University of Dundee

Development of an automated detection algorithm for patient motion blur in digital mammograms

Hill, Melissa L.; Whelehan, Patsy; Vinnicombe, Sarah J.; Tromans, Christopher E.; Evans, Andrew; Warwick, Violet R.

Development of an automated detection algorithm for patient motion blur in digital mammograms

Melissa L. Hill*[a], Patsy Whelehan[b,c], Sarah J. Vinnicombe[b,c], Christopher E. Tromans[a], Andrew Evans[b,c], Violet R. Warwick[b], J. Michael Brady[a], and Ralph P. Highnam[a]

[a]Volpara Solutions Ltd., Level 7, 44 Victoria Street, Wellington, 6011, NZ; [b]School of Medicine, University of Dundee, Dundee, DD1 9SY, UK; [c]NHS Tayside, Dundee, DD1 9SY, UK
*melissa.hill@volparasolutions.com

ABSTRACT

The purpose is to develop and validate an automated method for detecting image unsharpness caused by patient motion blur in digital mammograms. The goal is that such a tool would facilitate immediate re-taking of blurred images, which has the potential to reduce the number of recalled examinations, and to ensure that sharp, high-quality mammograms are presented for reading. To meet this goal, an automated method was developed based on interpretation of the normalized image Wiener Spectrum. A preliminary algorithm was developed using 25 cases acquired using a single vendor system, read by two expert readers identifying the presence of blur, location, and severity. A predictive blur severity score was established using multivariate modeling, which had an adjusted coefficient of determination, $R^2=0.63\pm0.02$, for linear regression against the average reader-scored blur severity. A heatmap of the relative blur magnitude showed good correspondence with reader sketches of blur location, with a Spearman rank correlation of 0.70 between the algorithm-estimated area fraction with blur and the maximum of the blur area fraction categories of the two readers. Given these promising results, the algorithm-estimated blur severity score and heatmap are proposed to be used to aid observer interpretation. The use of this automated blur analysis approach, ideally with feedback during an exam, could lead to a reduction in repeat appointments for technical reasons, saving time, cost, potential anxiety, and improving image quality for accurate diagnosis.

Keywords: Patient motion, unsharpness, blur, mammography, digital radiography, automated detection

1. INTRODUCTION

Achieving high mammographic imaging quality, both in terms of the physical image properties and patient positioning, is crucial towards maximizing diagnostic performance, while minimizing radiation dose and repeat examinations[1]. The key physical properties of image quality are the image contrast, noise, and sharpness. The apparent spatial resolution of a mammographic image determines its sharpness, and refers to the ability to visualize adjacent structures as being distinct. The apparent spatial resolution can be reduced from an ideal maximum due to several factors, including geometric distortion from the finite focal spot size, detector crystallization, light spread in the image receptor, and relative motion of one or more of the patient, the X-ray source, or the image receptor during image acquisition.

Most of the system-related sources of unsharpness can be identified through routine physics testing, and changes in apparent spatial resolution due to these factors are relatively infrequent[2]. In comparison, image unsharpness due to patient motion, or blur, occurs much more frequently[2]. It has been estimated that up to 20% of screening mammograms exhibit some elements of blur[3]. However, only a fraction of these are classified as clinically inadequate.

In the UK, the minimum standard is for less than 3% of exams be repeated for reasons of technical deficiency (technical repeats at the time of the examination and technical recalls by the readers), with a target of less than 2%[4]. Over a year in one UK screening service (40,954 clients), 0.88% of mammographic examinations were recalled by the reader and 1.16% were repeated by the mammographer at acquisition time (2.04%)[5]. More than half of the recalls were due to blur, but only 5% of repeats[5]. This indicates that motion blur is often not being identified at the time of the examination. One potential reason for fewer repeats due to blur at the time of the exam is a lower acquisition monitor resolution than

review monitor resolution (1 MP vs 5 MP)[5,6], and potentially poor viewing conditions. Anecdotal evidence suggests that mammographers also have a lack of time and confidence to identify blurred cases at the time of acquisition[5].

The purpose of this work is to develop and validate an algorithm to automate the detection of patient motion-related blur. The application of such an algorithm could support clinical decision-making related to motion blur; ideally at the time of the exam to allow for the potential of a retake, thereby avoiding the time, expense, and anxiety related to recalls. Furthermore, reduced motion blur should help to improve the potential for accurate diagnostic interpretation, for example by improving the conspicuity of microcalcifications.

2. METHOD

2.1 Blur detection method

An "ideal" mammogram has high contrast between adjacent tissues and has a sharp appearance, such that small features such as Cooper's ligaments, can be distinguished. In an image with patient motion, the tissue signal is expected to be effectively 'smeared' across the image in some fashion as it is recorded by a set of neighbouring detector elements during a given exposure, rather than a single element. The image contrast will thus be reduced in regions with motion blur, compared to the relatively sharp signal transitions in regions without motion blur. Note that one could not detect motion blur in a perfectly "smooth", or homogeneous, tissue region in a mammogram. Thus, we are interested in identifying reductions in a high frequency (HF) portion of the signal for blur analysis due to patient motion, and can remove a low-frequency approximation, S, to a mammogram region of interest (ROI), as follows:

$$\mathrm{HF}(x,y) = \mathrm{ROI}(x,y) - \mathrm{S}(x,y), \tag{1}$$

where S is any appropriate smooth approximation to the low-frequency signal in the ROI at image pixel locations, (x,y). A second order polynomial fit was used in this work. A useful method to interrogate the variation in signal intensities of HF is the noise power spectrum, which will give us information at multiple spatial frequencies. This requires taking the Fourier transform of HF. However, a direct Fourier transform of HF would result in ringing in the resulting spectrum due to sharp signal discontinuities at the ROI edge. This can be corrected by an appropriate "window" function, W, centred on HF, that tapers its signal towards zero at the edges. Following nomenclature and methodology used for the characterization of 'anatomical noise' in mammographic images, which includes breast tissue signal in the evaluation rather than pure noise, the normalized image Wiener Spectrum (nWS) of HF was computed[7]. Briefly, a linearized mammogram is broken up into a series of half-overlapping ROI, and used in the following computation:

$$\mathrm{nWS}(f_x, f_y) = \frac{p_x p_y}{N_x N_y} \langle \frac{|\mathcal{F}\{W_{xy}[\mathrm{HF}(x,y)]\}|^2}{W_{xy}[\mathrm{HF}(x,y)]^2} \rangle, \tag{2}$$

where: the Fourier transform ($\mathcal{F}$) is taken of each HF(x,y), scaled with a unit-energy-normalized Tukey window function W_{xy}, N and p are the ROI side lengths and pixel pitches in the x and y dimensions, respectively, and the angled brackets indicate an ensemble average over the individual nWS from all ROI.

For typical patient imaging conditions, low-frequency signal fluctuations are usually dominated by variations in tissue composition and thickness rather than by system noise. This manifests as a large nWS magnitude at low spatial frequencies for the swirl phantom image nWS compared to the uniform phantom image nWS in Fig. 1a, which decreases with increasing spatial frequency in a manner characteristic to the signal pattern, until the spectra converge at spatial frequencies dominated by quantum noise. This relationship between the WS and spatial frequency has been shown to be remarkably consistent for mammograms, with an average value of the WS slope of about 3 on a log-log scale[8–10]. A comparison of the nWS in Fig. 1a and 1b, demonstrates that the swirl phantom and the sharp mammogram have very similar spectral shapes. Furthermore, the effect on this slope of repositioning of the same breast between CC and MLO views has been shown to be small, with an average difference of less than 3%[11]. We have recently reported on noise estimation using the image WS[12], and believe that the WS carries useful information for mammographic image quality assessment as it measures both image physical properties and quantifies aspects of the tissue appearance. Thus, it is the premise of this work that a method to detect image unsharpness due to patient motion can be developed that evaluates the nWS within a region of the spectrum sensitive to contrast from tissue structure. It was hypothesized that in an unsharp image, the nWS should be lower than the nWS of an image of the same tissue with a sharp appearance. Fig. 1b demonstrates via the nWS of a blurry and sharp mammogram of the same anatomical view, that this can be true over at least a limited spatial frequency band.

A blur detection method was developed that performs comparative assessment of image nWS for all views acquired within a patient study. This approach relies on the evidence in the literature that the nWS arising from breast tissue is reasonably consistent between views and between breast laterality, such that the main source of disparity between the nWS from all views may be due to patient motion. One limitation to this method is that there is a relationship between the WS and breast density[10], so the nWS may vary substantially with tissue composition and complexity on a local scale for a given mammogram, and between breasts with different densities. This is one reason why the nWS comparison was restricted to intra-study analysis, thereby minimizing differences in tissue characteristics.

The exposure level and x-ray spectrum have a strong influence on the WS magnitude[11]. In order to mitigate these effects, and to allow for comparison of nWS between images, the WS is normalized by the linearized image magnitude as shown in Equation 2, and the nWS is also shifted to a common value at 4 mm^{-1}, where quantum noise dominates the spectra. This results in the similar nWS magnitudes between the blurry and sharp mammogram in Fig. 1b despite the images being acquired under different conditions.
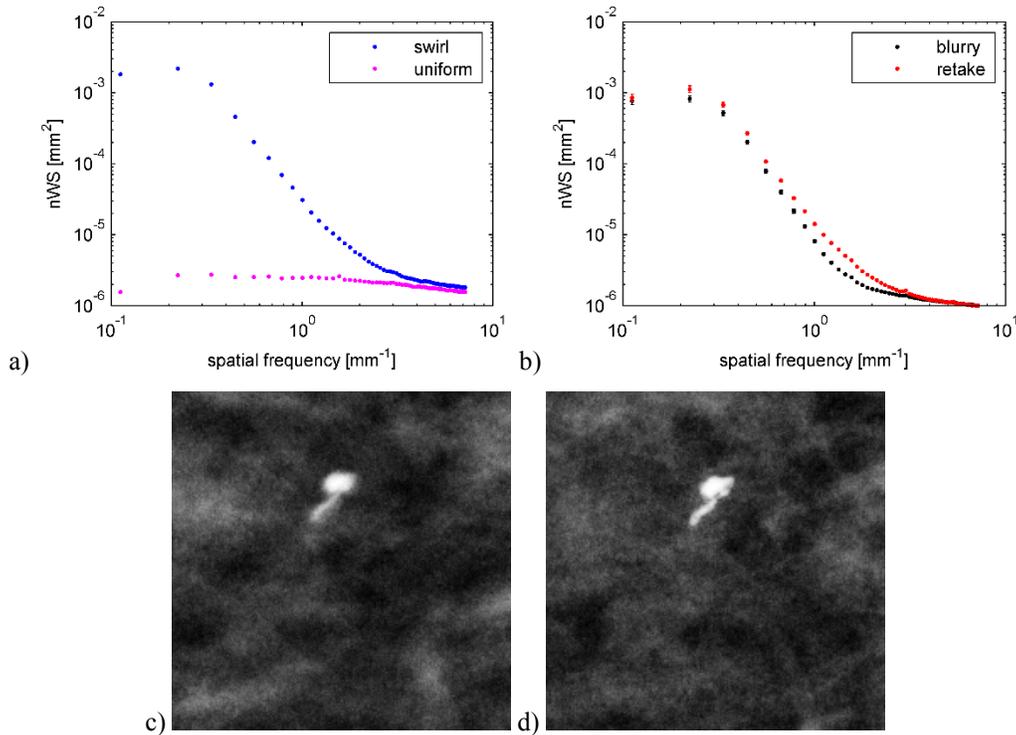

a) b)

c) d)

Figure 1. a) 1D radially averaged nWS of a 4 cm CIRS swirl phantom, and a 4 cm 50/50 uniform phantom acquired using the same technique factors; W/Rh, 28 kVp, 90 mAs. b) 1D radially averaged nWS of a blurred mammogram and the retake of the same anatomical view, shifted to a common nWS magnitude at 4 mm^{-1}. c) A 1.4 × 1.4 cm^2 region of the mammogram with blur; W/Ag, 30 kVp, 134 mAs, 71 mm. d) The corresponding tissue region from a retake of (c); W/Rh, 32 kVp, 153 mAs, 67 mm.

For the nWS computation, an image mask is first determined for each view, which extends to the skin-line, but excludes significant artefacts. The mask was delineated manually for this study. Next, the locations of 128×128 ROI are selected from within the mask, and the nWS are computed for all images in each exam. The image with the greatest average nWS over a defined spatial frequency range is selected as the reference 'sharp' image. The spatial frequency range was selected based upon an iterative optimization procedure for this study, as the range which gave the greatest R^2 value for multivariate linear regression of the algorithm output against the average reader blur severity VAS. The resulting spatial frequencies, 1.6 to 2.4 mm^{-1} were in good agreement with estimates in the literature of the minimum amount of simulated motion which can be perceived by human observers[13]. Next, the uncertainty in the nWS is computed, which depends on the ROI size and the number of ROI used in the nWS computation. The nWS from small regions of all mammograms are compared to the reference ensemble mean nWS. Regions with nWS that lie outside of a 95% confidence interval of the reference are identified as unsharp. Using this data, a number of metrics of image unsharpness

are computed, including: 1) mean blur magnitude, which is the average fractional difference between the reference nWS and the significantly different nWS across the selected spatial frequency window; 2) maximum blur magnitude, which is the maximum fractional nWS difference; 3) blur magnitude standard deviation, which is the standard deviation of the fractional nWS difference; 4) blur area, which is the absolute area in cm^2 where blur was detected; and 5) fractional blur area, which is the fraction of analyzed mammogram area that was identified as having blur above a minimum magnitude threshold. Finally, a 'blur heatmap' is output as a visual representation of the locations and relative magnitudes of significant nWS differences.

The methodology adopted in the blur identification algorithm was developed iteratively by comparing outputs of the various blur metrics under testing to reporting made by expert readers.

2.2 Clinical data

A total of 25 four-view Hologic Dimensions mammography exams, which included at least one blurred image, were selected by an expert radiologist, S.V., from among screening mammograms with both raw ('For Processing') and processed ('For Presentation') image formats available.

The institution's research governance manager deemed ethical approval unnecessary for this study. All images were de-identified for use and local NHS Caldicott Guardian approval was obtained.

2.3 Reader study

Two readers with substantial clinical experience assessed the images; one radiologist, A.E., and one radiographer screen-reader, P.W. The 'For Presentation' images were displayed using the MedXViewer (Scientific Computing Department, Royal Surrey County Hospital) in rooms with low ambient lighting on 5 MP monitors at default magnifications of 0.46 or 0.59, with an available option to zoom to full resolution.

No specific instructions were given to the readers, and they were not blinded to the study design. However, they did not know how many of the images had been pre-classified as blurred by the radiologist who selected the cases. Readers were asked to complete a form with their interpretation of each view, where they reported whether or not there was blur, and if there was blur whether or not a retake was required on grounds of blur. The means by which the blur was noticed was recorded, and a reason was given if no retake was required despite the presence of blur. The location of blur was sketched on a diagram, and a Visual Analogue Scale (VAS) was used to capture the interpretation of blur severity on a scale from 0 (no blur) to 100 ("worst blur imaginable"). The presence of any suspected abnormalities was noted, and scored for the level of suspicion on a 5-point scale. The Breast Imaging-Reporting And Data System (BI-RADS) breast density (5th Edition) was also recorded.

2.4 Data analysis

Correct blur identification by the algorithm was assessed by comparing the automated blur metrics with the expert opinion for each image. The mean blur severity VAS of the two readings was used as the 'ground truth' for algorithm development. Interobserver variability was assessed on this measure using the intraclass correlation coefficient (ICC), treating blur severity as a continuous outcome variable, and calculation based on a single-rater, absolute-agreement, 2-way random-effects model. Cohen's kappa was determined for reader agreement on blur presence and retake.

Univariate linear regression was performed on individual algorithm blur metrics against the average blur severity VAS as well as the scores for individual readers. Multiple linear regression using blur predictor variables was used to develop an automated blur severity score. The Akaike information criterion (AIC) was used to determine the predictors for use in the model. The uncertainty of the model was estimated using bootstrapping, by training using leave one-out analysis and testing the predictions with replacement.

The image area affected by blur was qualitatively categorized from reader sketches for comparison with algorithm output using the following descriptors: 1) less than 25% of the view (0 to <0.25); 2) less than 50% of the view (0.25 to <0.50); and 3) all or most of the view (0.5 to 1). In the case of disagreement, the maximum of the blur area categories was used, which favors more sensitive blur reporting, but avoids null area fractions for cases where only one reader identified blur. Correspondence between blur area category and continuous measures (e.g., automated blur fractional area) was tested using Spearman rank correlation.

3. RESULTS

The mean patient age was 61 [51 to 75], with a mean compressed breast thickness of 63 mm [40 to 86]. Using the Volpara algorithm v1.5.4 the mean volumetric breast density (VBD) was found to be 6.8% [2.5 to 16.5]. Visual BI-RADS assessment categorized the breast density of the 25 cases as 6 category A, 8 B and 11 C.

Of the 100 individual views, 35 were identified by at least one reader as having some amount of blur. Cohen's Kappa for the agreement of the readers on the presence of blur was 0.84, with 80% agreement. Of the 35 views with blur, 20 were identified for retake by one or more readers, with Cohen's Kappa of 0.22 based on 35% (7/20) agreement. The ICC for the VAS blur severity score was 0.84. Given the excellent agreement of the readers on the presence and severity of blur, the average blur severity VAS was used for development of the automated patient motion blur detection algorithm, while the prediction of a retake decision is out of the scope of the present work.

Table 1 lists the coefficients of determination, R^2, for the univariate linear regression of each blur algorithm metric vs the average reader blur severity VAS, and the scores of readers 1 and 2 individually for the 35 views with some amount of blur. The Spearman rank correlation between the reader blur area fraction categories and the individual reader blur severity VAS for these images is 0.71, and 0.87, for readers 1 and 2 respectively. Thus, it is believed that the blur severity VAS may reflect both blur magnitude and the size of the affected area, although it may be that when the physical movement "distance" is high, the proportion of the breast that moves is correspondingly high.

Table 1. Univariate coefficient of determination, R^2, for the linear regression of each blur algorithm metric against the VAS blur severity for the averaged reader scores, reader 1 scores, and reader 2 scores. All results are statistically significant, with $p<0.05$ for the F-statistic of the regression, unless indicated with the notation "ns" (not significant).

metric	R^2 (avg.)	R^2 (reader 1)	R^2 (reader 2)
Mean blur magnitude	0.30	0.19	0.32
Max. blur magnitude	0.35	0.19	0.41
Std. dev. blur magnitude	0.47	0.27	0.54
Blur area (absolute)	0.18	0.07 (ns)	0.25
Blur fractional area	0.52	0.29	0.61
VBD (%)	0.04 (ns)	0.07 (ns)	0.01 (ns)

A sample illustration of the automated patient motion blur algorithm output for two representative cases is given in Fig. 2, with reader blur reports, including location sketches, in Fig. 2b and 2d. The blur magnitude heatmap shows the difference between the nWS of the blurred and sharp reference image, over the sharp image nWS across the image.

Multiple linear regression with parameter selection by AIC minimization, and confirmation by cross validation, resulted in a predictive model for blur severity score using three variables: the standard deviation of blur magnitude ($p<0.1$), blur fractional area ($p<0.01$) and VBD ($p<0.01$). It should be noted that all blur metrics in Table 1 had significant cross correlations (not shown) except for VBD. Overall, the algorithm-generated predictive blur severity score had a coefficient of determination, $R^2=0.66$, for linear regression against the average reader blur severity VAS, as shown in Fig. 3a. For regression against the VAS of reader 1, $R^2=0.44$, and for reader 2, $R^2=0.70$. The adjusted coefficients of determination and their standard deviations are 0.63 ± 0.02, 0.38 ± 0.02, and 0.67 ± 0.02, between the predicted blur severity and the averaged, reader 1, and reader 2 VAS severity, respectively. The Spearman rank correlation between the maximum of the reader blur area fraction categories and the algorithm-estimated area fraction with blur is 0.70, with rank correlations of 0.64 and 0.72 between the categorical area fractions of readers 1 and 2, respectively. Fig. 3b demonstrates the correspondence between the algorithm-estimated area fraction with blur and the max. of the categorical reader-estimated blur area fractions with a box-and-whisker plot.
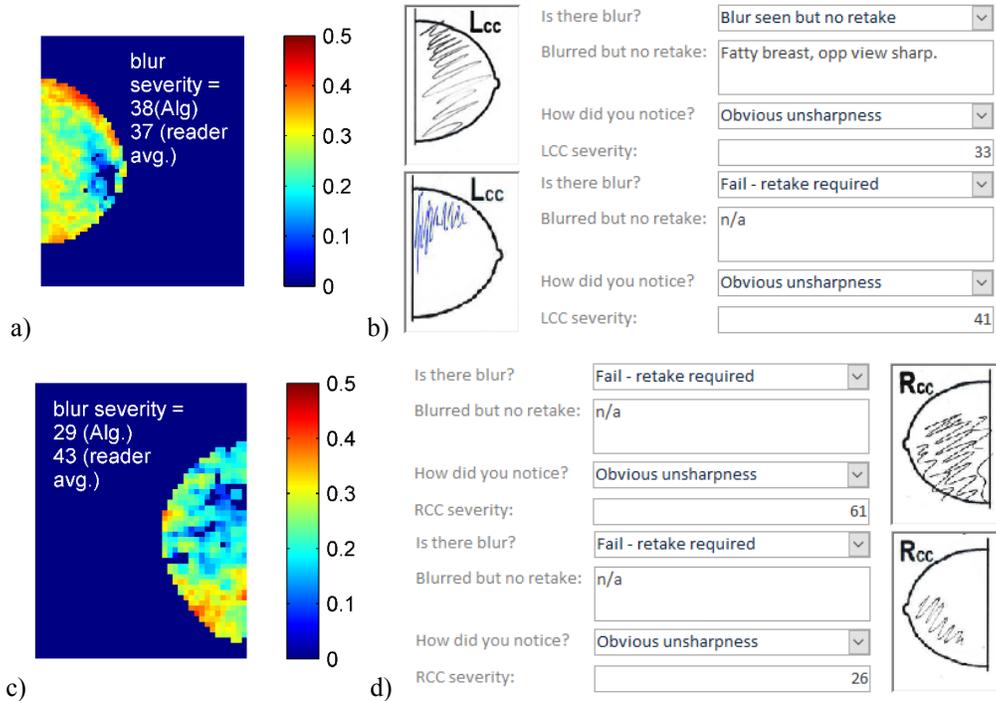
Figure 2. Example automated patient motion blur algorithm output for two mammograms. The blur magnitude heatmap for the blurry mammogram shown in Fig. 1 is given in (a), with reader reports in (b). The blur magnitude heatmap for a second blurred mammogram is shown in (c), with the corresponding reader blur reports in (d).
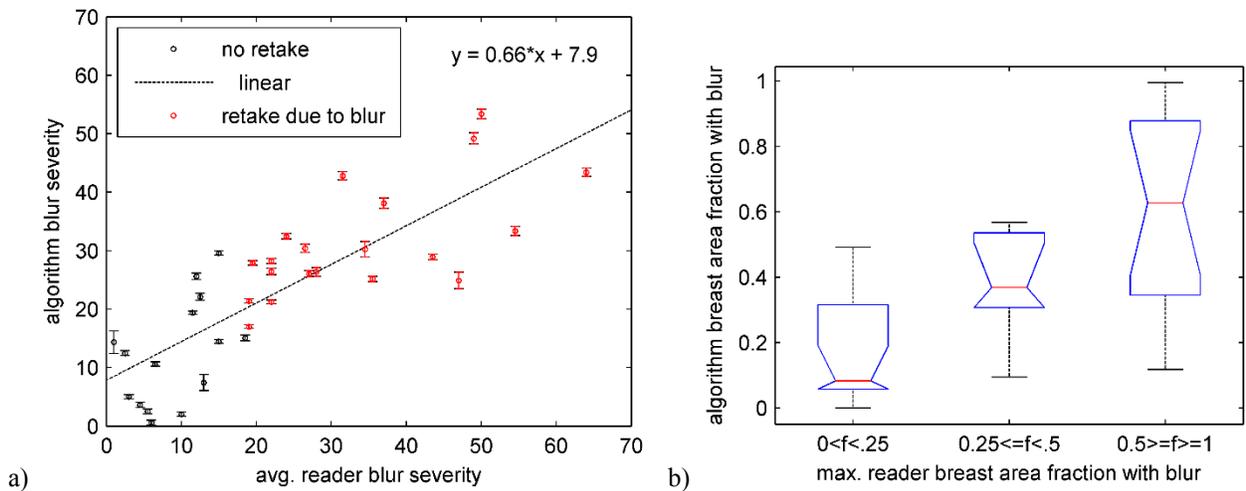

Figure 3. Algorithm performance compared to expert readers. a) Plot of algorithm-estimated blur severity score vs the average reader blur severity VAS for the 35 blurred views. The views called by at least one reader for retake are highlighted in red. Error bars denote standard deviation, as estimated by bootstrapping using leave-one-out model fits, with replacement for testing. b) Box-and-whisker plot showing the relationship between the algorithm-reported fractional area of the mammogram with blur, and the maximum of the area fraction category suggested by the readers' blur location sketches.

4. DISCUSSION

This study represents the initial phase of development and the limitations are numerous at this stage. For example, the number of cases was small, which limited the range of breast characteristics and variety of blur presentations. Only two readers participated in the study, who both practice at the same center. Although this was likely helpful to achieve reasonable consensus, it may have biased the blur severity prediction model towards local practices. Differences in individual reader behavior compared to the algorithm are evident based on the coefficient of determination for regression of the algorithm-reported blur severity against reader 1 VAS at 0.38, compared to 0.67 for reader 2. This difference in performance likely arose due to the apparent stronger association between the fractional area of blur and blur severity VAS for reader 2 compared to reader 1. An extended study with more readers is planned for future work to ensure that the algorithm performance is representative of a broad cross-section of readers. An evaluation of intra-observer variability would also help to estimate uncertainty for a single reader, which was not assessed here.

Only images from one system type were included, and performance on other technologies remains to be tested, especially systems using a different detector type. Furthermore, the results presented here should not be interpreted as being indicative of broader performance on Hologic images because an independent data set was not used to test the algorithm performance after training. Leave-one-out analysis was used to estimate uncertainty in the blur severity estimates, but the lack of high breast densities from the BI-RADS D category means that this model may not be comprehensive. This is likely reflected by the lack of univariate correlation between VBD and the averaged reader VAS blur severity, despite the significance of VBD as an independent variable in the predictive model for blur severity. Further algorithm development and testing on a larger dataset that includes breasts with BI-RADS category D densities is planned for future work.

The self-reference approach to blur identification, where at least one image within a patient study is assumed to be relatively sharp such that its nWS is selected as a reference, may result in false negatives for cases where all study views are blurred. It is believed that such cases of poor image quality for all views would be rare. No control cases were included, so the potential for false positives was not adequately evaluated. The lack of control cases may have also biased the interpretation of the readers towards a heightened sensitivity to blur compared to typical clinical practice.

All images were selected from screening cases, and without the presence of breast implants, thus, no foreign objects (e.g., clips, wires), scarring, or distortions from previous surgery were present. Furthermore, only the four routine anatomical views were included. Testing should be carried out on non-standard views, such as magnification views, to validate the approach for other applications such as morphological evaluation of microcalcifications.

Algorithm measurements were made on raw mammograms, while readings were done using processed images. We believe that the use of raw mammograms should permit high sensitivity to blur, but there is likely to be some disconnect between the blur visualized in the processed images compared to the raw images, as image processing may sharpen the appearance of the image, alter the contrast, and enhance the signal at the breast periphery. This remains to be evaluated in a future study on the manifestation of blur in different image formats.

One reason why the prediction of a retake decision was not addressed in the present work is that there was low agreement between the two readers on this decision at just 35%. An extended study with additional readers may improve the overall agreement on retakes. But there are also believed to be multiple additional factors considered in a retake decision, such as breast density and location of blur relative to features of interest that were not assessed in the present work. Nonetheless, Fig. 3a demonstrates that there is a strong correspondence between increasing values of the average reader-reported blur severity VAS and the likelihood that at least one reader identified a need for retake. Above a threshold of about 20 average VAS, all images had a least one reader recommend a retake on the basis of motion blur. A larger study is needed to confirm this relationship, but these results are encouraging for the potential utility of the algorithm-reported blur severity scores to assist retake decisions.

5. CONCLUSIONS

The good correlation between the algorithm-estimated blur severity score and the reader-reported scores, and the apparent accuracy of the blur heatmap suggest that this method should serve as a useful aid for blur identification and interpretation. Future work will involve extensive testing using images from multiple vendors, control cases, and additional readers to evaluate method performance under a wider range of conditions.

ACKNOWEDGEMENTS

REFERENCES

[1]   Saunders, R. S., Baker, J. A., Delong, D. M., Johnson, J. P. and Samei, E., "Does image quality matter? Impact of resolution and noise on mammographic task performance," Med. Phys. **34**(10), 3971–3981 (2007).

[2]   Choi, J. J., Kim, S. H., Kang, B. J., Choi, B. G., Song, B. and Jung, H., "Mammographic artifacts on full-field digital mammography," J. Digit. Imaging **27**(2), 231–236 (2014).

[3]   Hogg, P., Szczepura, K., Kelly, J. and Taylor, M., "Blurred digital mammography images," Radiography **18**(1), 55–56 (2012).

[4]   Public Health England., "NHS Breast Screening Programme Consolidated Standards," 28 (2017).

[5]   Kinnear, L. and Mercer, C., "The detection of visual blurring in 1MP and 5MP monitors within mammography clinical practice," Imaging Ther. Pract., M. Armstrong, Ed., 23–28 (2016).

[6]   Ma, W., Hilton, B., Borgen, R., Kelly, J. and Hogg, P., "Comparative analysis of visual blurring detection in mammography images - clinical room monitor versus reporting grade monitor: initial results," Eur. Congr. Radiol., C-0316, European Congress of Radiology 2016 (2016).

[7]   Bochud, F. O., Verdun, F. R., Hessler, C. and Valley, J.-F., "Detectability of radiological images: the influence of anatomical noise," SPIE Phys. Med. Imaging **2436**, H. L. Kundel, Ed., 156–164, International Society for Optics and Photonics (1995).

[8]   Burgess, A. E., "On the detection of lesions in mammographic structure," Proc. SPIE **3663**, 419–437 (1999).

[9]   Engstrom, E., Reiser, I. and Nishikawa, R., "Comparison of power spectra for tomosynthesis projections and reconstructed images," Med. Phys. **36**(5), 1753–1758 (2009).

[10]   Mainprize, J. G., Tyson, A. H. and Yaffe, M. J., "The relationship between anatomic noise and volumetric breast density for digital mammography," Med. Phys. **39**(8), 4660–4668 (2012).

[11]   Hill, M. L., Mainprize, J. G., Carton, A.-K., Saab-Puong, S., Iordache, R., Muller, S., Jong, R. A., Dromain, C. and Yaffe, M. J., "Anatomical noise in contrast-enhanced digital mammography. Part II. Dual-energy imaging," Med. Phys. **40**(8), 81907 (2013).

[12]   Hill, M. L., Mackenzie, A., Khan, N., Wang, K. and Highnam, R. P., "Validation of noise estimation for a clinical contrast-to-noise ratio for digital mammographic imaging," SPIE Phys. Med. Imaging **10573**, 199 (2018).

[13]   Ma, W. K., Aspin, R., Kelly, J., Millington, S. and Hogg, P., "What is the minimum amount of simulated breast movement required for visual detection of blurring? An exploratory investigation," Br. J. Radiol. **88**(1052), 20150126 (2015).