l</i> = 1 pixel	NR	LR	NR	R	R	
	COR		NR	R	NR	LR	R	
	ENG		NR	NR	NR	NR	R	
	ETP		NR	NR	NR	NR	LR	
	HCOR		R	NR	NR	R	R	
	INT		NR	NR	NR	NR	NR	
	IDM		NR	NR	NR	NR	NR	
	GLN		NR	NR	NR	R	R	
Run length	HGLRE		NR	R	NR	VR	R	
	LRE		NR	NR	NR	NR	R	
	LGLRE		NR	NR	NR	NR	LR	
	RLN		NR	NR	NR	NR	R	
	RP		NR	NR	NR	NR	NR	
	SRE		LR	NR	NR	NR	NR	
	Structural	LBP	$R = 1, P = 8$	VR	VR	VR	VR	NR
			$R = 2, P = 20$	VR	VR	VR	VR	NR
$R = 3, P = 36$			VR	VR	VR	VR	NR	
EEI		$R = 1$	NR	LR	R	LR	LR	
		$R = 5$	NR	NR	NR	NR	NR	
		$R = 10$	NR	NR	NR	LR	LR	
		$R = 15$	NR	NR	NR	LR	LR	
FD		NR	NR	NR	VR	LR		

Table 4 Feature robustness between the Hologic Selenia Dimensions and GE DS systems. VR, very robust; R, robust; LR, less robust; NR, not robust.

Feature group	Feature	Parameter	Window size (W , pixel)					
			15	31	63	127	255	
Gray-level histogram	5TH		R	R	R	VR	VR	
	5THM		R	R	R	R	R	
	95TH		NR	R	R	R	R	
	95THM		NR	LR	LR	NR	R	
	ETP		NR	NR	NR	VR	VR	
	KTS		VR	R	R	R	R	
	MAX		NR	NR	NR	NR	LR	
	MEAN		VR	R	R	R	R	
	MIN		LR	R	R	R	R	
	STD		NR	NR	NR	NR	R	
	SKEW		NR	NR	R	VR	VR	
	SUM		VR	R	R	R	R	
	Co-occurrence	CSD	Offset length $l = 1$ pixel	NR	NR	R	VR	VR
COR			NR	NR	NR	NR	R	
ENG			NR	NR	NR	NR	R	
ETP			NR	NR	NR	NR	R	
HCOR			NR	NR	NR	LR	R	
INT			NR	NR	NR	LR	R	
IDM			NR	NR	NR	LR	R	
Run length		GLN		NR	NR	NR	R	R
		HGLRE		NR	NR	R	R	VR
		LRE		NR	NR	NR	R	VR
	LGLRE		NR	LR	R	NR	VR	
	RLN		NR	NR	NR	VR	R	
	RP		LR	NR	NR	R	R	
	SRE		LR	NR	NR	R	R	
Structural	LBP	$R = 1, P = 8$	VR	VR	VR	R	LR	
		$R = 2, P = 20$	VR	R	VR	R	R	
		$R = 3, P = 36$	VR	R	VR	R	R	
	EEI	$R = 1$	NR	NR	LR	VR	R	
		$R = 5$	NR	NR	NR	NR	VR	
		$R = 10$	NR	NR	NR	LR	LR	
		$R = 15$	NR	NR	NR	LR	LR	
FD		R	R	VR	VR	LR		

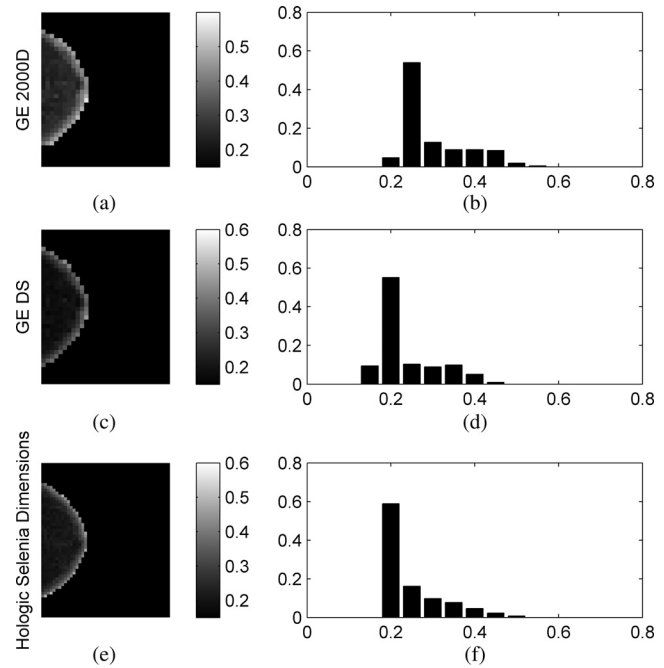


Fig. 4 Example feature maps [(a), (c), and (e)] and feature-value histograms [(b), (d), and (f)] of the co-occurrence inverse difference moment feature computed on the GE 2000D [(a) and (b)], GE DS [(c) and (d)], and Hologic Selenia Dimensions [(e) and (f)] digital mammography systems using a window size of 63 pixels² and offset of $l = 1$ pixel.

GE 2000D-GE DS comparisons. Figure 5 provides texture maps and histograms of the inertia feature on the three systems computed at offset length 7 and window size 127.

4 Discussion

The question of how to most appropriately normalize mammographic images obtained from different vendor systems and acquisition physics parameters for subsequent image analysis and feature extraction is an important active area of research. In our study, z-score normalization was applied to the log-transformed

Table 5 Effect of offset length, l , on feature robustness between the GE 2000D and GE DS systems for the co-occurrence texture features computed at window size 127. VR, very robust; R, robust; LR, less robust; NR, not robust.

Feature		Offset length (l , pixel)							
		1	3	5	7	9	11	13	
Co-occurrence	CSD	R	R	R	R	R	R	R	
	COR	R	R	R	R	R	R	R	
	ENG	NR	NR	NR	R	R	R	R	
	ETP	NR	NR	LR	R	R	R	R	
	HCOR	R	R	R	R	R	R	R	
	INT	NR	NR	NR	R	R	R	R	
	IDM		NR	NR	LR	R	R	R	R
			NR	NR	LR	R	R	R	R

Table 6 Effect of offset length, l , on feature robustness between the Hologic Selenia Dimensions and GE 2000D systems for the co-occurrence texture features computed at window size 127. VR, very robust; R, robust; LR, less robust; NR, not robust.

Feature		Offset length (l , pixel)						
		1	3	5	7	9	11	13
Co-occurrence	CSD	R	R	R	R	R	R	R
	COR	LR	NR	NR	NR	NR	NR	NR
	ENG	NR	R	R	R	LR	LR	LR
	ETP	NR	R	VR	R	R	R	R
	HCOR	R	R	R	R	R	R	R
	INT	NR	LR	VR	VR	VR	VR	VR
	IDM	NR	NR	VR	R	LR	LR	LR

Table 7 Effect of offset length, l , on feature robustness between the Hologic Selenia Dimensions and GE DS systems for the co-occurrence texture features computed at window size 127. VR, very robust; R, robust; LR, less robust; NR, not robust.

Feature		Offset length (l , pixel)						
		1	3	5	7	9	11	13
Co-occurrence	CSD	VR	VR	VR	VR	VR	VR	VR
	COR	NR	NR	NR	NR	NR	NR	NR
	ENG	NR	NR	NR	NR	NR	NR	NR
	ETP	NR	NR	NR	NR	NR	NR	LR
	HCOR	LR	LR	LR	LR	LR	LR	LR
	INT	LR	LR	LR	LR	R	R	R
	IDM	LR	VR	VR	R	R	R	R

images in order to remove linear differences in gray-level intensities between the various FFDM systems evaluated. Both log-transformation and z-score normalization are commonly utilized preprocessing steps when dealing with raw, FOR PROCESSING, mammograms as a means to alleviate interscan differences between studies via histogram alignment, while still maintaining the overall pattern, spatial relationship, and relative contrast of the pixels,²⁴ all of which are important in texture analysis. Our results suggest that the framework proposed in this work will be beneficial for studies utilizing texture analysis in digital mammography in clinical, multisite, multivendor data sets, including studies investigating breast cancer risk assessment, tissue classification, and computer-aided diagnosis.

As our study focuses on the effect of detector differences on texture quantification and not necessarily on classification tasks such as risk assessment using such textures, it is difficult to compare our extracted texture measures directly to prior work, although comparison of parameter selection choices can be reported. In previous studies using co-occurrence features,

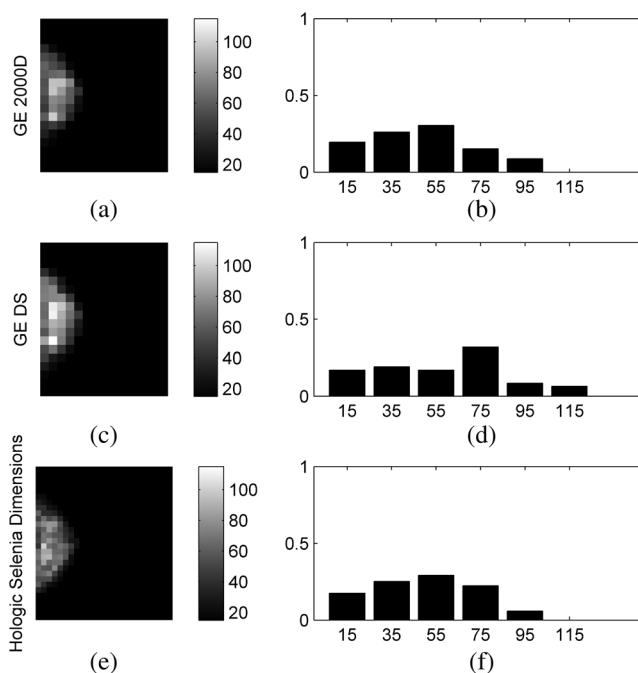


Fig. 5 Example feature maps [(a), (c), and (e)] and feature-value histograms [(b), (d), and (f)] of the co-occurrence inertia (INT) feature computed on the GE 2000D [(a) and (b)], GE DS [(c) and (d)], and Hologic Selenia Dimensions [(e) and (f)] digital mammography systems using a window size of 127 pixels² and offset of $l = 7$ pixels.

the offset length was generally chosen heuristically. For example, in mammographic imaging research, l is often chosen between 1 and 10 pixels with 1 being the most common value, such as in the work by Khuzi et al.³⁰ for mass identification and Li et al.⁷ for texture analysis of mammographic parenchymal patterns. Our results suggest that if the offset length is gradually increased to 7 pixels or larger, the majority of inherent system effects can be alleviated. The effects of the window size W and the offset length l on the texture feature images are shown in Figs. 6 and 7. For smaller window sizes, more granular features are obtained, while for larger windows, more smoothed (i.e., blurred) texture images are generated.

This observation may be partly explained via examination of the modulation transfer function (MTF) of the imaging system,⁴⁴ a measure of its inherent spatial resolution. The MTF is unity at the lowest spatial frequency and drops with increasing spatial frequency. Both of the GE systems in our study use indirect conversion detectors (i.e., cesium phosphor with thin-film transistor). Compared to a direct conversion process, such as the one utilized in the Hologic Selenia Dimensions system evaluated in our study, blurring effects are introduced by phosphor luminescence and can cause loss of the spatial resolution, which may be alleviated by the use of smoothed (i.e., less granular) texture features that do not consider the spatial relation of pixels (e.g., first-order histogram statistics), intensity-invariant structure-based features, and larger lattice window sizes. Furthermore, when taking this blurring effect into consideration, the differences between texture images for the two GE systems, and between the GE and Hologic systems, can potentially be enhanced within relatively small pixel neighborhood sizes, as pixels within such a small neighborhood can be highly spatially correlated, which could be the case when choosing a small offset length (e.g., $l = 1$ pixel) in the co-occurrence texture

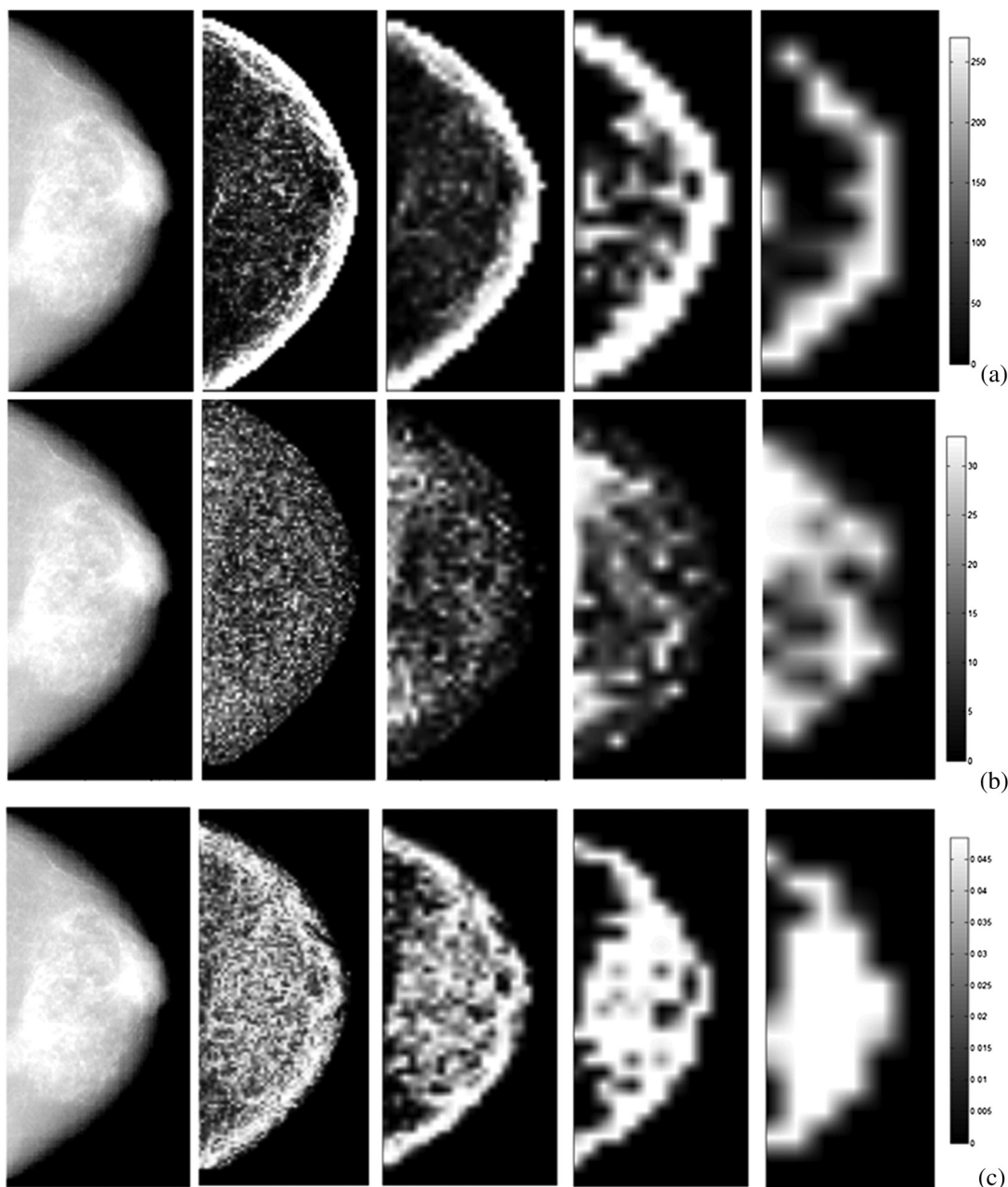


Fig. 6 Illustration of the effect of the window size (W) on the texture feature images. Representative examples shown in each row, from left to right, for GE 2000D raw images (after $-\log$ transformation) and corresponding texture feature images with increasing window sizes of 15, 31, 63, and 127 pixels² for texture features (a) STD (gray-level histogram); (b) GLN (run length); (c) EEI (structural). The texture feature images are generated from z-score normalized raw images acquired on the GE 2000D system.

feature calculation. This interpixel correlation is also likely to affect the robustness of many of the gray-level run-length features: as the correlation properties differ between images, so will the gray-level runs in an image as they are, in essence, representative of pixel-adjacency relationships. Similarly, as the lattice window size W increases, more features are labeled as robust between the three systems. This does not necessarily imply that features generated using a coarser lattice scheme (e.g., 63 pixels or larger) or larger co-occurrence offset length l will perform better in tissue characterization or breast cancer risk assessment. As the images are essentially smoothed when

analyzed with such coarser parameters, image noise may be removed, implying less confounding results due to system effects; however, the performance of these features for tissue characterization may be reduced, and this will need to be further investigated in future research studies.

One interesting finding from our study is the impact of vendor normalization on apparent gray-level values in an image. One potential explanation is that these differences are due to the fact that x-ray projection images are inherently not calibrated, as are most medical imaging modalities with a few exceptions, such as computed tomography images. In practice, this leads to

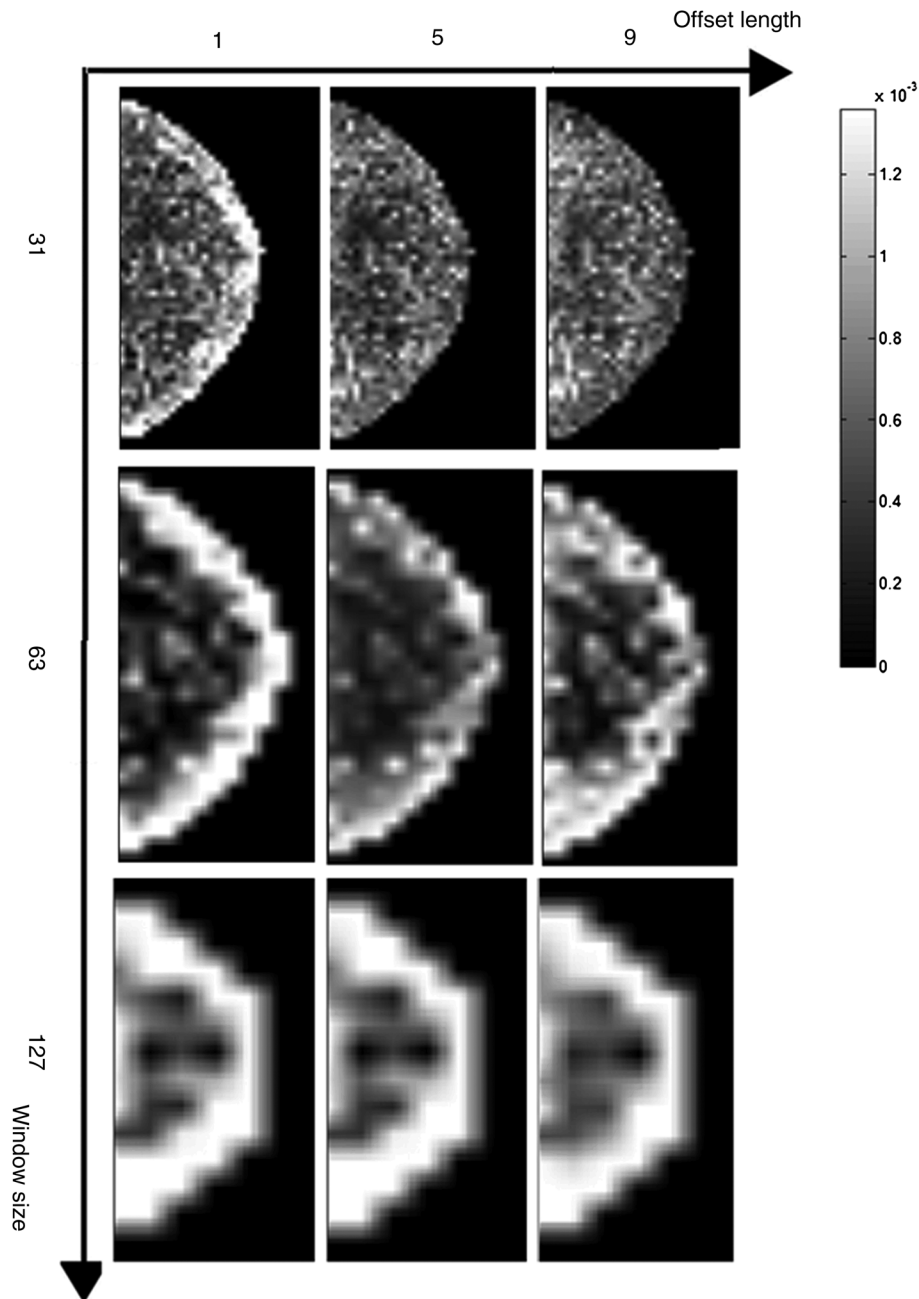


Fig. 7 Co-occurrence feature energy with varying window size and offset length. Each row indicates one window size (31, 63, 127 pixels²) and each column indicates one offset length (1, 5, 9 pixels).

vendor-specific mapping of observed x-ray attenuation values to gray-level pixel values using proprietary algorithms. This mapping can vary between the different generations of a given system, as is observed in this study, as well as between different vendors of a given modality. The existence of such differences demonstrates the necessity of running studies such as ours and the use of a consistent normalization scheme (i.e., log-transformation followed by z-score normalization) when investigating imaging features dependent on gray-level values, especially when images are acquired from a variety of data sources, in order to ensure that observed differences in an imaging feature are not affected by underlying confounding and bias.

One should note that although the phantom was not moved between the five acquisitions of a single system, it was physically moved between the three systems, and thus, the mammograms taken by the three systems were not spatially aligned in any systematic way, as, for example, by the means of image registration. In that context, our results also suggest that when a whole-breast texture analysis approach is taken, the exact alignment of the lattice regions of interest, and to a lesser extent, the dose and system resolution differences, seem to have a negligible impact on the robustness of a significant subset of features. This is likely due to the fact that any small variations in texture over the breast are averaged out when analyzing the entire breast, and supports the notion and utility of whole-breast

texture analysis. One limitation of our study is that while we do compare several detectors with different technological properties, two of the three systems evaluated, specifically the GE 2000D and DS, were developed by the same manufacturer. In addition, we only evaluate a single system with a direct detector, the Hologic Selenia Dimensions, and only utilize one physical phantom as ground truth for parenchymal texture in our analysis.

Last, one issue common to studies evaluating many metrics simultaneously is that of multiple comparisons. However, hypothesis testing for equivalence via a CI-based approach somewhat mitigates this multiplicity issue. A multiple-comparison adjustment (e.g., a Bonferroni type of approach) in this setting would imply the use of a wider CI for interscanner differences than the commonly used 95% CI. However, based on the definitions of robustness defined in our work, the use of wider CIs would have systematic effects on the results that would most likely and for the most part leave our findings unchanged. Specifically, given any arbitrarily wider CI as a multiple-comparisons correction, VR features would at most become R as the test statistic would still remain in the intrasystem CI and the wider CI would remain crossing $\Delta = 0$ even if the entire CI was no longer contained within the intrasystem CI. In contrast, NR feature CIs could at most become LR by crossing $\Delta = 0$, but a wider CI will not alter the point estimate of the test statistic nor allow for the CI to shift to be entirely contained within the intrasystem CI if it was not already. As such, even in the presence of multiple comparisons correction, the relative robustness categories of the features would likely hold, particularly in the case of NR features as they cannot, by definition, meet the criterion for R or VR features. The fact that data-driven, feature-specific margin parameters are used to define equivalence also minimizes the multiple-comparisons issue as the number of family-wise comparisons is drastically reduced relative to the case where only a single, study-wide margin is utilized.

Overall, the comparison between these imaging systems with a specific phantom serves as a proof-of-concept for the utility of our approach, and in principal, our proposed statistical framework could generalize to additional imaging systems and breast phantoms. For example, while the Rachel phantom used in this work was designed as a craniocaudal view phantom, extension of this analysis to include mediolateral-oblique view mammograms and phantoms, which may include the chest wall, would primarily depend on accurate delineation of the breast area so as to exclude the pectoralis muscle and would require no alterations to the proposed framework itself. Furthermore, while our experiments focus on detector differences when mammograms are acquired using an x-ray dose estimated using the standard automated exposure control setting, there exists the option to also perform low-dose analyses, such as with tomosynthesis projection data, or to acquire texture from other breast screening modalities altogether, such as with MRI and whole-breast ultrasound. As such, future work will, therefore, need to consider dose and intermodality differences in addition to model and detector differences within a given modality. In addition, our framework could also be used to examine a broader range of texture features and other parenchymal descriptors for different computation parameters than those considered in this work. Ultimately, optimization of these features should also consider the specific context of the clinical application at hand, as for example examining the performance

of these features in breast cancer risk assessment for different feature extraction parameters.

5 Conclusion

We proposed a statistical framework to evaluate the robustness of texture features across different FFDM imaging systems, focusing specifically on the effect of different x-ray systems on image intensity and texture. Our results suggest that proper choice of feature extraction parameters is crucial in generating robust texture features. The z-score normalization can alleviate linear differences in intensity profiles between different imaging systems. Once the images are normalized, gray-level histograms and structural features appear to be more robust than co-occurrence and run-length texture features. For all texture features, larger window sizes (i.e., >63 pixels²) tend to result in more robust texture features. Specifically for co-occurrence texture features, a larger offset length (i.e., >7 pixels) also appears to generate features more robust to inherent imaging system effects. The framework proposed in this work will be beneficial for studies utilizing texture analysis in digital mammography in clinical, multisite, multivendor data sets, including studies investigating breast cancer risk assessment, tissue classification, and computer-aided diagnosis.

Acknowledgments

This work was supported by an American Cancer Society (ACS) Research Scholar Grant (RSGHP-10-108-01-CPHPS), by research grants from the National Institutes of Health/National Cancer Institute (1U54CA163313-01, R01-CA161749-01A1), and by a Breast Cancer Alliance Young Investigator Research Grant.

References

1. R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2013," *CA Cancer J. Clin.* **63**, 11–30 (2013).
2. American Cancer Society, "Breast cancer facts and figures 2011–2012," 2011, <http://www.cancer.org/acs/groups/content/epidemiologysurveillance/documents/document/acspc-030975.pdf> (17 March 2015).
3. FDA, "MQSA National Statistics," 2014, <http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/facilityScorecard/ucm113858.htm> (21 January 2015).
4. S. Suryanarayanan et al., "Flat-panel digital mammography system: contrast-detail comparison between screen-film radiographs and hard-copy images," *Radiology* **225**(3), 801–807 (2002).
5. N. Boyd et al., "The association of breast mitogens with mammographic densities," *Br. J. Cancer* **87**(8), 876–882 (2002).
6. N. F. Boyd et al., "Mammographic densities and breast cancer risk," *Cancer Epidemiol. Biomarkers Prev.* **7**(12), 1133–1144 (1998).
7. H. Li et al., "Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms," *Acad. Radiol.* **12**(7), 863–873 (2005).
8. A. Manduca et al., "Texture features from mammographic images and risk of breast cancer," *Cancer Epidemiol. Biomarkers Prev.* **18**(3), 837–845 (2009).
9. J. N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern," *Cancer* **37**(5), 2486–2492 (1976).
10. V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," *Cancer Epidemiol. Biomarkers Prev.* **15**(6), 1159–1169 (2006).
11. J. J. Heine and P. Malhotra, "Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography. Part 1. Tissue and related risk factors," *Acad. Radiol.* **9**(3), 298–316 (2002).
12. J. J. Heine and P. Malhotra, "Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography. Part 2. Serial breast

- tissue change and related temporal influences," *Acad. Radiol.* **9**(3), 317–335 (2002).
13. C. Byrne et al., "Mammographic features and breast cancer risk: effects with time, age, and menopause status," *J. Natl. Cancer Inst.* **87**(21), 1622–1629 (1995).
 14. C. Byrne et al., "Effects of mammographic density and benign breast disease on breast cancer risk (United States)," *Cancer Causes Control* **12**(2), 103–110 (2001).
 15. C. J. Fabian and B. F. Kimler, "Mammographic density: use in risk assessment and as a biomarker in prevention trials," *J. Nutr.* **136**(10), 2705–2708 (2006).
 16. J. A. Tice et al., "Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population," *Breast Cancer Res. Treat.* **94**(2), 115–122 (2005).
 17. H. Li et al., "Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location," *Med. Phys.* **31**, 549 (2004).
 18. H. Li et al., "Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms," *Acad. Radiol.* **12**(7), 863–873 (2005).
 19. H. Li et al., "Power spectral analysis of mammographic parenchymal patterns for breast cancer risk assessment," *J. Digit. Imaging* **21**(2), 145–152 (2008).
 20. H. Li et al., "Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment," *Acad. Radiol.* **14**(5), 513–521 (2007).
 21. W. H. Kim et al., "Variability of breast density assessment in short-term reimaging with digital mammography," *Eur. J. Radiol.* **82**(10), 1724–1730 (2013).
 22. C. B. Caldwell and M. J. Yaffe, "Development of an anthropomorphic breast phantom," *Med. Phys.* **17**, 273 (1990).
 23. M. L. Marx and R. J. Larsen, *Introduction to Mathematical Statistics and Its Applications*, Pearson/Prentice Hall, Upper Saddle River, New Jersey (2006).
 24. B. M. Keller et al., "Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation," *Med. Phys.* **39**(8), 4903–4917 (2012).
 25. Y. Zheng et al., "A fully-automated software pipeline for integrating breast density and parenchymal texture analysis for digital mammograms: parameter optimization in a case-control breast cancer risk assessment study," *Proc. SPIE* **8670**, 86701B (2013).
 26. L. Ibanez et al., "The ITK software guide," 2003, <http://www.itk.org/ItkSoftwareGuide.pdf> (9 August 2014).
 27. U. Meier, "Nonparametric equivalence testing with respect to the median difference," *Pharm. Stat.* **9**(2), 142–150 (2010).
 28. M. Amadasun and R. King, "Textural features corresponding to textural properties," *Trans. Syst. Man Cybern.* **19**, 1264–1274 (1989).
 29. R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **3**(6), 610–621 (1973).
 30. A. M. Khuzi et al., "Identification of masses in digital mammogram using gray level co-occurrence matrices," *Biomed. Imaging Interv. J.* **5**(3), e17 (2009).
 31. M. M. Galloway, "Texture analysis using gray level run lengths," *Comput. Graph. Image Process.* **4**(2), 172–179 (1975).
 32. A. Chu, C. Sehgal, and J. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recognit. Lett.* **11**(6), 415–419 (1990).
 33. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002).
 34. C. B. Caldwell et al., "Characterisation of mammographic parenchymal pattern by fractal dimension," *Phys. Med. Biol.* **35**(2), 235 (1990).
 35. G. Karemore et al., "Anisotropic diffusion tensor applied to temporal mammograms: an application to breast cancer risk assessment," in *Annual Int. Conf. of the Engineering in Medicine and Biology Society*, pp. 3178–3181 (2010).
 36. Z. Huo et al., "Computerized analysis of digitized mammograms of BRCA1 and BRCA2 gene mutation carriers," *Radiology* **225**(2), 519–526 (2002).
 37. J. Weickert, "Coherence-enhancing diffusion filtering," *Int. J. Comput. Vis.* **31**(2), 111–127 (1999).
 38. T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006).
 39. Q. A. Holmes, D. R. Nuesch, and R. A. Shuchman, "Textural analysis and real-time classification of sea-ice types using digital SAR data," *IEEE Trans. Geosci. Remote Sens.* **GE-22**(2), 113–120 (1984).
 40. L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sens.* **37**(2), 780–795 (1999).
 41. A. K. Mohanty, S. Beberta, and S. K. Lenka, "Classifying benign and malignant mass using GLCM and GLRLM based texture features from mammogram," *Int. J. Eng. Res. Appl.* **1**(3), 687–693 (2011).
 42. R. van den Boomgaard, "Algorithms for non-linear diffusion," Technical Report no.1, Intelligent Sensory Information Systems, University of Amsterdam, The Netherlands.
 43. J. Hodges and E. L. Lehmann, "Hodges-Lehmann estimators," in *Encyclopedia of Statistical Sciences*, S. Kotz, N. L. Johnson, and C. B. Read, Eds., Vol. 3, pp. 463–465, John Wiley & Sons, New York (1983).
 44. M. J. Yaffe, "Detectors for digital mammography," in *Digital Mammography*, U. Bick and F. Diekmann, Eds., pp. 13–31, Springer, Berlin, Heidelberg (2010).

Brad M. Keller received his PhD from Cornell University in 2011. Currently, he is a research associate in the Computational Breast Imaging Group of the Department of Radiology at the University of Pennsylvania. His current research interest focuses on quantitative medical image analysis, applied epidemiological and statistical analysis of quantitative analytical software, and development of image-derivable biomarkers of the pathophysiology underlying breast density and its association to breast cancer risk.

Andrew Oustimov received his MPH (biostatistics) from the University of South Florida in 2014, after which he joined the Computational Breast Imaging Group in the Department of Radiology at the University of Pennsylvania. With expertise in pattern recognition, predictive modeling, analysis of survival, categorical, and longitudinal/multilevel data, he is currently interested in causal inference, biomedical image analysis, cancer genomics, and predictive modeling with respect to personalized screening, diagnosis, prognosis, and treatment of breast cancer.

Yan Wang received her PhD from the University of Pennsylvania in 2013. Her research primarily focused on the study of image-based local scale structure using nonlinear diffusion in various medical imaging modalities. She is currently a health economist with Grey Healthcare Group.

Jinbo Chen is an associate professor of biostatistics at the University of Pennsylvania. Her primary areas of research include the development of statistical methods for designing and analyzing two-phase epidemiologic studies, methodological and collaborative work in genetic association studies of complex diseases, and the development and evaluation of risk prediction models.

Raymond J. Acciavatti received his PhD from the University of Pennsylvania in 2013. He is currently a postdoctoral researcher in the Department of Radiology Physics Section of the University of Pennsylvania. His research focuses on image acquisition physics and the development of new techniques to improve image resolution.

Yuanjie Zheng received his PhD from Shanghai Jiaotong University in 2006. He is currently a senior research investigator in the Penn Image Computing & Science Lab at the University of Pennsylvania. His research focuses on the enhancement of patient care via the creation of algorithms for automatically quantifying and generalizing the information latent in various medical images for tasks such as disease analysis and surgical planning.

Shonket Ray received his PhD from the University of California, Davis, in 2012. He is currently a postdoctoral researcher in the Computational Breast Imaging Group at the University of Pennsylvania School of Medicine. His research interests focus on the development of novel techniques using advanced biomedical image analysis and

processing, such as image segmentation, classification, and feature extraction using various breast imaging modalities.

James C. Gee is an associate professor of radiologic science and computer and information science, and director of the Penn Image Computing and Science Laboratory at the University of Pennsylvania. His research interests are broadly in bio/medical image analysis and computing, with active work in all of the quantitative methods represented, including segmentation, registration, morphometry, and shape statistics, as applied to a variety of organ/model systems and the major modalities in biological and medical imaging.

Andrew D. A. Maidment is an associate professor of radiology at the Hospital of the University of Pennsylvania and chief of the Physics Section of the Department of Radiology. His research currently focuses on the development of advanced imaging modalities to improve breast cancer detection and clinical care.

Despina Kontos is an assistant professor of radiology and head of the Computational Breast Imaging Group at the University of Pennsylvania. Her current research focuses on assessing the role of imaging as a quantitative biomarker for improving personalized clinical decision making for breast cancer screening, prognosis, and treatment.