

Is it blur or is it texture? An analysis of computational blur and texture measures commonly used in medical image analysis

Andrew William Chen¹ and Murray H. Loew^{1*}

The George Washington University, Department of Biomedical Engineering, Washington, District of Columbia, United States

Abstract

Purpose: In this work, we endeavor to investigate how texture information may contribute to the response of a blur measure (BM) with motivation rooted in mammography. This is vital as the interpretation of the BM is typically not evaluated with respect to texture present in an image. We are particularly concerned with lower scales of blur (≤ 1 mm) as this blur is least likely to be detected but can still have a detrimental effect on detectability of microcalcifications.

Approach: Three sets of linear models, where BM response was modeled as a linear combination of texture information determined by texture measures (TMs), were constructed from three different datasets of equal-blur-level images; one of computer-generated mammogram-like clustered lumpy background (CLB) images and two image sets derived from the Brodatz texture images. The linear models were refined by removing those TMs that are not significantly non-zero across all three datasets for each BM. We use five levels of Gaussian blur to blur the CLB images and assess the ability of the BMs and TMs to separate the images based on blur level.

Results: We found that many TMs used frequently in the reduced linear models, mimicked the structure of the BMs that they modeled. Surprisingly, while none of the BMs could separate the CLB images across all levels of blur, a group of TMs could. These TMs occurred infrequently in the reduced linear models meaning that they rely on different information compared with that used by the BMs.

Conclusion: These results confirm our hypothesis that BMs can be influenced by texture information in an image. That a subset of TMs performed better than all BMs on the blur classification problem with the CLB images further shows that conventional BMs may not be the optimal tool for blur classification in mammogram images.

© 2023 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.10.S1.S11920](https://doi.org/10.1117/1.JMI.10.S1.S11920)]

Keywords: blur measures; texture measures; clustered lumpy background; mammography.

Paper 22361SSR received Dec. 19, 2022; revised manuscript received Apr. 4, 2023; accepted for publication May 1, 2023; published online May 24, 2023.

1 Introduction

When using a medical image as raw data for anything from tumor detection to disease stage grading, the investigator must at some point decide what quantitative information they will extract from the image and how they will do it. Texture measures (TMs) are an extremely popular method to extract features from an image or region of an image and then to perform some sort of classification.¹ In essence, a TM mathematically summarizes some property of the image pixel intensity values and/or distribution. TMs that quantify the heterogeneity of pixel intensity values of a tumor image may be useful in grading cancer stage for example. Blur measures (BMs) provide an additional method of extracting quantitative information from images; however, the implementation of BMs is often more concerned with image quality assessment as opposed to

*Address all correspondence to Murray H. Loew, loew@gwu.edu

the diagnostic application that TMs are typically used for.²⁻⁴ This paper examines the hypothesis that texture information can affect the output of BMs to the point where the output of the BM is not significantly dependent on the level of blur present in the image.

BMs have largely been used for applications such as shape from focus (SFF) or shape from defocus (SFDF).⁵⁻⁸ These methods would be referred to in the literature as full-reference BMs. A concise definition of full-, reduced-, and no-reference BMs can be found in Ref. 9. Briefly, a full-reference BM is computed using both a blurred image and a reference image such as the original unblurred scene. A reduced-reference BM has only partial access to information about the original image which is how SFF and SFDF operate. SFF and SFDF are depth detection methods that construct a map of a three-dimensional scene or object by controlling the focal length of the imaging device. For these methods, a blur level can be calculated for every pixel in the image using a BM of choice for all the images captured at each focal length. The focal length that minimizes the amount of local blur at a given pixel is used to assign that pixel a distance or depth. The full- and reduced-reference methodologies, by their nature, allow the BMs used to be generally robust as opposed to no-reference BMs. The robustness of the SFF/ SFDF methods comes from the fact that a BM must accurately characterize the blur level of a patch relative only to blurred versions of itself, not relative to other patches. In contrast to full- or reduced-reference BMs, a no-reference BM computes a value for blur in an image patch or over an entire image using only a single image. Using a no-reference approach is required for imaging scenarios such as mammograms, as there is no user-controllable focal length such as in SFF and SFDF applications. As no-reference BMs lack the robustness intrinsic to full-/reduced-reference BMs, it is critical to examine the interaction of texture with BMs in computer vision.

Blur in mammograms can be caused by motion of the patient, motion of the compression plates, or vibration translated through the machine.¹⁰ Blur in mammograms was the cause of 87% of repeat imaging according to an internal UK hospital audit.¹¹ In this study, we are particularly interested in small levels of blur (≤ 1 mm, see Sec. 2.2). This region of low levels of blur is of special interest in mammograms for two reasons. The first is that pathology-relevant information can exist at these low scales. Microcalcifications smaller than $500 \mu\text{m}$ are more likely to be malignant than larger ones,¹² and calcification clusters under 1 cm^2 are more suspicious than larger groupings.¹³ Using simulated blur in full-field digital mammography images, Abdullah et al. showed that blur levels as high as 1.5 mm to as low as 0.7 mm decrease the sensitivity of expert human observers of detecting microcalcifications.¹⁴ Compounding these detrimental effects of blur on mammograms is the finding by Ma et al., which shows that blur as low as 0.7 mm is at the lower limit of detection with human observers in mammograms, meaning that not only can low levels of blur affect the detection of microcalcifications, but also the radiologist may not even realize blur is present in the image.¹¹ Second, and directly related to the analysis of BMs and TMs we propose to perform, we hypothesize that blur may go undetected because of a BM's greater sensitivity to texture information than to small levels of blur.

Blur is a property of an image that occurs during image acquisition, usually due to the alignment of the optics of the imaging device or some relative motion of the subject and imaging device during image acquisition. As it applies to medical imaging, the blur information present in an image is regarded as the product of a physical process usually independent of the image subject. Blur affects the image in an undesirable manner and ultimately obfuscates clinically relevant information in an image. Unlike blur, for which there is wide agreement that it can be computationally summarized as low-pass filtering of the imaged subject prior to an additive noise process,^{5,15,16} texture has no real universal definition due to its wide use and variability.^{17,18} In attempts to reference a universal definition of texture, it is often said that intuitively, TMs will describe the smoothness or coarseness of an image or that a given texture will be composed of repeated or tessellated mutually related elements, these mutually related elements often labeled as "textons."¹⁹ Both BMs and TMs use information available in an image to compute some property about the image. A critical difference is the way these measures are interpreted. In practice, a TM is understood to provide information about the texture of the image subject as captured by the imaging apparatus. It is understood that external factors such as differences in medical imaging device make and models will produce different TMs for the same image subject. Computational methodologies have been developed specifically to harmonize the textural information extracted from patient cohorts imaged with different devices.^{20,21} However, when BMs are applied, there is

no consideration given to how the structure of the imaged anatomy may influence the result of a BM. That is to say, we intend the BM to accurately detect the level of blur in a mammogram, but it is imperative that we consider how, for example, very dense breast tissue and fatty breast tissue might affect the output of the BM. This paper examines the hypothesis that texture information can affect the output of BMs to the point where the output of the BM is not significantly dependent on the level of blur present, examining this hypothesis with particular attention to the context of medical imaging. This is an especially pertinent question as the fields of machine learning and deep learning have become increasingly popular over the last decade,²² and an increased pool of published feature extraction methods for blur and texture is available.

To the best of our knowledge, this is the first study that aims to characterize the sensitivity of a population of BMs to a population of various texture information, specifically in the regime of small amounts of blur. This is important not only for mammograms but also for other imaging applications where computer vision is poised to augment or perhaps automate processes, such as digital histopathology or remote sensing applications.

This paper is organized as follows. Section 2 contains the methods used to create our image sets and the means by which we explore the relationships between BMs and TMs. The results of this work are described in Sec. 3. Section 4 discusses the relationship between many BMs and TMs. Conclusions are given in Sec. 5. The measures used are described in more detail in the [Supplemental Material](#).

2 Methods and Data

Three datasets were used for this study. The first set of data was a computer-generated set of images using the clustered lumpy background (CLB) method as described by Bochud et al.²³ The images (Fig. 1) were generated to simulate 1 cm² patches of mammograms at a resolution of 24 px/mm. This size and scale of our images are based on the scale of the images in the INbreast database so as to ensure the effect of our applied blur is of similar scale to real world blur and that the results of the analysis on our set of BMs and TMs have salience in relation to actual mammograms.²⁴ We chose to use these CLB images instead of real world mammogram patches as this gives us control over the texture information present in the patches, as well as eliminating the question of latent blur present in these patches as could be the case with real-world mammograms. One hundred of each image type (scaled, dendrites, fibrous, and lessBlobs), shown in Fig. 1, were created totaling 400 computer-generated images. These images can be thought of as patches from a mammogram dataset. However, the computer-generated images have the

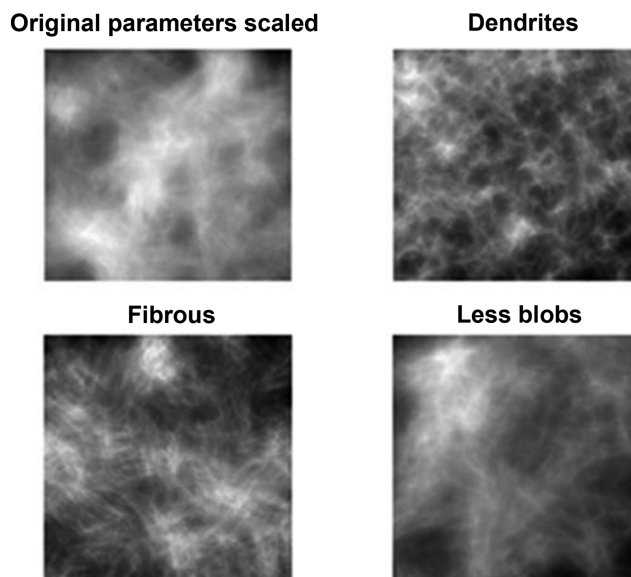


Fig. 1 Examples of the four texture groups of computer generated CLB images.

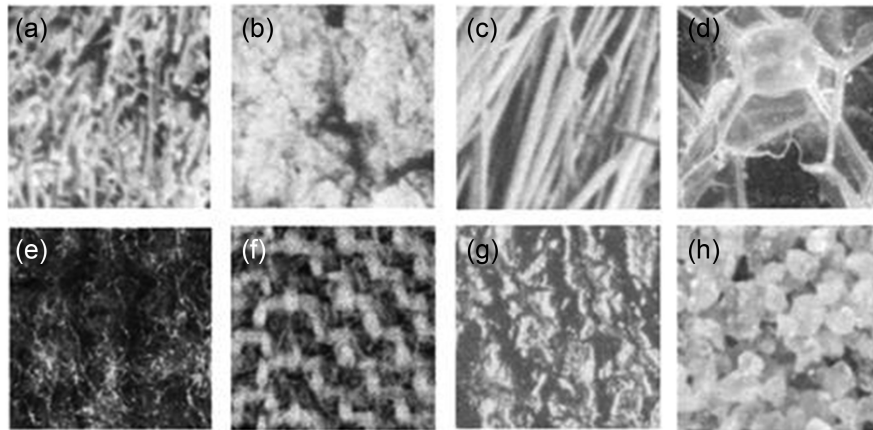


Fig. 2 Representative patches from the Brodatz images used for (a)–(d) the BDZ-1 and (e)–(h) BDZ-2 texture groups. The images from (a)–(h) are grass, bark, straw, bubbles, wool, herringbone, leather, and sand.

advantage of providing control over texture information and thus, precise knowledge of what images do in fact have unique texture, unlike real mammogram images where the ground truth about the texture content is unknown. Since we have generated these 400 images, we are also able to say that they are not blurred which we would not be able to assume if we used real mammogram patches.

A MATLAB script developed by The University of Arizona Wyant College of Optical Sciences was used to generate the CLB images.²⁵ The script provides the input parameters to the function to recreate the images as described by Bochud et al. The images in this study were created from the original Bochud et al. paper with various parameters changed. (The following changes need to be made to the original CLB parameters to obtain the texture groups used in this study. Scaled: $\text{dim} = 238$, $\text{Kbar} = 279$. The parameters changed in the scaled images are the same for all other image groups unless otherwise stated. Dendrites: $\text{Kbar} = 102$, $\text{sigma} = 85.71$. Fibrous: $\text{alpha} = 3$. lessBlobs: $\text{Lbar} = 10$.) This dataset is referred to as the CLB images.

Two additional sets of 400 images were created from real world images using the Brodatz texture images.²⁶ The Brodatz images were chosen as they present apparent and strikingly different image textures. In addition, the use of a variety of image texture datasets means that our later determination of any BM dependence on texture information will be more robust. Four different images were selected for each of the two additional datasets (Fig. 2). One hundred 128×128 image patches were created from each of the Brodatz texture images to create a total of 400 images for each of the two real world image sets. In generating patches from the original images, 39 pixels of overlap were needed between patches on all sides to generate a sufficient number of images. This means that between any two image patches a maximum of 30.5% of pixels would be shared. While we can assume that all of the Brodatz images are at the same blur level, these real world images do not provide any specific control over the scale of our image patches as with the CLB images. These datasets are referred to as BDZ-1 and BDZ-2.

2.1 Blur and Texture Measure Relationship Investigation

A total of 31 BMs²⁷ and 54 TMs^{28–33} were used. These BMs and TMs have broad applications from blur segmentation to scanned document quality assessment as well as a variety of biomedical applications including mammogram classification.^{34–38}

The TMs used in this study come from a number of families of TMs. One of the more common is the Haralick or gray level co-occurrence matrix measures, of which we use 19.³⁰ These features are calculated from the co-occurrence matrix of an image which can be thought of as a two-dimensional (2D) histogram of the frequency of occurrence of neighboring pixels with certain intensities. Neighboring pixels are defined by the user with an angle and distance parameter. Another class of TM is the Laws Kernel measures of which we used 16. The Laws

Kernel TMs are computed by applying a set of 2D convolution kernels to the image which themselves are computed from a set of one-dimensional convolution kernels. These kernels mimic and are designed to detect smooth pixel intensity level changes, edges, spots, and ridges. Another class of TM used is the gray level run length matrix (GLRLM) features of which 13 were used in this study. GLRLM features are computed from the run length matrix which can be thought of as a 2D histogram of the frequency of runs or sequential pixels of discrete gray level values of various lengths. The final class of TMs used in this study is calculated from the histogram of the entire image. Six features were calculated from the image histogram including the mean, variance, skewness, kurtosis, energy, and entropy. A detailed mathematical definition of all texture features used can be found in the [Supplemental Material](#).

While not all BMs used can be neatly categorized into families of measures as with TMs, there are a number of BMs derived in similar ways. One such way are discrete cosine transform (DCT) BMs which use the coefficients from the DCT of the image to compute a measure of pixel intensity variation known as energy. Similarly, a set of wavelet transform BMs are calculated by computing the discrete wavelet transform of the image. A set of Laplacian BMs are calculated from convolving the image with a series of Laplacian masks and performing some measure of pixel intensity energy or variance from the resulting convolution. As with the TMs, a detailed mathematical definition of all BMs used can be found in the [Supplemental Material](#).

All TMs and BMs were computed over the entire image to generate a single value for each of the image texture and blur. Artificial blur was added to the images using a Gaussian blur model¹⁸ using a built in MATLAB R2019a function. The blur levels were based on filter sizes of 0.2, 0.4, 0.6, 0.8, and 1 mm generated based on the scale of the CLB images and were created by defining appropriate values for the standard deviation. In our model of blur, we choose to ignore the reintroduction of additive noise after applying the Gaussian blur model. This is in part because we do not have access to the original level of noise present in the Brodatz images. The original unblurred Brodatz image patches do contain some noise which was generated by the camera originally used. Similarly, the CLB image-patch generation algorithm is designed to have a power spectrum that mimics actual mammograms and so has the high frequency spectral information based on actual mammogram noise built into the design. We decided, therefore, not to add an arbitrary amount of additive noise after the Gaussian blurring process.

Each of the 54 TMs and 31 BMs was first calculated on the three datasets of unblurred images. The TMs would inform us if we do indeed have four unique textural groups in each of our three datasets of images. Similarly, the result of the BMs on the unblurred image sets will inform us of the validity of the driving hypothesis of this work, that BMs may be inaccurate due to textural information. If we do have distinct textural groups and BMs are affected by the presence of different texture at equal-blur-level, then we would expect so be able to separate the image groups using the TMs and BMs. Separability was determined by one-way ANOVA performed between image types in each of the three datasets for each TM and BM individually.

To determine BM dependence on texture information, a generalized linear regression model (GLM) was created to predict the output of each BM individually where each model uses all 54 TMs on the unblurred images as predictor variables, thus each TM has 400 observations. Separate models were created for the CLB, BDZ-1, and BDZ-2 image sets. The model can be represented as

$$\mathbf{BM} = \sum_{i=1}^{54} \mathbf{TM}_i(\mathbf{I}), \quad (1)$$

where \mathbf{BM} represents an individual BM and $\mathbf{TM}_i(\mathbf{I})$ represents the output of one of the 54 TMs used on an individual image \mathbf{I} . A BM's dependence on any particular texture information in an image was determined by variables in the GLM having coefficients that are significantly non-zero. The adjusted R^2 values for the models are also reported.

A sensitivity analysis was performed on all of the 31 linear models. All predictor variables (TMs) were held constant at their mean value except for one TM that was varied across its 400 values, generating 400 responses. This was repeated for all 54 TMs across all 31 linear models. The sensitivity measure was defined as the ratio of the standard deviation of the response variables (BM) to the standard deviation of the corresponding varied TM.

To refine our results and obtain a better understanding of the effect that certain textural information has on the output of various BMs, we created a second set of linear models in which each GLM contained a subset of the original measurement variables. Only variables from the 31 original linear models that had significantly non-zero coefficients across all three datasets were used to create the reduced linear models. Thus, each of the reduced linear models contains a unique and reduced set of TMs to generate it. These reduced linear models were generated again to predict the output of each BM on the unblurred images and on all three datasets separately. The adjusted R^2 values of these reduced linear models are also reported.

All 54 TMs and 31 BMs were performed on the blurred CLB images; the size of each Gaussian kernel was created relative to the scale of each image. We report the TMs' and BMs' ability to separate blur levels within the four texture groups as well as to separate blur across the four texture groups.

3 Results

Results of the one-way ANOVA test on the CLB images for each TM are shown in Fig. 3(a). Each point on the plot represents the mean and standard deviation of the particular TM indicated on the horizontal axis, calculated on all images in one of the four texture groups as indicated by the point shape and noted in the figure legend. The red diamonds indicate those TMs that separate the four CLB image types significantly. Similar plots can be generated for the two BDZ datasets as well; for the sake of brevity, however, only the results for the CLB data set are shown. Of the 54 TMs used in this study, a total of 27, 20, and 36 TMs separated the four image groups

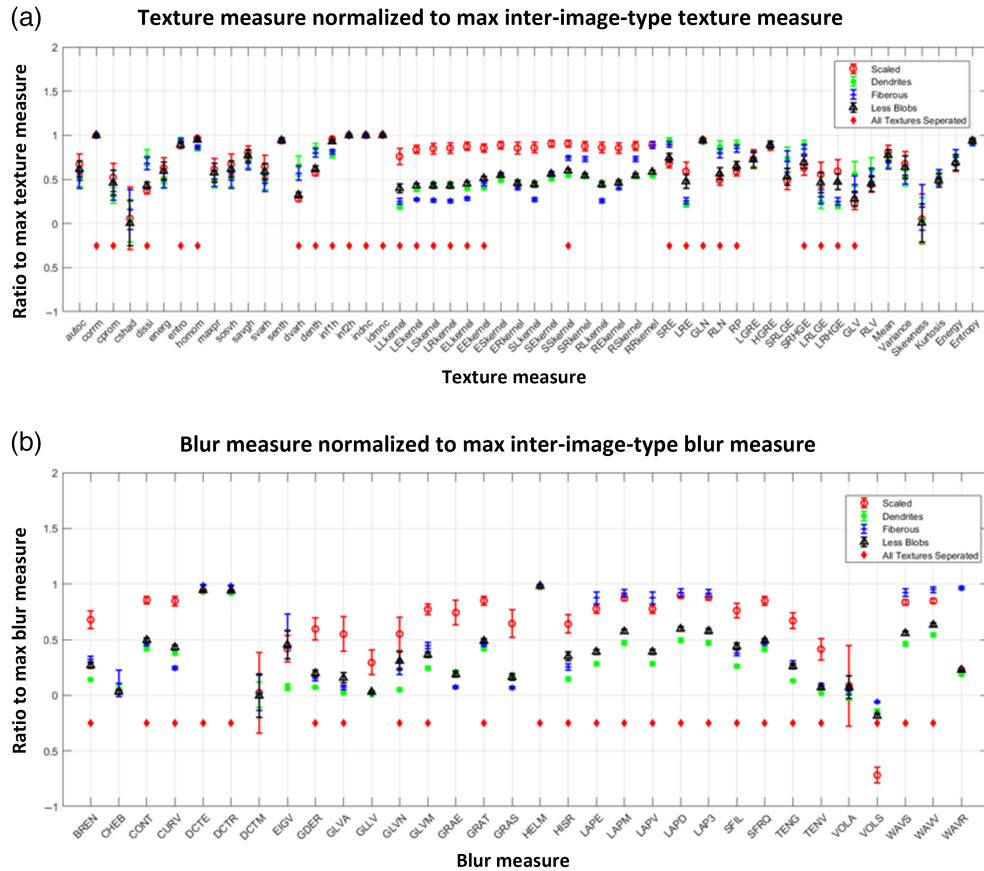


Fig. 3 (a) Normalized means and standard deviations for all TMs or “texture profiles” of the four unblurred CLB texture groups. (b) Normalized means and standard deviations for all BMs on the four unblurred CLB texture groups. In both plots, red diamonds indicate measures that separate all four texture groups with significance ($p < 0.05$).

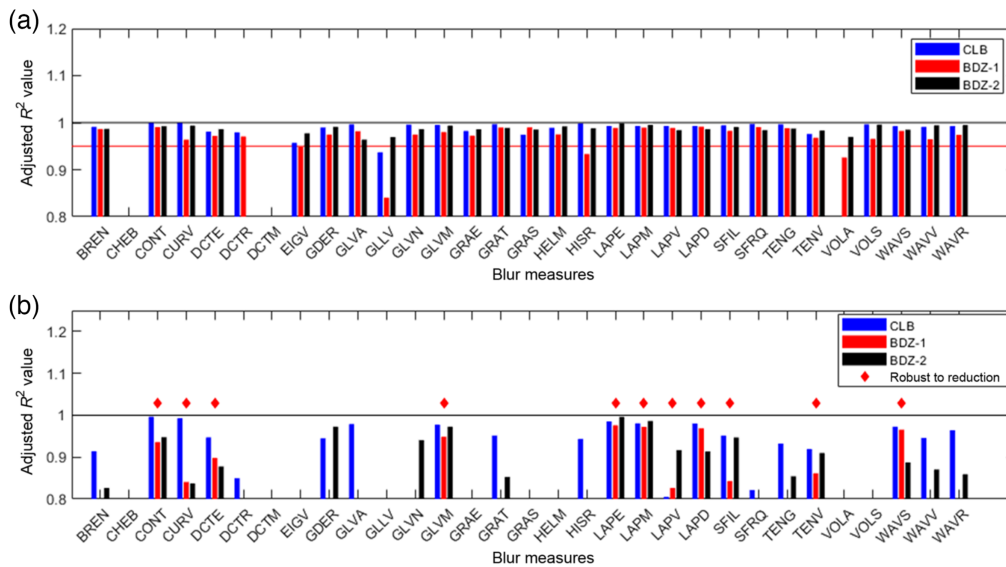


Fig. 4 The adjusted R^2 values of the linear model for each BM composed of (a) all TMs and (b) reduced TMs. TMs were reduced by using only those TMs that were significantly non-zero across all three linear models created for each dataset for that particular BM. Linear models that were considered robust to a reduction in variables are marked with red diamonds.

significantly for the CLB, BDZ-1, and BDZ-2 datasets, respectively. Similarly, Fig. 3(b) indicates BMs that separate the four CLB image types well as determined by a one-way ANOVA test and are denoted by red diamonds. Of the 31 BMs used 23, 16 and, 23 BMs separated the four image groups well for the CLB, BDZ-1, and BDZ-2 datasets, respectively. Significant separation was established at $p < 0.05$.

Figure 4(a) summarizes the performance (adjusted R^2) of the linear models as predictor models for each BM using all 54 TMs. The red line in Fig. 4(a) is at the adjusted R^2 value of 0.95 which is the accepted standard for a linear model with good fit. Only 6 of the 31 BMs investigated (CHEB, DCTR, DCTM, GLLV, HISR, and VOLA) were not modeled well as linear combinations of all 54 TMs for at least one of the three datasets; the rest of the linear models were good linear fits across all three datasets. Figure 4(b) displays the adjusted R^2 results for all 31 BMs across all three datasets using a reduced number of TMs as described in Sec. 2.2. A value of 0.8 for the adjusted R^2 value was chosen as the threshold value for linear models that were considered robust to a reduction in the number of measurement variables used to construct them. This reduced threshold for the adjusted R^2 value on the linear models reflects the reduction in features used to calculate the linear models and allows for a larger final pool of features for further analysis. Eleven of the 31 BMs were robust to the reduction of TMs with this threshold and are boldfaced in Table 1 as well as marked with red diamonds in Fig. 4(b). The TMs that were used to create the reduced linear models is reported in Table 2 along with the number of individual reduced linear models each TM appeared in.

A one-way ANOVA was performed on the output of all of the BMs and TMs across all six blur levels for each of the four CLB image groups. Though some BMs were able to separate the lower blur levels from all levels of blur, none of the BMs was able to separate all six groups from each other with significance. This meant that the higher levels of blur (0.6, 0.8, and 1 mm) were often inseparable from each other. Interestingly, 11 TMs were able to separate all blur levels with significance across all of the CLB images. These were cprom, denth, inflh, SRE, LRE, RLN, RP, SRLGE, SRHGE, LRLGE, and LRHGE.

4 Discussion

The analysis undertaken in this work is primarily dependent on our datasets having unique texture information. This is confirmed by way of their texture profiles [Fig. 3(a)]. The fact that

Table 1 BMs used in this study.

BM abbreviation	Shared vars. in linear models (%)	BM abbreviation	Shared vars. in linear models (%)
BREN	0.27	HELM	0.11
CHEB	0.00	HISR	0.22
CONT*	0.31	LAPE*	0.50
CURV*	0.18	LAPM*	0.40
DCTE*	0.47	LAPV*	0.35
DCTR	0.40	LAPD*	0.40
DCTM	0.00	SFIL*	0.18
EIGV	0.15	SFQR	0.05
GDER	0.23	TENG	0.33
GLVA	0.21	TENV*	0.33
GLLV	0.29	VOLA	0.00
GLVN	0.27	VOLS	0.00
GLVM*	0.36	WAVS*	0.47
GRAE	0.17	WAVV	0.21
GRAT	0.16	WAVR	0.19

All BMs used in this study and the percentage of shared significantly non-zero variables (TMs) across the linear models of all three datasets. BMs in bold and with a star (*) indicate robustness to reduction of variables for their linear models.

Table 2 TMs used in the reduced models.

TM abbreviation	Occurrences in reduced linear models	TM abbreviation	Occurrences in reduced linear
corr [*]	3	SEkernel [†]	3
savgh [*]	1	SSkernel [†]	6
denth [*]	1	SRkernel [†]	5
inf1h [*]	1	RLkernel [†]	4
inf2h [*]	1	REkernel [†]	2
LLkernel [†]	8	RSkernel [†]	3
LEkernel [†]	2	SRE [‡]	1
LSkernel [†]	2	RLN [‡]	2
LRkernel [†]	4	Variance [◇]	2
ELkernel [†]	2		
EEkernel [†]	1		
ESkernel [†]	1		
ERkernel [†]	1		
SLkernel [†]	3		

Key: ^{*}, Co-Occurrence; [†], Laws Kernel; [‡], Run Length; [◇], Histogram.

The TMs used in the reduced linear models and the number of occurrences out of the 11 BMs that were considered robust to a reduction of variables.

each of the four unblurred image types in each of our three image sets can be separated well based on the output of more than half of the BMs used in this study demonstrates that textural information can influence the output of a BM and therefore that users of BMs need to practice caution in the way results are interpreted.

One of the surprising results of this study was how well the majority of BMs can be modeled as a linear combination of the 54 TMs. The high (≥ 0.95) measure of goodness of fit for the majority of BMs regardless of dataset used to generate the TMs [Fig. 4(a)] suggests an underlying relationship that links the population of TM outputs to BM performance which is agnostic to the content of the image. To investigate what this structure might be, and so gain a better understanding of how BMs may be affected by textural information, we reduced the number of TMs used to create each GLM.

The reason we chose to reduce the criterion for robustness of a model from an R^2 threshold of 0.95 to 0.8 for the reduced linear models was to account for the possibility of a particular texture having less variability in one of the three image datasets used. Our criteria for retaining a TM in the reduced linear models of a particular BM was that the TM has a significantly non-zero coefficient in all three linear models for that BM across the three image sets. It is conceivable that a TM may contribute significantly to two out of three models for a particular BM and in the third model there is a much lower variability of the particular type of texture detected by that TM in the image dataset used to compute the model. This is also a reason that we chose to perform this study with three distinct datasets, in the event a particular dataset does not contain a high variability of a particular texture that a TM is sensitive to the other two datasets would allow us to uncover the relationship of a particular type of texture influencing the performance of a BM. The reduction in threshold of R^2 can mitigate the effect of a single dataset not having enough variability of a particular texture type minimizing the contribution of a TM to a linear model.

Included in the BMs that could be modeled well with a reduced number of variables were all of the Laplacian based operators. As the structure of the Laplacian BMs and the Laws kernel TMs are similar, it is reasonable that the Laws kernel TMs were so heavily represented in these reduced linear models. For example, the LAPE (Laplacian energy) [Fig. 5(a)] BM shares similar structures with the Laws LLkernel [Fig. 5(c)] and SSkernel [Fig. 5(d)] TMs in their radial symmetry and zero-sum nature, respectively. While the local energy of the image convolved with the

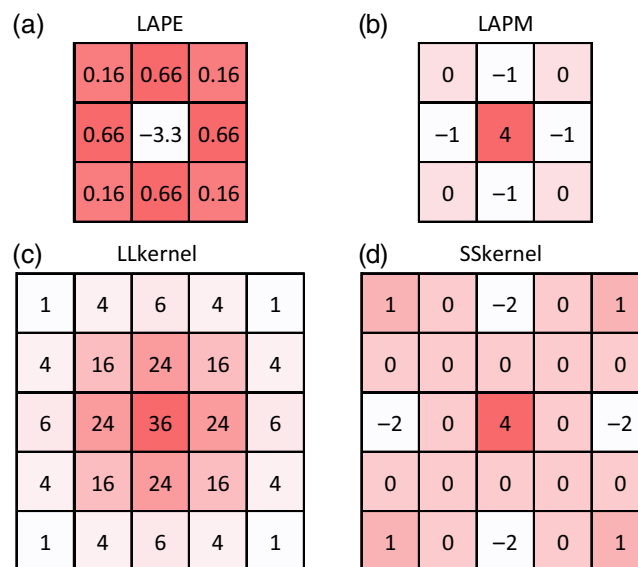


Fig. 5 The matrices used in the computation of the (a) LAPE and (b) LAPM BMs and the (c) LLkernel and (d) SSkernel TMs. These features share properties such as radial symmetry and often their zero sum nature.

Laplacian operator is intended to indicate the image sharpness at each pixel, the various Laws kernels were devised to detect gray level intensity, edges, spots, and ripples in horizontal and vertical directions in the image.³¹ In the Laws kernel TM abbreviations, the order of the capital letters indicates which feature is being detected in which direction, for example, the SRkernel feature is detecting spots in the vertical direction and ripples in the horizontal direction.

In view of the connections between some of the BMs and TMs, it was surprising to find that none of the Laplacian BMs as well as none of the Laws kernel TMs was able to separate the six blur levels when the textural groups were combined. Upon an initial analysis, it may seem more surprising that a few of the co-occurrence matrix and run length matrix TMs were able to separate the images based on the six blur levels when the textural groups were combined, while these measures were not significantly represented in the reduced linear models. However, it is precisely the fact that these TMs did not contribute to the reduced linear models that their ability to separate images will based on blur level is not surprising as the measures robust to reduction could not separate the images based on blur level. The run length matrix TMs provide information on the length of runs of constant gray level in an image; the expectation is that different textures will produce different-length runs of gray levels. While our model of blur is limited in that it does not re-introduce additive noise from the imaging system as would occur in the real-world scenario, the re-introduction of this noise would act uniformly on all measures and therefore would not significantly alter the mechanism by which these features are performing. For example, with or without additive noise, the blurring of an image would certainly cause an increase in the uniformity of gray levels in a neighborhood of the image and so increase the lengths of gray level runs.

5 Conclusion

In this work, we have shown with the use of linear models that texture information can affect the output of BMs. This was confirmed on three separate image sets and confirms our hypothesis that BMs can be influenced by texture in an image. To probe a more universal relationship of TMs to BMs independent of the image data used to create the linear models, a new set of reduced linear models were created using only TMs that were significantly non-zero across all three image data sets for each linear model. Of the linear models that were considered robust to this reduction of variables, there were clear structural similarities between the BMs and TMs that modeled their response. Applying small levels of blur (0.2 to 1 mm) to our CLB image data set and attempting to separate the images on the six levels of blur using both BMs and TMs revealed that none of the BMs were able to separate the six levels of blur; however, a population of TMs was able to. Further, these TMs were not large contributors to the reduced linear models. If we wished to perform classification of blur level in our CLB images, then according to our results some combination of run length matrix TMs would likely provide the best feature set for this task, not the suite of BMs in published literature. This suggests that conventional BMs may not be the optimal tool for blur classification in mammogram images.

These results should encourage investigators to spend time to explore and analyze optimal features for their dataset and application. This is especially important in the age of open source tools where it is tempting to use these tools at face value however, as we have shown, the solution could produce suboptimal or potentially misleading results and should be validated for each individual application. Simply because a calculation on an image is called a BM does not mean it will perform best in blur detection tasks, nor does it mean it will detect only blur. In reality, BMs and TMs are classified as such because of the use case for which they were created. This does not mean that they will not perform well in detection tasks other than the ones they were created for as we have shown. These results should encourage investigators to challenge the classic definitions of these texture, blur, and other measures and to apply measures as their label suggests without further validation.

Disclosure

No conflicts of interest, financial or otherwise, are declared by the authors.

References

1. R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology* **278**, 563–577 (2016).
2. L. S. Chow and R. Paramesran, "Review of medical image quality assessment," *Biomed. Signal Process. Control* **27**, 145–154 (2016).
3. N. Kamona and M. Loew, "Automatic detection of simulated motion blur in mammograms," *Med. Phys.* **47**, 1786–1795 (2020).
4. H. Luo et al., "Motion blur detection in radiographs," *Proc. SPIE* **6914**, 275–282 (2008).
5. S. K. Nayar, *Shape From Focus*, Carnegie-Mellon University Robotics Inst., Pittsburgh, Pennsylvania (1989).
6. N. Persch et al., "Physically inspired depth-from-defocus," *Image Vis. Comput.* **57**, 114–129 (2017).
7. S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.* **46**(5), 1415–1432 (2013).
8. M. Subbarao and J.-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 864–870 (1998).
9. R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.* **18**, 717–728 (2009).
10. W. K. Ma et al., "Analysis of motion during the breast clamping phase of mammography," *Br. J. Radiol.* **89**, 20150715 (2016).
11. M. K. Ma et al., "What is the minimum amount of simulated breast movement required for visual detection of blurring? An exploratory investigation," *Br. J. Radiol.* **88**, 20150126 (2015).
12. P. Henrot et al., "Breast microcalcifications: the lesions in anatomical pathology," *Diagn. Interv. Imaging* **95**, 141–152 (2014).
13. S. O'Grady and M. P. Morgan, "Microcalcifications in breast cancer: from pathophysiology to diagnosis and prognosis," *Biochim. Biophys. Acta-Rev. Cancer* **1869**, 310–320 (2018).
14. A. K. Abdullah et al., "The impact of simulated motion blur on lesion detection performance in full-field digital mammography," *Br. J. Radiol.* **90**, 20160871 (2017).
15. J. Kautsky et al., "A new wavelet-based measure of image focus," *Pattern Recognit. Lett.* **23**, 1785–1794 (2002).
16. C.-H. Shen and H. H. Chen, "Robust focus measure for low-contrast images," in *Digest of Technical Papers Int. Conf. Consum. Electron.*, IEEE, pp. 69–70 (2006).
17. R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall (2002).
18. M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, Cengage Learning (2014).
19. B. Julesz, "Textons, the elements of texture perception, and their interactions," *Nature* **290**, 91–97 (1981).
20. H. Horng et al., "Improved generalized ComBat methods for harmonization of radiomic features," *Sci. Rep.* **12**, 19009 (2022).
21. R. N. Mahon et al., "ComBat harmonization for radiomic features in independent phantom and lung cancer patient computed tomography datasets," *Phys. Med. Biol.* **65**, 015010 (2020).
22. J. S. Duncan, M. F. Insana, and N. Ayache, "Biomedical imaging and analysis in the age of big data and deep learning," *Proc. IEEE* **108**, 3–10 (2020).
23. F. O. Bochud, C. K. Abbey, and M. P. Eckstein, "Statistical texture synthesis of mammographic images with clustered lumpy backgrounds," *Opt. Express* **4**, 33–43 (1999).
24. I. C. Moreira et al., "Inbreast: toward a full-field digital mammographic database," *Acad. Radiol.* **19**, 236–248 (2012).
25. The University of Arizona, "Image Quality Toolbox," (2001). <https://wp.optics.arizona.edu/cgri/objectives/image-quality-toolbox/> (accessed 20 June 2019).
26. P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, by Phil Brodatz. Dover Publications (1966).
27. S. Pertuz, "Shape from focus - file exchange - MATLAB central," (2019). https://www.mathworks.com/matlabcentral/fileexchange/55103-shape-from-focus?s_tid=prof_contriblnk (accessed 7 October 2019).

28. D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Can. J. Remote Sens.* **28**, 45–62 (2002).
29. E. A. Gibson et al., "Multiphoton microscopy for ophthalmic imaging," *J. Ophthalmol.* **2011**, 1–11 (2011).
30. R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **SMC-3**(6), 610–621 (1973).
31. K. I. Laws, "Textured image segmentation," University of Southern California Los Angeles Image Processing INST (1980).
32. 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**, 780–795 (1999).
33. G. Thibault et al., "Shape and texture indexes application to cell nuclei classification," *Int. J. Pattern Recognit. Artif. Intell.* **27**, 1357002 (2013).
34. U. Ali and M. Mahmood, "Analysis of blur measure operators for single image blur segmentation," *Appl. Sci.* **8**, 807 (2018).
35. A. Depeursinge and J. Fageot, "Biomedical texture operators and aggregation functions: a methodological review and user's guide," in *Biomedical Texture Analysis*, A. Depeursinge et al., Eds., pp. 55–94, Elsevier (2017).
36. M. Moeller et al., "Variational depth from focus reconstruction," *IEEE Trans. Image Process.* **24**, 5369–5378 (2015).
37. M. Rusiñol, J. Chazalon, and J.-M. Ogier, "Combining focus measure operators to predict OCR accuracy in mobile-captured document images," in *11th IAPR Int. Workshop on Document Anal. Syst.*, IEEE, pp. 181–185 (2014).
38. A. S. Setiawan et al., "Mammogram classification using Law's texture energy measure and neural networks," *Proc. Comput. Sci.* **59**, 92–97 (2015).

Andrew William Chen is currently a postdoctoral researcher in the Department of Radiology at the University of Pennsylvania, Philadelphia, Pennsylvania (PA), USA. He received his PhD in biomedical engineering from George Washington University, Washington D.C., in 2021. He received his BS degree with majors in mathematics and physics from Dickinson College, Carlisle, PA, in 2016. His research emphasizes medical image analysis, radiomics (with a focus on lung cancer outcomes prediction), texture measures, and quantitative ultrasound.

Murray H. Loew is a professor and chair of the Department of Biomedical Engineering at George Washington University, Washington, D.C. His research emphasizes medical image analysis, hyperspectral imaging, and machine learning. His applications include image registration and fusion (for medicine and art), and image interpretation and diagnosis (for medicine and for conservation in cultural heritage). He is a fellow of IEEE, SPIE, and AIMBE and was the inaugural Fulbright Distinguished Chair in Advanced Science and Technology in 2014.