FFDM image quality assessment using computerized image texture analysis

Rachelle Berger, Ann-Katherine Carton, Andrew D. A. Maidment, and Despina Kontos

University of Pennsylvania, Department of Radiology, 3400 Spruce St., Philadelphia PA 19104
bergerra@mail.med.upenn.edu
{Ann-Katherine.Carton | Andrew.Maidment | Despina.Kontos} @uphs.upenn.edu

ABSTRACT

Quantitative measures of image quality (IQ) are routinely obtained during the evaluation of imaging systems. These measures, however, do not necessarily correlate with the IQ of the actual clinical images, which can also be affected by factors such as patient positioning. No quantitative method currently exists to evaluate clinical IQ. Therefore, we investigated the potential of using computerized image texture analysis to quantitatively assess IQ. Our hypothesis is that image texture features can be used to assess IQ as a measure of the image signal-to-noise ratio (SNR). To test feasibility, the "Rachel" anthropomorphic breast phantom (Model 169, Gammex RMI) was imaged with a Senographe 2000D FFDM system (GE Healthcare) using 220 unique exposure settings (target/filter, kVs, and mAs combinations). The mAs were varied from 10%-300% of that required for an average glandular dose (AGD) of 1.8 mGy. A 2.5cm^2 retroareolar region of interest (ROI) was segmented from each image. The SNR was computed from the ROIs segmented from images linear with dose (*i.e.,* raw images) after flat-field and off-set correction. Image texture features of skewness, coarseness, contrast, energy, homogeneity, and fractal dimension were computed from the *Premium View*^{TM} post-processed image ROIs. Multiple linear regression demonstrated a strong association between the computed image texture features and SNR (R^2=0.92, $p \leq 0.001$). When including kV, target and filter as additional predictor variables, a stronger association with SNR was observed (R^2=0.95, $p \leq 0.001$). The strong associations indicate that computerized image texture analysis can be used to measure image SNR and potentially aid in automating IQ assessment as a component of the clinical workflow. Further work is underway to validate our findings in larger clinical datasets.

Keywords: Digital mammography, image quality, signal-to-noise ratio (SNR), image texture analysis.

1. INTRODUCTION

Mammography poses a demanding task for the radiologist, as low-contrast masses and small micro-calcifications can be obscured in the superposition of overlapping breast tissues. Good image quality (IQ) is therefore critical for lesion detection and characterization[1,2]. Currently no method exists to quantitatively assess clinical IQ in support of the diagnostic interpretation. Quantitative measures of IQ are mainly obtained during routine system evaluations for characterizing and monitoring the performance of specific imaging systems. Common quantitative measures include signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), and the noise equivalent quanta (NEQ)[3,4]. Although useful for assessing the performance of specific imaging systems, these measures do not necessarily correlate with the IQ of actual clinical images[5-7]. Patient-specific image quality can also be affected by factors such as breast positioning, compression, and potential image artifacts, which are accounted for by radiologists in a qualitative assessment[5,6].

We investigated the potential use of computerized breast image texture analysis to quantitatively assess IQ. Image texture features, such as skewness, coarseness, contrast, energy, homogeneity, and fractal dimension, have been used extensively in mammographic image analysis, to characterize parenchymal patterns[8-12] and for estimating breast density[13,14]. Our hypothesis is that image texture features can also be used to assess IQ, as measured by the image SNR. To test feasibility, we determined the relationship between computer-extracted image texture features and SNR as a function of image acquisition parameters for a full-field digital mammography (FFDM) system. Our long-term goal is to develop a Computer-Aided Diagnostic Image Quality (CAD*iq*) tool that can be integrated into the clinical workflow to provide quantitative IQ measures for clinical images in support of the diagnostic interpretation.

Medical Imaging 2010: Physics of Medical Imaging, edited by Ehsan Samei, Norbert J. Pelc,
Proc. of SPIE Vol. 7622, 762213 · © 2010 SPIE · CCC code: 1605-7422/10/$18 · doi: 10.1117/12.845297

2. METHODS

2.1 Image acquisition

The "Rachel" anthropomorphic breast phantom (Model 169, Gammex RMI, Madison, WI)[15,16] was imaged with a Senographe 2000D FFDM system (GE Healthcare, Chalfont St. Giles, UK). Images were acquired with various combinations of target/filter (Mo/Mo, Rh/Rh, Mo/Rh), kV (25 – 34 kV), and mAs, resulting in 220 unique exposure settings. For each combination of acquisition settings two phantom images were acquired in order to compute a difference image for estimating noise (Eq. 1). The mAs was varied from 10%-300% of that required for a standard phototimed exposure yielding an average glandular dose (AGD) of 1.8 mGy. Thus, for each target/filter and kV combination, eight mAs settings were acquired relative to the reference mAs (10%, 20%, 40%, 70%, 100%, 150%, 200%, and 300%).

2.2 SNR calculation

To calculate SNR, a 2.5 cm^2 (256×256 pixels at 0.1 mm/pixel resolution) retroareolar region of interest (ROI) was segmented from each image (Fig. 1). The retroareolar region was selected for its texture richness, due to the underlying prominence of the ductal network[17]. The ROI was placed in the exact same position in the retroareolar region in each acquired image using automated software. SNR was computed from the ROIs of the images linear with dose (*i.e.,* raw images) after flat-field and off-set correction. Signal intensity was computed as the average pixel value. Noise (σ) was computed as the root mean square (RMS) difference of each pair of images acquired with the same exposure settings as,

$$\sigma = \frac{\sigma_\Delta}{\sqrt{2}} \qquad \textbf{(Eq. 1)}$$

where σ_Δ is the per-pixel standard deviation of the difference ROI.

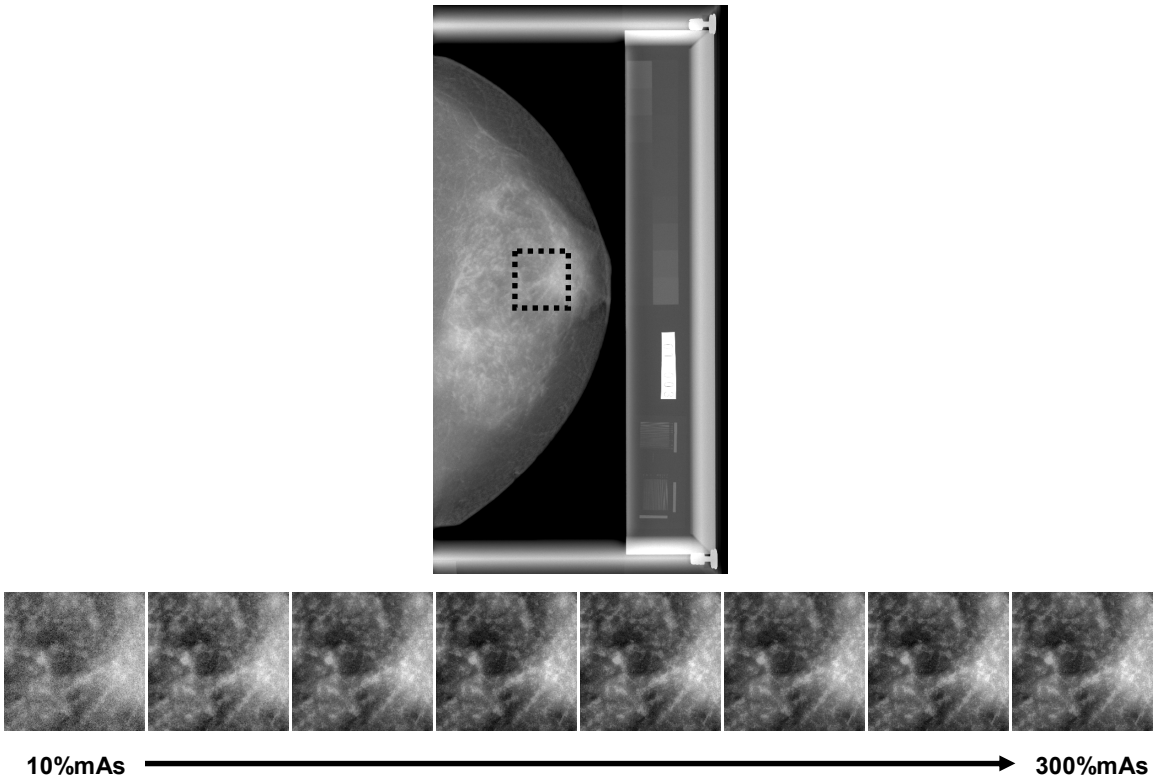

Figure 1. FFDM image of the "Rachel" anthropomorphic breast phantom acquired with Mo/Mo, 28kV, at 100% of the AGD. The segmented ROI is outlined, and displayed below over the entire range of mAs exposure settings.

2.3 Texture analysis

Image texture features of skewness, coarseness, contrast, energy, homogeneity, and fractal dimension (FD) were computed from ROIs of the *Premium View*^{TM} (GE Healthcare, Chalfont St. Giles, UK, v. ADS_43.10.1) post-processed images[10,12,18-21].

Skewness was computed as the third statistical moment of the gray-level histogram:

$$skewness = \frac{w_3}{w_2^{3/2}}, \quad w_k = \sum_{i=0}^{g_{max}} n_i \left(i - \bar{i}\right)^k \Big/ N, \quad N = \sum_{i=0}^{g_{max}} n_i, \quad \bar{i} = \sum_{i=0}^{g_{max}} \left(i n_i / N\right), \qquad \textbf{(Eq. 2)}$$

where n_i represents the number of pixels in the ROI with gray-level value i, g_{max} is the maximum gray-level value, and N is the total number of pixels. Skewness measures the asymmetry of the gray-level histogram around the mean. For radiographic breast images, largely bright or dense regions tend to have a negative measure of skewness, whereas dark or fatty regions tend to have higher, positive values of skewness[9,21].

Coarseness is based on the neighborhood gray tone difference matrix (NGTDM), *v(i),* computed as[9,17,21]:

$$v(i) = \left\{ \begin{array}{c} \sum \left| i - \overline{L}_i \right| \; for \; i \in \{n_i\} \; if \; n_i \neq 0 \\ 0 \quad otherwise \end{array} \right\} \qquad \textbf{(Eq. 3)}$$

Here, *v(i)* is derived from the difference between each pixel's gray-level value i and the average gray-level value ($\overline{L_i}$) in the neighborhood window around the pixel (Eq. 4). The set $\{n_i\}$ contains all pixels with gray-levels equal to i, so that the NGTDM = 0 if there are no pixels with gray-level value i.

$$\overline{L}_i = \frac{1}{S-1} \sum_{k=-t}^{t} \sum_{l=-t}^{t} j(x+k, y+l) \qquad \textbf{(Eq. 4)}$$

In equation (4), *j(x,y)* is the pixel located at *(x,y)* with gray-level value i, *(k,l)* $\neq$ *(0,0)*, and $S = (2t+1)^2$, with $t = 1$ specifying the neighborhood size around pixel j. Coarseness is calculated based on the NGTDM as:

$$coarseness = \left(\sum_{i=0}^{g_{max}} p_i v(i) \right)^{-1} \qquad \textbf{(Eq. 5)}$$

where g_{max} is the maximum gray-level value in the ROI, and p_i is the probability that gray-level i occurs. Coarseness is inversely proportional to the amount of local variation in gray-level pixel values within the neighborhood. Hence, a high value of coarseness corresponds with little variation in local gray-levels, describing quantitatively the texture pattern humans perceive as coarse.

Contrast, energy, and homogeneity, proposed originally by Haralick *et. al.*[19], require the computation of second-order statistics from the gray-level co-occurrence matrix, which measures the frequency at which two given gray levels occur with a certain spatial separation in a specified direction. Contrast, energy, and homogeneity were computed as:

$$contrast = \sum_{i=0}^{g_{max}} \sum_{j=0}^{g_{max}} \left| i - j \right|^2 C(i, j) \qquad \textbf{(Eq. 6)}$$

$$energy = \sum_{i=0}^{g_{max}} \sum_{j=0}^{g_{max}} C(i, j)^2 \qquad \textbf{(Eq. 7)}$$

$$homogeneity = \sum_{i=0}^{g_{max}} \sum_{j=0}^{g_{max}} \frac{C(i, j)}{1 + \left| i - j \right|} \qquad \textbf{(Eq. 8)}$$

Equations (6-8) are summed over every pair of gray-level values *(i,j)*, with g_{max} being the maximum gray-level value in the ROI, and *C(i,j)* being the normalized co-occurrence matrix. To optimize the computation of the gray-level co-occurrence statistics, gray-level quantization was implemented[10,12]. The co-occurrence frequencies were calculated symmetrically in the four directions around each pixel using a displacement vector, $d=(dx,dy)$, along x and y dimensions, where $dx=dy=1$ pixel offset. The texture features calculated in each of these four directions were averaged to create a single measure that was used in our experiments[12].

Fractal dimension (FD) was calculated based on the power spectrum of the Fourier transform of the image[18,20]. The 2D discrete Fourier transform was performed using the fast-Fourier transform (FFT) algorithm as:

$$F(u,v)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} I(m,n)e^{-j(2\pi/\mathrm{M})um}e^{-j(2\pi/N)vn}, \quad u=0,1,...,M-1 \quad v=0,1,...N-1 \qquad \textbf{(Eq. 9)}$$

where *I* is the image ROI of size *(M,N),* and *u* and *v* are the spatial frequencies in the *x* and *y* directions. The power spectral density, *P*, was calculated from *F(u,v)* as:

$$P(u,v)=\left|F(u,v)\right|^2 \qquad \textbf{(Eq. 10)}$$

To compute the FD, *P* was averaged over radial slices spanning the FFT frequency domain. The frequency space was uniformly divided in 24 directions, with each direction uniformly sampled at 30 points along the radial component. To calculate the FD, the least-squares fit of log(P_f) versus log(*f*) was estimated, where $f=\sqrt{u^2+v^2}$ denotes the radial frequency. The FD is related to the slope, β, of this log-log plot by:

$$FD=\frac{3D_T+2-\beta}{2}=\frac{8-\beta}{2} \qquad \textbf{(Eq. 11)}$$

where D_T is the topologic dimension, equal to 2 for a 2D image. FD indicates the measure of self-similarity in the texture pattern, and the overall texture roughness at different scales.

2.4 Statistical analysis

The Pearson's correlation coefficient (*r*) was computed to evaluate the relationship between individual texture features and SNR for each target/filter combination, over the full range of kV and mAs settings. Stepwise multiple linear regression was applied to model the association between i) combinations of the texture features versus SNR and ii) combinations of the texture features in addition to kV, target and filter versus SNR. Variable normalization using the z-score was performed prior to inclusion in the model. A threshold of p-value ≤ 0.01 was set during stepwise feature selection. Both correlation and regression analyses were performed after logarithmic transformation of the variables.

3. RESULTS

Individual texture features correlate strongly with SNR. Table 1 summarizes the Pearson's correlation coefficients between the individual texture features and SNR, over the entire range of kV and mAs exposure settings. The strongest association was observed between energy and SNR ($|r|$=0.91-0.96, $p\leq0.001$), and the weakest association was observed between contrast and SNR ($|r|$=0.64-0.83, $p\leq0.001$). Figure 2 illustrates these trends for the specific texture features for the different target/filter combinations at 28 kV. With the exception of homogeneity, all texture features have an approximately linear relationship with SNR at low-dose settings until reaching a plateau at approximately 70% percent of the reference mAs. Homogeneity decreases continuously as a function of SNR.

Table 1. Pearson's correlation coefficients (*r*) for each texture feature and SNR, over the entire range of kV and mAs, for each target and filter combination, after logarithmic transformation of variables. All reported coefficients are statistically significant ($p \leq 0.001$).

		Pearson correlation coefficient (*r*)					
		Skewness	**Coarseness**	**Contrast**	**Energy**	**Homogeneity**	**FD**
SNR	Mo/Mo	0.94	0.96	-0.83	-0.96	-0.87	0.87
	Rh/Rh	0.83	0.89	-0.64	-0.91	-0.86	0.77
	Mo/Rh	0.90	0.94	-0.74	-0.96	-0.85	0.85

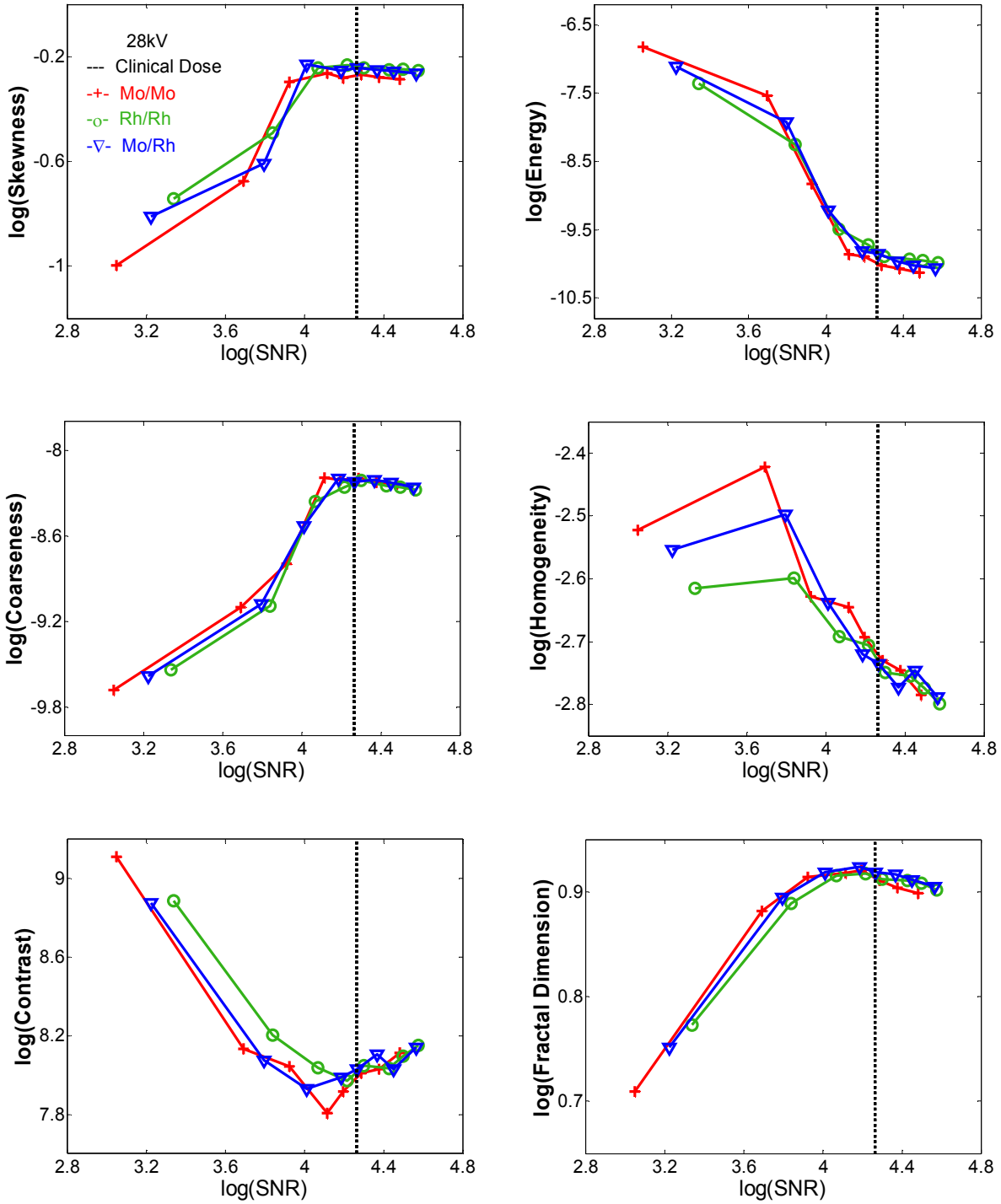

Figure 2. Relationship between individual texture features and SNR for the different target/filter combinations at 28 kV. Data points shown correspond to the eight different mAs settings investigated. The dotted line marks the reference AGD at 1.8 mGy.

Stepwise multiple linear regression models demonstrated a strong, significant association between combinations of the computed image texture features and SNR. Model (I) considers only combinations of texture features, while Model (II) considers combinations of texture features and image acquisition settings of kV, target and filter. Model (I) demonstrated strong predictive association between image texture features and SNR (R^2=0.92, p<0.001), with contrast, energy and FD selected as the most significant input variables predictive of SNR ($p \leq 0.001$). In Model (II), the observed association was slightly stronger when kV, target and filter were also considered as predictor variables (R^2=0.95, p<0.001), where all texture features were selected as significant predictors after the addition of kV, target and filter. Full model statistics are included in Tables 2 and 3. Figure 3 shows the values of the SNR predicted by the multiple linear regression Model (II) versus the actual SNR, over the entire range of administered dose, illustrating a highly accurate predictive association.

Table 2. Model (I): Stepwise multiple linear regression of combined texture features vs. SNR.

Model Summary				
R	R^2	Adjusted R^2	Std. Error	
0.96	0.92	0.92	0.013	
Coefficients				
	B	Std. Error	t	Sig.
(intercept)	4.101	0.008	523.623	0.000
Contrast	0.291	0.033	8.742	0.000
Energy	-0.302	0.014	-21.217	0.000
FD	0.374	0.038	9.953	0.000

Table 3. Model (II): Stepwise multiple linear regression of combined texture features, kV, target and filter vs. SNR.

Model Summary				
R	R^2	Adjusted R^2	Std. Error	
0.974	0.949	0.95	0.009	
Coefficients				
	B	Std. Error	t	Sig.
(intercept)	4.101	0.006	639.891	0.000
Skewness	-0.185	0.034	-5.442	0.000
Coarseness	-0.483	0.082	-5.871	0.000
Contrast	0.389	0.059	6.597	0.000
Energy	-1.250	0.150	-8.310	0.000
Homogeneity	0.319	0.067	4.747	0.000
FD	0.443	0.039	11.317	0.000
kV	0.098	0.010	9.474	0.000
Target	0.039	0.008	4.949	0.000
Filter	0.029	0.009	3.294	0.001

4. DISCUSSION

Clinical imaging procedures operate under the assumption that a well calibrated and well characterized imaging system will also produce images of good diagnostic quality, while no method exists to validate the actual IQ for the obtained clinical images. In order to assess the feasibility of a fully automated, quantitative method for assessing IQ of digital mammography images, we investigated the potential of using computerized image texture analysis to quantitatively assess IQ. Our results show that mammographic image texture is indicative of the corresponding IQ, as measured by SNR. Strong associations are observed between the texture features and image SNR, particularly when accounting for kV, target and filter acquisition settings (Fig 3). These observations could have significant implications in advancing technological developments for IQ assessment in clinical practice. If the observed association also holds true in clinical images, as compared to the phantom images analyzed in this proof-of-concept study, then computerized image texture analysis could be used to provide fully-automated, quantitative measures of clinical IQ on a per-patient basis at the time of imaging. Our ultimate goal is to develop a Computer-Aided Diagnostic Image Quality (CAD*iq*) tool, based on computerized image texture analysis that can be integrated as a component of the clinical workflow.

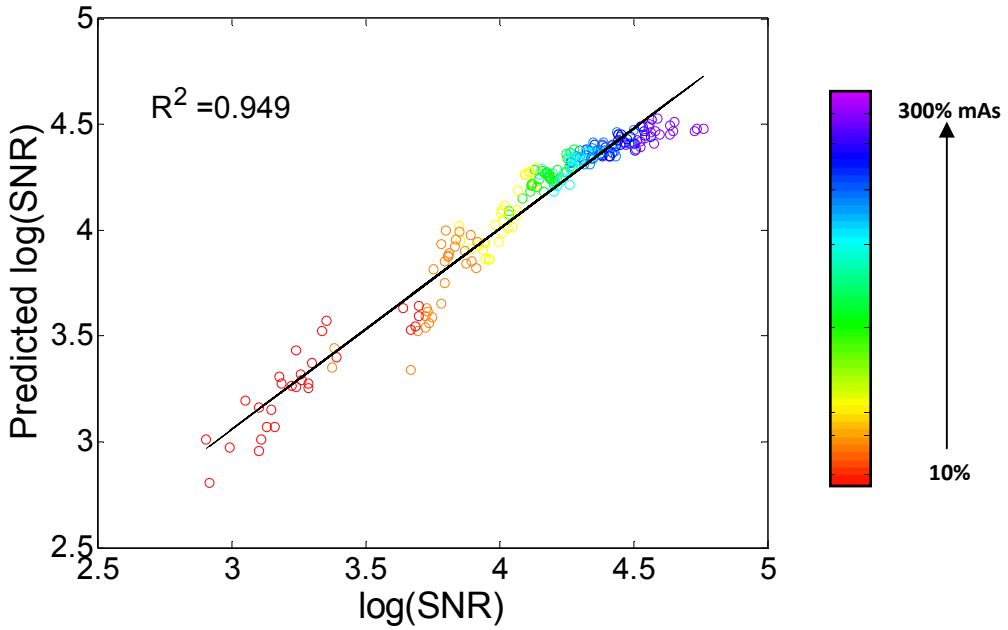

Figure 3. Multiple linear regression, showing strong association between combination of texture features, kV, target/filter vs. SNR.

The primary advantage of the envisioned CAD*iq* tool is that it shifts the current practice towards personalized, patient-specific IQ control. A recent paper published by Van Ongeval *et al.*[22], highlighting the need to use practical criteria to assess clinical IQ and verify dose settings for digital mammography, proposed a set of parameters to assess IQ including factors such as contrast, sharpness and saturation of image regions; nevertheless the assessment was still based on the qualitative assessment of the radiologist and the assignment of rating scores on a small quantitative scale. With a CAD*iq* tool, the SNR determined from texture analysis could be compared with the SNR expected from the amount of radiation used. In this way, texture analysis could be used to compare the achieved IQ with that which was desired. Such a dose-efficiency measure could be used to establish objective guidelines for IQ control during clinical imaging procedures.

In our experiments, relatively little change in texture was observed beyond 70% of the reference AGD. This may indicate that no additional anatomical texture information can be extracted from the image beyond these dose levels. Although this observation should be taken with caution, given that our results are restricted to the use of the specific breast phantom, a single x-ray system, and a particular image post-processing algorithm used in our experiments, the point at which texture features reach this threshold may indicate that anatomy dominates over quantum noise. This would indicate that sufficient dose has been used to extract quantitative information with respect to the particular texture feature. This could result in additional implications. For example, growing evidence supports the use of mammographic texture as a biomarker for breast cancer risk assessment[10,12,20]. If a dose of 40%-70% of the AGD is sufficient to quantify specific such textural characteristics from clinical images, then texture-based risk-related information could be measured at a lower radiation dose. However, it is uncertain whether the same plateau behavior of the texture features would be preserved in clinical images. The resulting plateau behavior of certain texture features versus SNR, for example, could be attributed to the limited spatial resolution of the phantom. It is also unclear if the natural texture variation of the breast tissue between women would dominate the observed variation due to the imaging physics that could potentially be observed on an individual basis. Further work is warranted to validate these findings in larger clinical datasets over a range of different breast parenchymal textures, imaging systems, and image-processing methods.

5. CONCLUSION

To the best of our knowledge, our study is the first to investigate the potential of using computerized image texture analysis to quantitatively assess digital mammography IQ. Our results show that computer-extracted image texture features are indicative of IQ, as measured by SNR. In particular, multiple linear regression of image texture features can predict SNR measurements with high accuracy (R^2>0.92, p≤0.001). When including kV, target and filter as additional predictor variables, a stronger association with SNR is obtained (R^2=0.95, p≤0.001). To justify general applicability, larger clinical studies are deemed necessary that would investigate the association between clinical digital

mammographic image texture and IQ, and study the variation of texture features over a range of different breast parenchymal texture patterns, imaging systems, and image acquisition physics. Our results serve as a feasibility study for ultimately developing a CAD*iq* tool based on computerized image texture analysis. Such a quantitative IQ assessment tool could be integrated into the clinical workflow to provide quantitative measures of clinical IQ on a per-patient basis, at the time of imaging, and could be used to establish objective clinical IQ guidelines for supporting diagnostic interpretation.

ACKNOWLEDGEMENTS

This work was partially supported by Susan G. Komen for the Cure Foundation Postdoctoral Fellowship (KG080756), and by the DOD Concept Award (BC086591), and the American Association of Physicists in Medicine (AAPM) Summer Undergraduate Fellowship Program of 2008. We would also like to thank Dr. Johnny Kuo for developing and maintaining the RSNA Medical Imaging Resource Center (MIRC) image archive.

REFERENCES

[1] Williams MB, Yaffe MJ, Maidment ADA, Martin MC, Seibert A, Pisano ED. Image Quality in Digital Mammography: Image Acquisition. Journal of the American College of Radiology. 2006; 3(8):20.
[2] Haus AG, Yaffe MJ. Screen-film and digital mammography: Image Quality and Radiation Dose Considerations. Radiologic Clinics of North America. 2000; 38(4):871-98.
[3] European Guidelines for Quality Assurance in Mammography Screening. 3 Ed: European Commission, 2001.
[4] Senographe 2000D QAP: Quality Control Tests For MQSA Facilities. In: Co. GH, ed. 3, 2001.
[5] Bloomquist AK, Yaffe MJ, Pisano ED, et al. Quality control for digital mammography in the ACRIN DMIST trial: Part 1. Medical Physics. 2006; 33(3):719-36.
[6] Yaffe MJ, Bloomquist AK, Mawdsley GE, et al. Quality control for digital mammography: Part II recommendations from the ACRIN DMIST trial. Medical Physics. 2006; 33(3):737-52.
[7] Loo LND, Doi K, Metz CE. A Comparison of physical image quality indexes and observer performance in the radiographic detection of nylon beads. Physics in Medicine and Biology. 1984; 29(7):837-56.
[8] Karssemeijer N. Automated classification of parenchymal patterns in mammograms. Physics in Medicine and Biology. 1998; 43(2):365-78.
[9] Huo ZM, Giger ML, Wolverton DE, Zhong WM, Cumming S, Olopade OI. Computerized analysis of mammographic parenchymal patterns for breast cancer risk assessment: Feature selection. Medical Physics. 2000; 27(1):4-12.
[10] Kontos D, Bakic PR, Maidment ADA. Texture in digital breast tomosynthesis: A comparison between mammographic and tomographic characterization of parenchymal properties. In: M.L. Giger, N. Karssemeijer, eds. SPIE Medical Imaging: Computer Aided Diagnosis. San Diego, CA 2008.
[11] Huo ZM, Giger ML, Olopade OI, et al. Computerized analysis of digitized mammograms of BRCA1 and BRCA2 gene mutation carriers. Radiology. 2002; 225(2):519-26.
[12] Kontos D, Bakic PR, Carton AK, Troxel AB, Conant EF, Maidment ADA. Parenchymal Texture Analysis in Digital Breast Tomosynthesis for Breast Cancer Risk Estimation: A Preliminary Study. Academic Radiology. 2009; 16(3):283-98.
[13] Oliver A, Freixenet J, Marti R, et al. A novel breast tissue density classification methodology. Ieee Transactions on Information Technology in Biomedicine. 2008; 12(1):55-65.
[14] Zhou C, Chan HP, Petrick N, et al. Computerized image analysis: Estimation of breast density on mammograms. Medical Physics. 2001; 28(6):1056-69.
[15] Caldwell CB, Yaffe MJ. Development of an anthropomorphic breast phantom. Med.Physics. 1990; 17(2):273-80.
[16] GAMMEX 169 "Rachel" Anthropomorphic Breast Phantom. (http://www.gammex.com/n-portfolio/ detail. asp? Id =240&category=Mammography&name=Gammex+169+%22Rachel%22+Antrhopomorphic+Breast+Phantom); Accessed January 25, 2010.
[17] Li H, Giger ML, Huo ZM, et al. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: Effect of ROI size and location. Medical Physics. 2004; 31(3):549-55.
[18] Li H, Giger ML, Olopade OI, Chinander MR. Power spectral analysis of mammographic parenchymal patterns for breast cancer risk assessment. Journal of Digital Imaging. 2008; 21(2):145-52.
[19] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics. 1973; 3:12.
[20] Li H, Giger ML, Olopacle OI, Lan L. Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment. Academic Radiology. 2007; 14(5):513-21.
[21] Li H, Giger ML, Olopade OI, Margolis A, Chinander MR. Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms. Academic Radiology. 2005; 12(7):863-73.
[22] Van Ongeval C, Van Steen A, Geniets C, Dekeyzer F, Bosmans H, Marchal G. Clinical image quality criteria for full field digital mammography: A first practical application. Radiation Protection Dosimetry. 2008; 129(1-3):265-70.