

RESEARCH ARTICLE

Comparison of methods for texture analysis of QUS parametric images in the characterization of breast lesions

Laurentius O. Osapoetra^{1,2,3,4}, William Chan⁵, William Tran^{1,2,6}, Michael C. Kolios⁷, Gregory J. Czarnota^{1,2,3,4,7*}

1 Department of Radiation Oncology, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada, **2** Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada, **3** Physical Sciences, Sunnybrook Research Institute, Toronto, Ontario, Canada, **4** Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, **5** University of Waterloo, Toronto, Ontario, Canada, **6** Evaluative Clinical Sciences, Sunnybrook Research Institute, Toronto, Ontario, Canada, **7** Department of Physics, Ryerson University, Toronto, Ontario, Canada

* Gregory.czarnota@sunnybrook.ca



OPEN ACCESS

Citation: Osapoetra LO, Chan W, Tran W, Kolios MC, Czarnota GJ (2020) Comparison of methods for texture analysis of QUS parametric images in the characterization of breast lesions. *PLoS ONE* 15(12): e0244965. <https://doi.org/10.1371/journal.pone.0244965>

Editor: Guy Cloutier, University of Montreal, CANADA

Received: May 11, 2020

Accepted: December 18, 2020

Published: December 31, 2020

Copyright: © 2020 Osapoetra et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available upon request according to Sunnybrook Research Institute institutional policy and process. Data requests can be sent to cancer.research@sunnybrook.ca. The Data are from "Pilot Investigation of Ultrasound Imaging and Spectroscopy for Characterizing Breast Masses", REB# 232-2014.

Funding: This research was supported in part by the Collaborative Health Research Projects (CHRP 538814-19) initiative between the Natural Sciences

Abstract

Purpose

Accurate and timely diagnosis of breast carcinoma is very crucial because of its high incidence and high morbidity. Screening can improve overall prognosis by detecting the disease early. Biopsy remains as the gold standard for pathological confirmation of malignancy and tumour grading. The development of diagnostic imaging techniques as an alternative for the rapid and accurate characterization of breast masses is necessitated. Quantitative ultrasound (QUS) spectroscopy is a modality well suited for this purpose. This study was carried out to evaluate different texture analysis methods applied on QUS spectral parametric images for the characterization of breast lesions.

Methods

Parametric images of mid-band-fit (MBF), spectral-slope (SS), spectral-intercept (SI), average scatterer diameter (ASD), and average acoustic concentration (AAC) were determined using QUS spectroscopy from 193 patients with breast lesions. Texture methods were used to quantify heterogeneities of the parametric images. Three statistical-based approaches for texture analysis that include Gray Level Co-occurrence Matrix (GLCM), Gray Level Run-length Matrix (GRLM), and Gray Level Size Zone Matrix (GLSZM) methods were evaluated. QUS and texture-parameters were determined from both tumour core and a 5-mm tumour margin and were used in comparison to histopathological analysis in order to classify breast lesions as either benign or malignant. We developed a diagnostic model using different classification algorithms including linear discriminant analysis (LDA), *k*-nearest neighbours (KNN), support vector machine with radial basis function kernel (SVM-RBF), and an artificial neural network (ANN). Model performance was evaluated using leave-one-out cross-validation (LOOCV) and hold-out validation.

and Engineering Research Council (NSERC) of Canada and the Canadian Institutes of Health Research (CIHR), the Terry Fox Foundation, and the Hecht Foundation.

Competing interests: The authors have declared that no competing interests exist.

Results

Classifier performances ranged from 73% to 91% in terms of accuracy dependent on tumour margin inclusion and classifier methodology. Utilizing information from tumour core alone, the ANN achieved the best classification performance of 93% sensitivity, 88% specificity, 91% accuracy, 0.95 AUC using QUS parameters and their GLSZM texture features.

Conclusions

A QUS-based framework and texture analysis methods enabled classification of breast lesions with >90% accuracy. The results suggest that optimizing method for extracting discriminative textural features from QUS spectral parametric images can improve classification performance. Evaluation of the proposed technique on a larger cohort of patients with proper validation technique demonstrated the robustness and generalization of the approach.

Introduction

Breast cancer demonstrates a high incidence and leads to high morbidity in women [1,2]. In 2017, there were 250,520 newly diagnosed cases of female breast cancer and 42,000 death in the United States [3]. Early detection of breast carcinoma through screening can enhance prognosis as appropriate treatments are provided to patients at an earlier stage of the disease [2]. For this purpose, accurate and precise diagnostic techniques are required.

Breast cancer diagnosis is based on clinical examination, imaging findings, and confirmed by histopathological results [2]. The current imaging workflow for breast cancer diagnosis begins with x-ray mammography, followed by standard ultrasound imaging (B-mode US imaging), dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) as needed, followed by core-needle biopsy, as required [2]. Mammography is susceptible to both providing false positive readings and concealing an underlying malignancy because of superimposition of normal breast parenchyma [4]. Biopsy remains as the gold standard for pathological confirmation of malignancy and tumour grade characterization [2]. However, as biopsies are invasive in nature, they are associated with pain and a hypothetical increased risk of tumour cell migration [5]. Furthermore, the low specificity of B-mode US images resulted in a trend of increasingly performing unnecessary biopsies [4,6]. The specificity of breast cancer detection may be increased using Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) [7]. However, DCE-MRI is not always available for rapid diagnosis purposes. The development of imaging techniques that can perform rapid and accurate characterization of breast lesions is highly beneficial for early detection of breast carcinoma and triaging patients in a screening workflow [8,9].

Previously, sonographic characteristics of solid breast nodules have been used in the characterization of breast lesions [10]. In addition, morphologic features of the breast tumours have been utilized for developing computer aided diagnosis (CAD) systems for the characterization of breast lesions using artificial neural network (ANN) [11,12]. Recently, deep learning approaches have also been applied for breast mass classification in sonography [13,14]. As these studies used B-mode US images, which are instrument- and operator-dependent, the sonographic features and other quantitative features derived from them are influenced by acquisition system settings. QUS spectroscopy may address these limitations. QUS

spectroscopy estimate spectral-based parameters through analysis of raw radiofrequency (RF) signal and utilize normalization procedure to remove instrument-dependent effects [15,16]. Furthermore, attenuations because of propagation through intervening tissue layers and the tumour are also compensated prior to estimation of tumour scattering parameters. This results in the attenuation-corrected normalized power spectrum (NPS) or the backscatter coefficient (BSC) [16]. Linear parametrization of the attenuation-corrected NPS results in QUS spectral parameters including mid-band-fit (MBF), spectral-slope (SS), and 0-MHz intercept (SI) [17,18]. These parameters are linked to the scattering power, size, and shape of acoustic scatterers [19,20]. Furthermore, fitting of theoretical acoustic scattering models to the measured BSC allows for estimation of scatterer property estimates: average scatterer diameter (ASD) and average acoustic concentration (AAC) [20–22].

The utilities of QUS spectroscopy have been demonstrated in the assessment of tumour responses to cancer therapies both pre-clinically and clinically [23–27], characterization of different types of tissues such as prostate, liver, and retina [28–33], determination of blood-clot and various intravascular plaque components [34–36], and detection of tumour deposits in *ex vivo* lymph nodes [37]. Spontaneously occurring mammary fibroadenomas (benign lesions) and mammary carcinomas (malignant lesions) have also been differentiated using QUS techniques pre-clinically [21]. In addition, the methods have also been utilized to differentiate different types of mammary cancers including carcinoma and sarcoma [22]. Furthermore, QUS spectroscopy have also been used in clinical research to differentiate breast tumours from the surrounding normal tissues in patients with locally-advanced breast cancer (LABC) [38]. Recently, QUS spectral parametric imaging, along with texture and novel derivative texture analysis, have been used in the characterization of breast lesions [8,9].

Tumour micro-environment, physiology and metabolism exhibit spatial heterogeneities that offer diagnostic and prognostic values [39–43]. These have been demonstrated using different imaging modalities, such as MRI [44], positron emission tomography (PET) [45,46], computerized tomography (CT) [47,48], and diffuse optical spectroscopy (DOS) [26]. Texture analysis methods can quantify such heterogeneities [49]. Texture analysis using GLCM techniques has been applied to B-mode US images for breast lesions characterization, as benign and malignant lesions often demonstrate homogeneous and heterogeneous textures, respectively [50–54]. However, as these images are system- and operator-dependent, the quantitative texture measures do not represent independent intrinsic properties of a tumour. Application of texture analysis on QUS spectral parametric images alleviates this limitation, providing texture parameters that represent intrinsic tumour characteristics.

In earlier studies, we only used the GLCM method to analyze texture of QUS spectral parametric images [8,9]. In this study, different texture methods were applied to QUS spectral parametric images encompassing the breast mass and its 5-mm margin. As there are diverse approaches for analyzing texture, here we evaluated three statistical-based texture analysis methods that have been commonly used in the literature [55]. These include the Gray Level Co-occurrence Matrix (GLCM) [49], the Gray Level Run Length Matrix (GRLM) [56–60], and the Gray Level Size Zone Matrix (GLSZM) [61] methods. GLCM methodology quantifies texture using second-order statistics of gray scale image histograms [49,53–55]. The GRLM method characterizes texture images based on the run-length of image gray levels [56–60], whereas the GLSZM method measures the size of homogeneous zones for each gray level in an image [61]. These approaches were applied here on a larger cohort of 193 patients with breast lesions. QUS-based texture analysis of tumour margins has been demonstrated in the *a priori* prediction of response and survival in LABC patients undergoing neoadjuvant chemotherapy (NAC) [62]. Margin information is further potentially useful for characterizing breast lesions

as has been shown recently using QUS spectroscopy [9], US Nakagami shape parameters and texture features of B-mode US images [63].

The study here developed a diagnostic model to classify breast lesions as either benign or malignant. Specifically, our work used the different texture methods described above along with standard and advanced classification algorithms that include linear discriminant analysis (LDA), *k*-nearest neighbours (KNN), support vector machine with radial basis function kernel (SVM-RBF), and a shallow artificial neural network (ANN). The performance of the diagnostic model using standard classification algorithms was evaluated using leave-one-out cross-validation (LOOCV) and split-sample/hold-out validation. Evaluation of the proposed approach on the independent hold-out testing set demonstrates the generalization of our model. On the other hand, ANN implementation partitioned the data into training, validation, and testing subsets and evaluated performance of the trained network on the hold-out testing set. Classification performance was assessed using the receiver operating characteristics (ROC) analysis to obtain metrics of sensitivity, specificity accuracy, AUC, positive predictive value (PPV), and negative predictive value (NPV). The ground truth about the nature of lesions as either benign or malignant was obtained from clinical patient reports comprising of MR images and biopsy results, for the biopsied lesions. The results suggest that QUS spectral parametric imaging, along with optimized texture analysis methods, is a potential imaging modality for the rapid, accurate, and non-invasive characterization of breast lesions.

Methods

Study protocol & data acquisition

This study was conducted based on institutional-research-ethics board approval (Sunnybrook Health Sciences Center). US RF Data were acquired from 193 patients (benign and malignant) with breast lesions at the Rapid Diagnostic Unit (RDU) of the Louise Temerty Breast Cancer Center at Sunnybrook Health Sciences Center, Toronto, Ontario, Canada upon obtaining written informed consent. Data acquisition was performed by an experienced sonographer using a Sonix Touch US system (Ultrasonix, Vancouver, Canada). The system was equipped with a linear array transducer (L14-5/60W) that operates at 6.5 MHz center frequency and 3–8 MHz bandwidth. Beam-formed RF data were digitized using 40 MHz sampling frequency. Data acquisition was performed along 512 scan lines, spanning a 6cm lateral field-of-view (FOV) and a 4cm depth, obtained using a high line density option. This feature allows acquisition of beamformed A-lines from 512 transmit-receive apertures through application of electronic time delay. The focal depth was set at the center of the tumour. US images were acquired at approximately 5-mm intervals across the tumour volume via hand translation.

A radiologist with experience in interpreting breast US images performed contouring of the tumour regions of interest (ROI) on B-mode US images. QUS spectroscopy and texture analyses were performed on selected ROIs covering tumour core and a 5-mm tumour margin. The margin is an extension of the tumour from the core up to a 5-mm maximum distance into the surrounding area (peri-tumoural region). A 5-mm margin was chosen as it previously provided the best characterization results in breast cancer patients in other QUS applications [62].

The inclusion criterion of this study is sonographically identified breast lesions after the masses has been identified on clinical examination, in combination with imaging findings. The ground truth lesions identification as either benign or malignant was obtained from clinical reports that include results from MR images and biopsy, for lesions that underwent biopsy. Patients were excluded if the lesion was not able to be identified during the US scan. The goal of this study is to demonstrate that QUS spectroscopy and different texture methods can extract imaging biomarkers that are distinct between benign and malignant breast lesions.

Feature extraction: Linear regression & acoustic form-factor parameters

QUS spectral parametric images were created using a sliding window technique with a 2-mm by 2-mm kernel and a 94% window overlap between adjacent kernels in the axial and lateral directions. The kernel size was chosen to include enough number of acoustic wavelengths for reliable spectral estimation, while preserving image texture. At 6.5MHz center frequency, the kernel includes 8 wavelengths axially and 17 scan-lines laterally.

Individual RF scan lines within the window were gated along the beam direction using a Hanning function for spectral analysis. We used the Fast Fourier Transform (FFT) algorithm to extract the power spectrum of the sample. Several independent adjacent RF signals within the window were used to obtain an averaged power spectrum that better represents the true power spectrum of the sample. Normalization procedure was performed using the reference phantom technique to remove instrument-dependent effects and to account for transmission path factors [15,16,20]. The reference phantom was composed of 5–30 μm glass beads embedded in a homogeneous medium of oil droplets that were immersed in gelatin. The measured attenuation coefficient and speed of sound of the phantom were 0.786dB/cm/MHz and 1,540m/s, respectively (University of Wisconsin, Department of Medical Physics, Madison, WI, USA). Prior to estimating spectral parameters, attenuation correction was performed. We assumed an attenuation coefficient of 1dB/cm/MHz for the intervening breast tissues [64,65] and estimated the local attenuation coefficient of the tumour (ACE) using a spectral difference method [66]. It estimates the rate of change in the log-transformed spectral power magnitude over depth in the ROI (over the tumour region) relative to the reference phantom for each frequency within the analysis bandwidth [66]. The power of the frequency dependence of the attenuations was assumed to be linear over the analysis bandwidth [66]. Our choice of 2-mm by 2-mm kernel size and typical ROI lengths greater than 35λ satisfy the requirement for optimal attenuation estimation, prescribed by Labyed *et al.* [66]. Specifically, Labyed *et al.* concluded that window sizes greater than 5λ and ROI sizes greater than 35λ resulted in the mean and STD errors of the ACEs that are less than 15% and 10%, respectively [66]. The measured BSC from the sample $\sigma_m(f)$ was calculated using [16,20]

$$\sigma_m(f) = \sigma_r(f) \frac{|S_m(f)|^2}{|S_r(f)|^2} e^{\{4(\alpha_m - \alpha_r)(R + \frac{\Delta z}{2})\}} \propto NPS(f), \tag{1}$$

where $\sigma_r(f)$ is the BSC of the reference phantom, $S_m(f)$ and $S_r(f)$ are the RF spectra from the sample and the reference phantom, respectively. Parameters α_m and α_r are the attenuation functions from the sample and the reference phantom, respectively. Parameter R is the distance from the transducer face to the proximal side of the ROI window, and Δz is the kernel length. The MBF, SS, and SI parameters were obtained from linear regression analysis of the attenuation-corrected NPS. Subsequently, using more complex acoustic scattering models of soft tissues, acoustic scattering parameters can be obtained. We fitted theoretical BSC from spherical Gaussian acoustic form factor model $\sigma_{theor}(f)$ to the measured BSC to obtain average scatterer diameter (ASD) a_{eff} and average acoustic concentration (AAC) n_z parameters [20,67]. The AAC represented the net scattering strength [19,21,22,67]. It is defined as the product of average number density of scatterers $-n$ and squared of the fractional difference in the acoustic impedance between the scatterer and surrounding tissue γ_0^2 [19,21,22,67]. The theoretical BSC is given as [20,67]

$$\sigma_{theor}(f) = Cf^4 a_{eff}^6 n_z F(f, a_{eff}), \tag{2}$$

where $C = \frac{\pi^2}{36c_l^4}$ and c_l is the speed of sound. $F(f, a_{eff})$ is the form factor that captures the

frequency dependence of the scattering. These analyses resulted in parametric images of MBF, SS, SI, ASD, and AAC. From these images, mean-value parameters were obtained from tumour core and tumour margin. Core and margin analyses were reflected in core-to-margin ratio (CMR) and core-to-margin-contrast ratio (CMCR) metrics in order to compare pixel intensities between the two regions:

$$CMR = \frac{\text{mean}(ROI_{Core})}{\text{std}(ROI_{Margin})} \quad (3)$$

$$CMCR = \frac{|\text{mean}(ROI_{Core}) - \text{mean}(ROI_{Margin})|}{\frac{1}{2}(\text{std}(ROI_{Core}) + \text{std}(ROI_{Margin}))} \quad (4)$$

[9]. Mean-value parameters, CMR, and CMCR parameters of each parametric image were subsequently used as potential features for classification.

Feature extraction: Texture analysis methods

GLCM (Gray Level Co-occurrence Matrix) method. The GLCM method realizes second-order statistical analysis by studying the spatial relationship between neighbouring pixels in an image [49]. The full range of gray levels in each parametric image was linearly scaled into 16 discrete gray levels. We evaluated symmetric GLCM matrices from each parametric image at inter-pixel distances: 1, 2, 3, 4, 5 pixels and at four angular directions: 0°, 45°, 90°, and 135°. From these GLCM matrices, we extracted GLCM features that include:

$$Contrast = \sum_{i,j=0}^{N_g} |i - j|^2 p(i, j) \quad (5)$$

$$Correlation = \frac{1}{\sigma_i \sigma_j} \sum_{i,j=0}^{N_g} (i - \mu_i)(j - \mu_j) p(i, j) \quad (6)$$

$$Energy = \sum_{i,j=0}^{N_g} p^2(i, j) \quad (7)$$

$$Homogeneity = \sum_{i,j=0}^{N_g} \frac{p(i, j)}{1 + |i - j|} \quad (8)$$

In Eqs 5–8, $p(i, j)$ is the gray level matrix element that represents the probability of having neighbouring pixels with intensities i and j in the image. N_g denotes the number of gray levels, while μ and σ are the mean and standard deviation for row i or column j of the GLCM matrix. Textural features are subsequently averaged over distances and angular directions. Textural measures were assumed to be reflected in these averaged values [49]. Contrast quantifies local gray level variations in the parametric image. Smoother image results in lower contrast, while coarser image produces higher contrast. Correlation represents linear correlation between neighbouring pixels. Energy measures textural uniformity between neighbouring pixels, while homogeneity quantifies the incidence of pixel pairs of different intensities [8].

GRLM (Gray Level Run-length Matrix) method. The GRLM method characterizes the coarseness of texture based on run-length of image gray levels [56–60]. A gray level run is a set of consecutive, collinear pixels having the same gray level i in a prescribed direction θ (flat zone) [56–60]. For a given image, the size of the run-length matrix is the number of gray levels N_G by the number of run-length N_R . A run-length matrix element $p_{RL}(i, j|\theta)$ is defined as the number of runs of j pixels with gray level intensities i in the direction θ . The GRLM method

was applied on each parametric image. Each parametric image was quantized into 16 discrete gray levels prior to texture estimation. Subsequently, run-length matrices were evaluated for directions $\theta = 0^\circ, 45^\circ, 90^\circ,$ and 135° . From each run-length matrix, we can extract run-length features [56–60]. In the following, let the total number of run-length in the image $s = \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} p_{RL}(i, j|\theta)$, while $\mu_i = \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} p_{RL}(i, j|\theta)i$, $\mu_j = \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} p_{RL}(i, j|\theta)j$, and $r(j|\theta) = \sum_{i=1}^{N_G} p_{RL}(i, j|\theta)$.

Short Run Emphasis (SRE):

$$SRE = \frac{1}{s} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} \frac{p_{RL}(i, j|\theta)}{j^2} = \frac{1}{s} \sum_{j=1}^{N_R} \frac{r(j|\theta)}{j^2} \tag{9}$$

Long Run Emphasis (LRE):

$$LRE = \frac{1}{s} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} j^2 p_{RL}(i, j|\theta) = \frac{1}{s} \sum_{j=1}^{N_R} j^2 r(j|\theta) \tag{10}$$

Gray Level Nonuniformity (GLN):

$$GLN = \frac{1}{s} \sum_{i=1}^{N_G} \left(\sum_{j=1}^{N_R} p_{RL}(i, j|\theta) \right)^2 \tag{11}$$

Run Length Nonuniformity (RLN):

$$RLN = \frac{1}{s} \sum_{j=1}^{N_R} \left(\sum_{i=1}^{N_G} p_{RL}(i, j|\theta) \right)^2 = \frac{1}{s} \sum_{j=1}^{N_R} r^2(j|\theta) \tag{12}$$

Run Percentage (RP):

$$RP = \frac{1}{N} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} p_{RL}(i, j|\theta) = \frac{1}{N} \sum_{j=1}^{N_R} r(j|\theta) \tag{13}$$

Low Gray Level Run Emphasis (LGRE):

$$LGRE = \frac{1}{s} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} \frac{p_{RL}(i, j|\theta)}{i^2} \tag{14}$$

High Gray Level Run Emphasis (HGRE):

$$HGRE = \frac{1}{s} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} i^2 p_{RL}(i, j|\theta) \tag{15}$$

Short Run Low Gray Level Emphasis (SRLGE):

$$SRLGE = \frac{1}{s} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} \frac{p_{RL}(i, j|\theta)}{i^2 j^2} \tag{16}$$

Short Run High Gray Level Emphasis (SRHGE):

$$SRHGE = \frac{1}{s} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} \frac{i^2 p_{RL}(i, j|\theta)}{j^2} \tag{17}$$

Long Run Low Gray Level Emphasis (LRLGE):

$$LRLGE = \frac{1}{s} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} \frac{j^2 p_{RL}(i, j|\theta)}{i^2} \tag{18}$$

Long Run High Gray Level Emphasis (LRHGE):

$$LRHGE = \frac{1}{s} \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} i^2 j^2 p_{RL}(i, j | \theta) \quad (19)$$

Gray Level Variance (GV):

$$GV = \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} p_{RL}(i, j | \theta) (i - \mu_i)^2 \quad (20)$$

Run-length Variance (RV):

$$RV = \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} p_{RL}(i, j | \theta) (j - \mu_j)^2 \quad (21)$$

Run-length Entropy (RE):

$$RE = - \sum_{i=1}^{N_G} \sum_{j=1}^{N_R} p_{RL}(i, j | \theta) \log_2(p_{RL}(i, j | \theta)) \quad (22)$$

Texture measures were subsequently averaged over directions. SRE quantifies the distribution of short run-lengths, with greater value indicating the presence of shorter run-lengths in the parametric map which characterizes finer textures [60]. On the other hand, LRE measures the distribution of longer run-lengths, with greater values indicating the presence of longer run-lengths in the parametric image which represents coarser structural textures [60]. RLN assesses the similarity of run-lengths in the parametric image, with a lower value indicating more homogeneity among run-lengths in the image. RP Measures the coarseness of textures by taking the ratio of the number of runs with the total number of pixels in the image. SRE, LRE, RLN, and RP are typical features of run-length statistics [56,58]. However, these run-length features are defined by $r(j|\theta)$ the total number of runs of j pixels for all possible gray levels i , in the prescribed direction θ . Since for a given value of $r(j|\theta)$, the composition of runs can vary for different gray levels, features that depend solely on $r(j|\theta)$ would not be able to detect variation in gray levels [58]. In order to overcome this, LGRE and HGRE features were introduced [58]. LGRE and HGRE make use of the distribution of gray levels of runs [58]. LGRE measures the distribution of pixels with lower gray levels, with a greater value indicating a greater concentration of low gray levels in the parametric image. On the other hand, HGRE quantifies the distribution of pixels with higher gray levels, with a greater value indicating a greater concentration of high gray levels in the image. In a later study, features that measure joint distribution of run-length and gray levels were also introduced [59]. These include SRLGE, SRHGE, LRLGE, and LRHGE. SRLGE measures the joint distribution of shorter run-length with low gray levels. SRHGE quantifies the joint distribution of shorter run-length with high gray levels. LRLGE measures the joint distribution of longer run-length with low gray levels. LRHGE quantifies the joint distribution of longer run-length with high gray levels.

GV measures the variance of gray levels in the runs. RV is a measure of the variance in run for run-length. Run entropy (RE) quantifies the randomness in the distribution of run-lengths and gray levels. A higher value of RE indicates more texture randomness in the image.

GLSZM (Gray Level Size Zone Matrix) method. The GLSZM method quantifies texture by measuring the size of homogeneous zones for each gray level in an image [60,61]. A gray level zone is defined as an area of connected pixels with the same gray level. In the GLSZM matrix, $p_{SZ}(i, j)$ represents the number of gray level zones with gray level i and size j appearing in the image. In contrast to GLCM and GRLM methods, the GLSZM technique is direction independent. The computation of GLSZM matrix is based on run-length matrix calculation. From the GLSZM matrix, zone size features can be determined [60,61]. In the following

equations, let $\mu_i = \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} p_{sz}(i, j)i$, $\mu_j = \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} p_{sz}(i, j)j$, and N_S is a dynamic number that represents the size of the largest flat zone in the image.

Small Area Emphasis (SAE):

$$SAE = \frac{1}{N_Z} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} \frac{p_{sz}(i, j)}{j^2} \quad (23)$$

Large Area Emphasis (LAE):

$$LAE = \frac{1}{N_Z} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} j^2 p_{sz}(i, j) \quad (24)$$

Gray Level Nonuniformity (GLN):

$$GLN = \frac{1}{N_Z} \sum_{i=1}^{N_G} \left(\sum_{j=1}^{N_S} p_{sz}(i, j) \right)^2 \quad (25)$$

Size Zone Nonuniformity (SZN):

$$SZN = \frac{1}{N_Z} \sum_{j=1}^{N_S} \left(\sum_{i=1}^{N_G} p_{sz}(i, j) \right)^2 \quad (26)$$

Zone Percentage (ZP):

$$ZP = \frac{1}{N} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} p_{sz}(i, j) \quad (27)$$

Low Gray Level Zone Emphasis (LGLZE):

$$LGLZE = \frac{1}{N_Z} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} \frac{p_{sz}(i, j)}{i^2} \quad (28)$$

High Gray Level Zone Emphasis (HGLZE):

$$HGLZE = \frac{1}{N_Z} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} i^2 p_{sz}(i, j) \quad (29)$$

Small Area Low Gray Level Emphasis (SALGE):

$$SALGE = \frac{1}{N_Z} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} \frac{p_{sz}(i, j)}{i^2 j^2} \quad (30)$$

Small Area High Gray Level Emphasis (LAHGE):

$$SAHGE = \frac{1}{N_Z} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} \frac{i^2 p_{sz}(i, j)}{j^2} \quad (31)$$

Large Area Low Gray Level Emphasis (LALGE):

$$LALGE = \frac{1}{N_Z} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} \frac{j^2 p_{sz}(i, j)}{i^2} \quad (32)$$

Large Area High Gray Level Emphasis (LAHGE):

$$LAHGE = \frac{1}{N_Z} \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} i^2 j^2 p_{sz}(i, j) \quad (33)$$

Gray-level Variance (GLV):

$$GLV = \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} p_{SZ}(i, j) (i - \mu_i)^2 \quad (34)$$

Zone Variance (ZV):

$$ZV = \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} p_{SZ}(i, j) (j - \mu_j)^2 \quad (35)$$

Zone Entropy (ZE):

$$ZE = - \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} p_{SZ}(i, j) \log_2(p(i, j)), \quad (36)$$

where $N_z = \sum_{i=1}^{N_G} \sum_{j=1}^{N_S} p_{SZ}(i, j)$ is the total number of zones in the image.

SAE quantifies the distribution of small size zones. The greater value for SAE indicates that the image consists of more smaller size zones or finer textures. On the other hand, LAE measures the distribution of large area size zones. The greater value for LAE indicates an image with coarser textures. GLN quantifies the variability of gray level intensities in an image. A higher value for GLN indicates less homogeneity in the image. SZN quantifies the variability of size zones in the image with a higher value for SZN indicating less homogeneity in size zone areas. ZP quantifies the coarseness of the texture.

LGLZE measures the distribution of lower gray level size zones with higher values indicating a greater proportion of size zones distribution of low gray levels. HGLZE measures the distribution of higher gray level size zones. Higher values indicate a greater proportion of size zone distributions with high gray levels. SALGE estimates in the image the proportion of smaller size zones with lower gray levels. SAHGE measures the proportion of smaller size zones with higher gray levels in the image. LALGE measures the proportion of larger size zones with lower gray levels in an image whereas LAHGE estimates the proportion of larger size zones of higher gray levels in the image. Parameter GLV estimates the variance of gray levels for the zones whereas ZV measures the variance of size zones for the zones [60]. Parameter ZE assesses the randomness in the distribution of size zones and gray levels in the image. The higher values in ZE indicate more heterogeneity in the texture image.

Classification algorithms

Mean-value parameters and textural features derived from GLCM, GRLM, and GLSZM methods were determined from each scan plane and averaged over all scan planes based on the ROI size. For each feature, we performed statistical analysis using MATLAB (Mathworks, Natick, Mass., USA) to check for any statistically significant difference between benign and malignant groups. To determine which tests to use, a Shapiro-Wilk normality test was performed on each feature to decide if it followed a normal distribution [27]. An unpaired *t*-test was used for a normally distributed feature. Otherwise, a non-parametric Mann-Whitney U-Test (two-sided, 95% confidence) was utilized. For these tests, *p*-values correction was not performed. The purpose of the statistical tests was solely to demonstrate the presence of discriminating features available for subsequent feature selection. These tests can gauge the resulting model performance as classification model developed using discriminating features will in general perform better compared to that developed using less discriminating features.

Using the GLCM method, a total of 25 mean-value and texture features were available for classification using either core or margin information. For the combined core and margin information, there were a total of 60 mean-value, texture, along with CMR and CMCR image quality features available for classification. Using the GRLM and GLSZM methods, a total of

75 mean-value and texture features were available for classification using either core or margin information. Combining both core and margin information, a total of 160 features were available for classification. These include mean-value, texture, along with CMR and CMCR image quality features.

The classification model was developed using the best combination of 10 features maximally. This was chosen based on the 10% rule of thumb to prevent overfitting [68]. Feature selection was performed using a forward sequential-feature-selection (SFS) that adds the feature one at a time up to a combination of 10 features. In each step, classification performance was evaluated. The selected features are those that provide the highest F1-Score (the harmonic average of precision and sensitivity) on the training set. We evaluated model performance using both LOOCV and hold-out validation. Leave-one-out cross-validation trains the model using all observations except one [27]. The process is repeated until all observations are left out for testing at least once [27]. The left-out observations are subsequently used for testing the developed model. As there are 193 observations in our cohort, this allows us to implement hold-out/split-sample validation. Hold-out validation will avoid performance over-estimation typically present using LOOCV [68]. This is appropriate to demonstrate the generalizability of the model to unseen testing sets. Hold-out validation randomly splits the data set into 70% training and 30% test sets. Model development was performed on the training set, while performance was evaluated on the unseen testing set. To account for the random partitioning process, several realizations were evaluated. The classification performance was found through averaging the results over ten different realizations of the data.

Standard classification algorithms were utilized and implemented using a custom software in MATLAB. These included: LDA, KNN, and nonlinear classifier in SVM-RBF. In addition, we also implemented an ANN using neural network pattern recognition tool (nprtool) in MATLAB. The performance of these classification algorithms was assessed using the ROC analysis utilizing sensitivity, specificity, accuracy, AUC, PPV, and NPV metrics. Using probabilistic generative models, LDA can be described as estimating the posterior probability of assigning an input vector \mathbf{x} into one of the two classes by assuming that probability density function of each class is a Gaussian [69]. A KNN is an instance-based learning algorithm that predicts class association of a test point in the feature space based on the majority of the points neighbouring the test point and the distance between those points to the test point. The KNN classifier used $k = 1, 3, 5$ nearest neighbours. The SVM-RBF creates a model that maximizes the margin between the two classes and predicts class association of the testing data based on which side of the gap they fall on [70]. RBF kernel is used to map the input data into a higher-dimensional space where the data are supposed to have better distribution, prior to selecting an optimal separating hyperplane in this higher-dimensional feature space. The soft margin parameter C and free parameter γ are the parameters of the kernel. These parameters were optimized using a grid search method.

Implementation of ANN using nprtool randomly partitions the data into training, validation, and testing subsets with 70%, 15%, and 15% proportion in each subset. The process is also repeated ten times, and classification performances were averaged over different partition realizations. The developed network is a shallow two-layer feedforward network that consists of a single hidden layer and an output layer. The network uses a sigmoid transfer function in the hidden layer and a soft-max transfer function in the output layer. The number of neurons in the hidden layer was set to 20. Training of the network involves updating weight and bias values to optimize the network's performance. The default performance function for feedforward network is the mean square error: averaged square error between the network output and the target output. Training was implemented in batch mode using an algorithm that updates the weight and bias values according to Levenberg-Marquardt optimization (trainlm

function in MATLAB). The batch size was 193. In Levenberg-Marquardt algorithm, the Jacobian of the performance with respect to weight and bias variables was calculated using back-propagation algorithm. The default trainlm training parameters were used. These include: the maximum number of epochs to train = 1000, performance goal = 0, minimum gradient = $1e-7$, maximum validation failures = 6, initial adaptive value $\mu = 0.001$, μ decrease factor = 0.1, μ increase factor = 10, and maximum $\mu = 1000$. Validation sets were used to suspend training early, if the performance on validation errors fails to improve or remains the same for 6 times in a row. Testing sets were used to further check for the generalizability of the network.

Results

US RF data were acquired from 193 patients in this study. Patients were aged 20 to 89 and 92 patients had benign masses and 101 patients had malignant masses in the research group. Patient and breast mass characteristics are provided in S1 and S2 Tables. In addition, Breast Imaging Reporting and Data Systems (BI-RADS) distribution among the lesions is also presented in S3 Table. Fig 1 presents representative B-mode US and parametric images of ASD, AAC, MBF, SS, and SI from both benign and malignant groups. The benign lesions in this study were diagnosed as predominantly fibroadenomas ($n = 46$) and cysts/complicated cysts ($n = 21$). The malignant lesions were diagnosed as invasive ductal carcinoma (IDC) ($n = 80$) and invasive mammary carcinoma ($n = 7$), respectively. Mean-value parameters were determined as well as GLCM, GRLM, and GLSZM texture parameters, along with image quality features from these parametric images and were evaluated based on their performance as imaging biomarkers associated with discriminating between benign and malignant lesions. Sonographically, benign lesions demonstrated better defined borders and appeared overall less

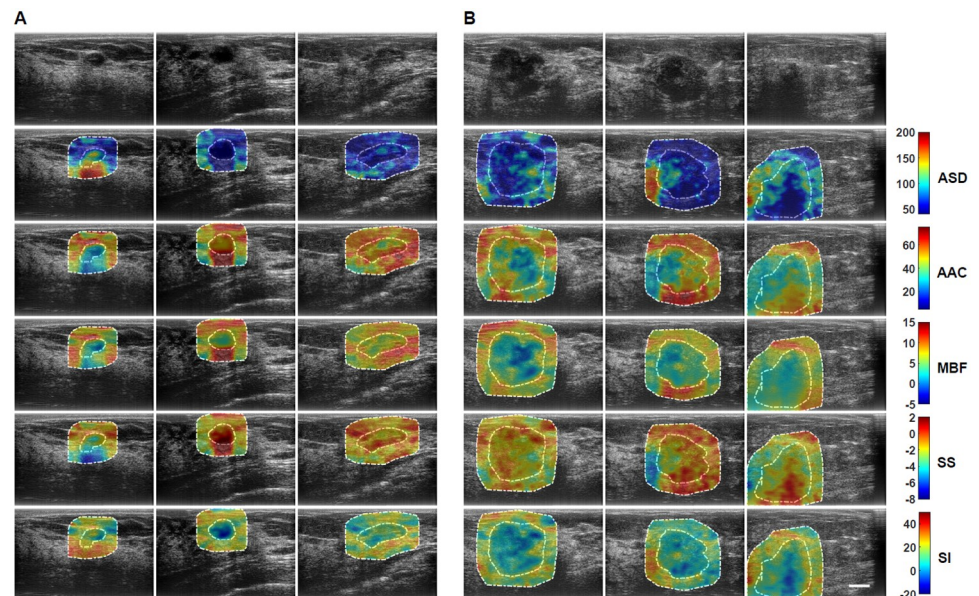


Fig 1. Representative B-mode and QUS spectral parametric images of ASD, AAC, MBF, SS, and SI from A benign (left three columns) and B malignant (right three columns) breast lesions. The colour-bar range is $160 \mu\text{m}$ for ASD, $70 \text{ dB}/\text{cm}^3$ for AAC, 20 dB for MBF, $10 \text{ dB}/\text{MHz}$ for SS, and 70 dB for SI. The scale bar represents 1 cm . The benign breast lesions were diagnosed as fibroadenomas, and complicated a cyst, respectively. The malignant lesions were diagnosed as invasive ductal carcinomas (IDC), invasive mammary carcinoma, and invasive lobular carcinoma (ILC). Using these parametric images, mean-value, textural, and image quality features were determined as imaging biomarkers for the characterization of breast lesions.

<https://doi.org/10.1371/journal.pone.0244965.g001>

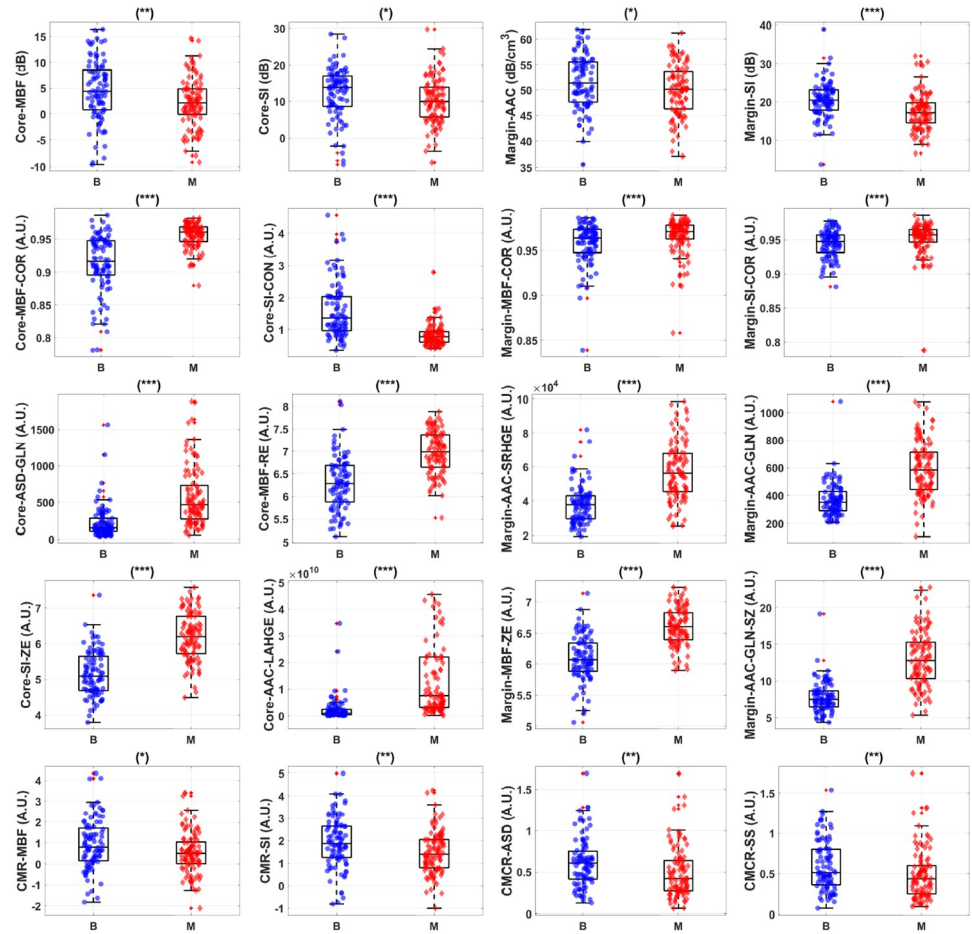


Fig 2. Representative box and scatter plots of features that demonstrate statistically significant difference (p -values < 0.05) between benign ('B') and malignant ('M') lesion groups. The first row shows core and margin mean-value parameters. The second row depicts representative core and margin GLCM features that showed discriminative power. The third row shows representative core and margin GRLM features that discriminate the two lesion groups. The last row depicts representative GLSZM features that provided the most discriminative power. There is a total of 160 features from tumour core and 5-mm margin, including 10 image quality features, available for feature selection. Among these features, 4 mean-values, 125 textural, and 1 image quality features demonstrate statistically significant difference between the two lesions. Statistically significant ($p < 0.05$), highly significant ($p < 0.01$), and extremely significant ($p < 0.001$) are shown with (*), (**), and (***), respectively.

<https://doi.org/10.1371/journal.pone.0244965.g002>

spiculated. In the parametric images, benign lesions demonstrated less obvious heterogeneity than was apparent in malignant lesions.

Fig 2 shows representative box and scatter plots of mean-value, GLCM, GRLM, and GLSZM texture values, along with image quality features that demonstrated statistically significant differences ($p < 0.05$) between benign and malignant breast lesions. Six mean-value, 24 GLCM, 127 GRLM, 126 GLSZM texture, and 4 image quality features demonstrated statistically significant differences ($p < 0.05$). Features were further subclassified based on their degree of statistical significance. Statistically significant ($p < 0.05$), highly significant ($p < 0.01$), and extremely significant ($p < 0.001$) features are indicated with (*), (**), and (***), respectively. Among the mean-value parameters from the core, MBF, SI, and AAC demonstrated statistically significant differences ($p < 0.05$). The MBF, SI, and AAC parameters from the core were 4.4 ± 0.6 dB versus 2.2 ± 0.5 dB, 13.8 ± 0.7 dB versus 10.0 ± 0.6 dB, and 46.9 ± 0.9 dB/cm³ versus 43.6 ± 0.7 dB/cm³ for benign and malignant lesions, respectively.

Table 1. Optimum feature set for classification using both core and margin information utilizing GLCM, GRLM, and GLSZM texture methods and SVM-RBF classification algorithm. A maximum of 10 features was selected for classification. Model performance was evaluated using LOOCV method. Features were selected using forward SFS based on F1-score metric. Textural features, for example Core-MBF-CON: GLCM contrast parameter of MBF parametric image from core ROI and Margin-MBF-SALGE: GLSZM small area low gray level emphasis parameter of MBF parametric image from margin ROI, were the dominant features that contributed to hybrid biomarkers that best separated the two lesion types.

GLCM Selected Features	GRLM Selected Features	GLSZM Selected Features
Core-MBF-CON	Core-MBF-SRE	Margin-MBF-GLN-SZ
Margin-MBF	CMCR-MBF	Margin-SI-GLN-SZ
CMR-AAC	Margin-AAC-SRE	Margin-MBF-LGZE
Margin-SS	Margin-MBF-RP	Core-MBF-GV-SZ
CMCR-AAC	Margin-AAC-RP	Core-SS-GLN-SZ
Margin-AAC	CMCR-AAC	Margin-AAC-GV-SZ
Core-SS-CON	Core-AAC-SRE	Core-SS-LGZE
Margin-ASD-CON	Core-MBF-RP	Margin-MBF-SALGE
Core-ASD-CON	Core-AAC-RP	Margin-SS-SZN
CMR-SS	Margin-SI-RP	Core-ASD-SAE

<https://doi.org/10.1371/journal.pone.0244965.t001>

Among the mean-value parameters from the margin, the MBF, SI, and AAC also demonstrated statistically significant differences ($p < 0.05$). The MBF, SI, and AAC from the margin are $11.2 \pm 0.3\text{dB}$ versus $9.2 \pm 0.3\text{dB}$, $20.4 \pm 0.5\text{dB}$ versus $17.1 \pm 0.5\text{dB}$, and $51.4 \pm 0.6\text{dB/cm}^3$ versus $50.1 \pm 0.5\text{dB/cm}^3$ for benign and malignant lesions, respectively.

Table 1 lists an optimum set of features from GLCM, GRLM, and GLSZM methods that contributed to a hybrid biomarker that best separated benign from malignant lesions using breast mass core and margin information. The best classification performance using the SVM-RBF was achieved using features derived from the GLSZM methodology: Margin-MBF-GLN-SZ, Margin-SI-GLN-SZ, Margin-MBF-LGZE, Core-MBF-GV-SZ, Core-SS-GLN-SZ, Margin-AAC-GV-SZ, Core-SS-LGZE, Margin-MBF-SALGE, Margin-SS-SZN, and Core-ASD-SAE. Texture features dominated the best hybrid biomarker which separated patients into whether they had benign or malignant characteristics.

Table 2A and 2B tabulate classification performance utilizing core GLCM features evaluated using LOOCV and hold-out validation, respectively. Using LOOCV, the SVM-RBF provided the best classification performance of 84% sensitivity, 78% specificity, 81% accuracy, 0.88 AUC, 81% PPV, and 82% NPV. Using hold-out validation, the ANN resulted in the best performance of 89% sensitivity, 77% specificity, 83% accuracy, 0.92 AUC, 81% PPV, and 86% NPV.

Table 2. A: Core classification results of GLCM methodology using LOOCV. **B:** Core classification results of GLCM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	86%	73%	80%	0.84	78%	83%
KNN	75%	71%	73%	0.77	74%	72%
SVM-RBF	84%	78%	81%	0.88	81%	82%
LDA	79%	65%	72%	0.81	71%	74%
KNN	72%	67%	70%	0.77	71%	69%
SVM-RBF	87%	69%	79%	0.82	76%	83%
ANN	89%	77%	83%	0.92	81%	86%

<https://doi.org/10.1371/journal.pone.0244965.t002>

Table 3. A: Margin classification results of GLCM methodology using LOOCV. B: Margin classification results of GLCM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	68%	74%	71%	0.75	74%	68%
KNN	75%	66%	71%	0.76	71%	71%
SVM-RBF	81%	75%	78%	0.80	78%	78%
LDA	60%	71%	65%	0.73	70%	61%
KNN	61%	60%	61%	0.61	64%	58%
SVM-RBF	67%	65%	66%	0.69	68%	64%
ANN	70%	80%	75%	0.84	81%	72%

<https://doi.org/10.1371/journal.pone.0244965.t003>

Table 3A and 3B show classification performance utilizing margin GLCM features evaluated using LOOCV and hold-out validation, respectively. Using LOOCV, the SVM-RBF attained the best performance of 81% sensitivity, 75% specificity, 78% accuracy, 0.80 AUC, 78% PPV, and 78% NPV. Using hold-out validation technique, the ANN resulted in the best classification performance of 70% sensitivity, 80% specificity, 75% accuracy, 0.84 AUC, 81% PPV, and 72% NPV.

Table 4A and 4B present classification performance utilizing core and margin GLCM features evaluated using LOOCV and hold-out validation, respectively. Using LOOCV, the SVM-RBF provided the best classification performance of 86% sensitivity, 83% specificity, 84% accuracy, 0.90 AUC, 84% PPV, and 84% NPV. Using hold-out validation, the ANN resulted in the best classification performance of 88% sensitivity, 78% specificity, 83% accuracy, 0.92 AUC, 83% PPV, and 87% NPV. Core GLCM features alone resulted in the highest 89% sensitivity detection of malignancy and overall accuracy and AUC of 83% and 0.92, respectively.

Table 5A and 5B tabulate classification performance utilizing core GRLM features evaluated using LOOCV and hold-out validation, respectively. Using LOOCV, the SVM-RBF provided

Table 4. A: Core and margin classification results of GLCM methodology using LOOCV. B: Core and margin classification results of GLCM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	83%	78%	81%	0.86	81%	81%
KNN	82%	72%	77%	0.84	76%	79%
SVM-RBF	86%	83%	84%	0.90	84%	84%
LDA	74%	67%	71%	0.80	72%	71%
KNN	72%	67%	69%	0.74	71%	68%
SVM-RBF	78%	64%	71%	0.81	72%	73%
ANN	88%	78%	83%	0.92	83%	87%

<https://doi.org/10.1371/journal.pone.0244965.t004>

Table 5. A: Core classification results of GRLM methodology using LOOCV. B: Core classification results of GRLM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	85%	76%	81%	0.87	80%	82%
KNN	85%	76%	81%	0.84	80%	82%
SVM-RBF	90%	83%	87%	0.87	85%	88%
LDA	69%	77%	73%	0.82	78%	70%
KNN	70%	67%	68%	0.74	70%	67%
SVM-RBF	72%	71%	72%	0.75	74%	70%
ANN	86%	82%	84%	0.93	84%	86%

<https://doi.org/10.1371/journal.pone.0244965.t005>

Table 6. A: Margin classification results of GRLM methodology using LOOCV. B: Margin classification results of GRLM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	74%	86%	80%	0.82	85%	75%
KNN	85%	83%	84%	0.82	84%	84%
SVM-RBF	85%	86%	85%	0.87	87%	84%
LDA	65%	85%	74%	0.83	83%	69%
KNN	72%	78%	75%	0.79	80%	72%
SVM-RBF	73%	67%	70%	0.76	72%	70%
ANN	90%	84%	87%	0.93	87%	88%

<https://doi.org/10.1371/journal.pone.0244965.t006>

the best classification performance of 90% sensitivity, 83% specificity, 87% accuracy, 0.87 AUC, 85% PPV, 88% NPV. Using hold-out validation, the ANN resulted in the best classification performance of 86% sensitivity, 82% specificity, 84% accuracy, 0.93 AUC, 84% PPV, and 86% NPV.

Table 6A and 6B tabulate classification performance utilizing margin GRLM features evaluated using LOOCV and hold-out validation, respectively. Using LOOCV, the SVM-RBF resulted in the best classification performance of 85% sensitivity, 86% specificity, 85% accuracy, 0.87 AUC, 87% PPV, 84% NPV. Using hold-out validation, the ANN obtained the best classification performance of 90% sensitivity, 84% specificity, 87% accuracy, 0.93 AUC, 87% PPV, 88% NPV.

Table 7A and 7B tabulate classification performance utilizing core and margin GRLM features evaluated using LOOCV and hold-out validation, respectively. Using LOOCV, the SVM-RBF achieved the best classification performance of 86% sensitivity, 85% specificity, 85% accuracy, 0.87 AUC, 86% PPV, 85% NPV. Using hold-out validation, the ANN achieved the best classification performance of 92% sensitivity, 86% specificity, 89% accuracy, 0.95 AUC, 88% PPV, 90% NPV. Margin GRLM features performed better than core GRLM features. Combining features from both the core and the margin resulted in improved classification performance with 92% sensitivity and accuracy and AUC of 89% and 0.95, respectively.

Table 8A and 8B tabulate classification performance using core GLSZM features evaluated using LOOCV and hold-out validation, respectively. Using LOOCV, the SVM-RBF resulted in the best classification performance of 90% sensitivity, 85% specificity, 88% accuracy, 0.89 AUC, 87% PPV, 89%. Using hold-out validation, the ANN achieved the best performance of 93% sensitivity, 88% specificity, 91% accuracy, 0.95 AUC, 90% PPV, 92% NPV.

Table 9A and 9B tabulate the classification performance utilizing margin GLSZM features using LOOCV and hold-out validation, respectively. Using LOOCV, the SVM-RBF resulted in the best classification performance of 90% sensitivity, 90% specificity, 90% accuracy, 0.91

Table 7. A: Core and margin classification results of GRLM methodology using LOOCV. B: Core and margin classification results of GRLM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	77%	85%	81%	0.82	85%	77%
KNN	86%	85%	85%	0.83	86%	85%
SVM-RBF	86%	85%	85%	0.87	86%	85%
LDA	70%	81%	75%	0.84	81%	72%
KNN	74%	77%	75%	0.79	79%	73%
SVM-RBF	71%	74%	72%	0.77	77%	70%
ANN	92%	86%	89%	0.95	88%	90%

<https://doi.org/10.1371/journal.pone.0244965.t007>

Table 8. A: Core classification results of GLSZM methodology using LOOCV. B: Core classification results of GLSZM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	82%	87%	84%	0.87	87%	82%
KNN	84%	80%	82%	0.82	83%	82%
SVM-RBF	90%	85%	88%	0.89	87%	89%
LDA	77%	76%	77%	0.85	79%	75%
KNN	75%	74%	75%	0.78	77%	73%
SVM-RBF	75%	72%	74%	0.80	75%	72%
ANN	93%	88%	91%	0.95	90%	92%

<https://doi.org/10.1371/journal.pone.0244965.t008>

Table 9. A: Margin classification results of GLSZM methodology using LOOCV. B: Margin classification results of GLSZM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	84%	89%	87%	0.89	89%	84%
KNN	87%	85%	86%	0.90	86%	86%
SVM-RBF	90%	90%	90%	0.91	91%	89%
LDA	69%	87%	78%	0.88	87%	72%
KNN	75%	79%	77%	0.81	81%	75%
SVM-RBF	74%	86%	80%	0.88	87%	76%
ANN	89%	91%	90%	0.95	92%	88%

<https://doi.org/10.1371/journal.pone.0244965.t009>

AUC, 91% PPV, 89% NPV. Using hold-out validation, the ANN provided the best performance of 89% sensitivity, 91% specificity, 90% accuracy, 0.95 AUC, 92% PPV, 88% NPV.

Table 10A and 10B tabulate classification performance utilizing core and margin GLSZM features evaluated using LOOCV and hold-out validation. Using LOOCV, the SVM-RBF resulted in the best classification performance of 90% sensitivity, 90% specificity, 90% accuracy, 0.90 AUC, 91% PPV, 89% NPV. Using hold-out validation, the ANN achieved the best performance of 89% sensitivity, 91% specificity, 90% accuracy, 0.96 AUC, 92% PPV, 89% NPV.

These results suggested that GRLM and GLSZM features outperform those of GLCM features. This supports our hypothesis that one texture analysis method may perform better than the others. Using LOOCV, our results indicated that core classification performs better than margin classification in the case of GLCM and GRLM. For the GLSZM, there was a slight improvement in performance by combining core and rim features, compared to that using core features alone. Using hold-out validation, core classification performed better than margin classification using GLCM features. However, using GRLM or GLSZM features to develop

Table 10. A: Core and margin classification results of GLSZM methodology using LOOCV. B: Core and margin classification results of GLSZM methodology using hold-out validation.

Classifier	Sensitivity	Specificity	Accuracy	AUC	PPV	NPV
LDA	80%	93%	87%	0.87	93%	81%
KNN	85%	87%	86%	0.87	88%	84%
SVM-RBF	90%	90%	90%	0.90	91%	89%
LDA	71%	90%	80%	0.87	90%	74%
KNN	72%	80%	76%	0.80	81%	72%
SVM-RBF	80%	84%	82%	0.87	86%	79%
ANN	89%	91%	90%	0.96	92%	89%

<https://doi.org/10.1371/journal.pone.0244965.t010>

a model, resulted in better margin classification than core classification. Overall, combining both core and margin information resulted in improved classification performance. Among the validation techniques, LOOCV led to better performance than hold-out validation. Although decreases in classification performance were observed using the latter validation method, the best averaged performance of 91% accuracy and 0.95 AUC were obtained utilizing GLSZM features from the tumour core, using the ANN. GLSZM features also attained 90% accuracy and 0.96 AUC with core and margin information using the ANN. Advanced machine learning classifier in ANN proves to be more robust in generalizing diagnostic model inference, compared to those of standard classifiers.

Discussion

In this study, the performance of different texture analysis methods applied on QUS spectral parametric images for the characterization of breast lesions was demonstrated, for the first time. Textural features derived from the GLCM, GRLM, and GLSZM methods were used as imaging biomarkers to develop a diagnostic model for classifying breast lesions as either benign or malignant. In addition to analyzing features from tumour core, analyses conducted also included peri-tumoural tissue (5-mm margin extending from tumour core). In invasive tumours, the rim contains infiltrating components that extend from tumour core into the surrounding tissue [71]. Earlier, tumour rim analysis has been used to predict the response to NAC [62]. Here, rim analysis was used to characterize breast lesions. This study builds upon previous studies through a significant expansion of the cohort and a comparison of different texture methods. Earlier, the cohort consisted of 78 patients with breast lesions (46 benign and 32 malignant cases) [8]. Recently, novel derivative texture methods were also evaluated on a larger cohort of patients with breast lesions [9]. In those studies, however, only the GLCM method was used to quantify texture of the parametric images. In the current study, different texture methods were applied here on a larger cohort of 193 patients with breast lesions (92 benign and 101 malignant cases). Larger cohort allowed assessment of the model performance using both LOOCV and hold-out validation. The latter demonstrates model generalizability to independent testing sets. Findings from this study suggest that different texture methods can affect classification performance. Specifically, tumour core features derived from the GLSZM method demonstrated the best classification performance of 93% sensitivity, 88% specificity, 91% accuracy, 0.95 AUC, 90% PPV, and 92% NPV with hold-out validation, utilizing the ANN.

In a previous study, the average-values of MBF, SI, and AAC images did not show statistically significant differences ($p < 0.05$) [8]. In the study here, however, these parameters from tumour core and a 5-mm margin showed statistically significant differences ($p < 0.05$). This can be attributed to the increased size of the cohort. The same trend was also observed in our recent study [9]. Malignant lesions exhibited lower MBF, SI, and AAC compared to those of benign lesions. An earlier study also observed the same trend of lower QUS spectral parameters in cancerous versus those of normal breast tissues [38]. Furthermore, this observation is also generally consistent with sonographic features of B-mode US images of breast nodules where marked hypo-echogenicity was observed in the malignant lesions compared to the benign lesions [10]. The MBF and SI represent tissue microstructural characteristics that include the size, shape, number, and organization of acoustic scatterers, along with their elastic properties [19]. On the other hand, the AAC reflects scatterers number density, organization, and their elastic properties [19]. Histopathological analysis has demonstrated that related tissue structural properties are distinct between benign and malignant lesions [72]. A more regular arrangement of cells is observed in benign lesions [72]. In contrast, malignant lesions exhibit cellularity-rich areas with a tendency to form cell clusters [72].

As average-based parameters do not preserve information regarding tumour heterogeneity, texture analysis is needed. Texture analysis of QUS spectral parametric images can quantify lesion heterogeneities that includes variations in size, density, and distribution of acoustic scatterers. Better discrimination of different histological tissue types is potentially achievable using these imaging biomarkers compared to mean-value parameters. Among the GLCM features, 24 biomarkers showed statistically significant differences ($p < 0.05$). Among the GRLM features, 127 biomarkers demonstrated statistically significant differences ($p < 0.05$). Among the GLSZM features, 126 biomarkers showed statistically significant differences ($p < 0.05$). In addition, QUS spectral and texture analyses of tumour core and its 5-mm margin allowed us to obtain image quality features including the CMR and CMCR. Here, the CMR of MBF and SI demonstrated statistically significant differences ($p < 0.05$). Additionally, the CMCR of ASD and SS also showed statistically significant differences ($p < 0.05$).

Previously, evaluation of mean-value parameters and GLCM texture features for breast lesion characterization on a smaller subset of 78 patients achieved the best classification performance of 96% sensitivity, 84% specificity, 91% accuracy and 0.97 AUC [8]. However, application of the same approach on a larger cohort of 193 patients was only able to achieve the best classification performance of 84% sensitivity, 78% specificity, 81% accuracy, and 0.88 AUC, as shown in Table 2A. This suggests that generalization of GLCM-based texture analysis for breast lesions characterization is not optimum. However, when different texture methods were considered, results like that of a recent study that include mean-value parameters, GLCM texture, and novel GLCM texture-derivate features of QUS spectral parametric images were achieved [9]. Past studies on classification using different texture methods suggested that GLCM-based features performed the least optimum in comparison to those using run-length (GRLM) and size zone (GLSZM) features [61]. This is consistent with the observations in the study here where GLSZM proved to be the optimum texture analysis approach for breast lesion characterization (91% accuracy and 0.95 AUC using core analysis and the ANN). The GRLM method marginally underperformed the GLSZM method (89% accuracy vs 91% accuracy). This is not unexpected as both techniques have similar matrix construction, albeit different interpretation. The results suggest that optimizing methods for extracting discriminative textural features can improve classification performance. Further application of derivative texture methods using GLSZM texture analysis on the QUS spectral parametric images can potentially improve the classification performance further. This type of investigation will be conducted in future studies.

Here, we evaluated model performance using LOOCV and hold-out validation. As expected, LOOCV led to better classification results than those of hold-out validation. However, hold-out validation is necessary to demonstrate the generalizability of the model. In terms of the classification algorithms, nonlinear classifiers in the SVM-RBF and ANN proved to be more robust to random data partitioning compared to the LDA and KNN, allowing for better generalization. Using LOOCV, the best classification performance of 90% sensitivity, 90% specificity, 90% accuracy, 0.91 AUC, 91% PPV, and 89% NPV was achieved, utilizing margin GLSZM texture features and the SVM-RBF. Using hold-out validation, core GLSZM features resulted in the best average performance of 93% sensitivity, 88% specificity, 91% accuracy, 0.95 AUC, 90% PPV, and 92% NPV using the ANN. Although random partitioning of the data can result in sub-optimal classification performance, the network was still able to pick up necessary patterns in the training data and generalize in predicting class association of the testing set. As more breast US RF data from patients are acquired over time, ANN and deep learning techniques may prove to be more effective classification algorithms for maximizing the classification performance.

Past studies have demonstrated the use of B-mode US images and texture analysis of these images in the characterization of breast lesions. Stavros *et al.* performed manual classification of solid breast nodules in 750 patients (625 benign and 125 malignant) using B-mode US images [10]. Using sonographical features of the lesions (for example echogenicity, shape, contour, and surrounding tissue), they achieved 98% sensitivity, 68% specificity, and 73% accuracy [10]. Tsui *et al.* analyzed the statistics of backscattered echo envelope using Nakagami statistical model, attaining 92% sensitivity, 72% specificity, and 82% accuracy in the characterization of 100 patients with breast tumours (50 benign and 50 malignant) [73]. Furthermore, Destrempes *et al.* explored various combinations of features from shear wave elastography (SWE), RF spectral analysis, and echo envelope statistical analysis, along with BI-RADS score in the classification of 103 suspicious solid breast lesions from 103 patients (BI-RADS 3–4) [74]. They observed that the combination of SWE, QUS, and BI-RADS scoring led to an AUC of 0.97, with 76% specificity at 98% sensitivity [74]. In addition, Dobruch-Sobczak *et al.* also found that the combination of echo envelope statistics features and BI-RADS scoring achieved 100% sensitivity, 55% specificity, and an AUC of 0.97 in the classification of 107 solid or cystic-solid breast lesions from 78 patients [75]. Gomez *et al.* utilized GLCM textural features from 436 breast US images (219 benign and 217 carcinoma) in the characterization of breast lesions, achieving classification performance of 70% sensitivity, 77% specificity, and 74% accuracy [54]. Further works involving the applications of ANN on breast US images resulted in an improved classification performance with 92% sensitivity, 91% specificity, and 91% accuracy [11,12]. Recently, Han *et al.* utilized a deep learning framework to differentiate benign from malignant breast lesions using the GoogLeNet convolutional neural network (CNN) on a large data set comprised of 7,408 breast US images from 5,151 patients [13]. They obtained 90% accuracy, 86% sensitivity, 96% specificity, and 0.90 AUC [13]. In another study, Byra *et al.* also developed a deep learning-based approach using deep CNN to classify breast lesions on 882 breast US images [14]. The trained network achieved 0.94 AUC on the test set of 150 cases [14]. Recently, Osapoetra *et al.* performed characterization of breast lesions using the combination of several single biomarkers from mean-value parameters, texture, and texture-derivate features of QUS spectral parametric images, achieving 90% sensitivity, 92% specificity, and 91% accuracy, and 0.93 AUC using the SVM-RBF classifier [9]. In that study, the GLCM method was used to extract texture and texture-derivate features from tumour core and its 5-mm margin. In this study, implementation of hold-out validation for assessing model performance resulted in the best classification performance of 93% sensitivity, 88% specificity, 91% accuracy, and 0.95 AUC using mean-value parameters, GLSZM texture, and image quality features. This demonstrates the generalizability of the QUS spectroscopy framework and texture methods in the characterization of breast lesions. Our results suggest that different methods for extracting textural features of QUS spectral parametric images can result in the same classification performance to that obtained using more computationally intensive derivative texture methods [9].

Conclusion

QUS-based techniques, along with optimized texture methods, provided an improved classification performance for the characterization of breast lesions compared to past work utilizing sonographical features of B-mode US images and other RF-based work. This result can be attributed to the fact that QUS techniques measure independent intrinsic acoustic and mechanical properties of tissue microstructure that are distinct between benign and more structurally disorganized malignant lesions. In addition, QUS spectral analysis allows measurement of instrument- and operator-independent tissue properties through a normalization

procedure. In the work here, the classification of breast lesions using these imaging biomarkers obtained from different texture methods resulted in a more robust classification model. Evaluation of QUS spectroscopy and texture analysis methods in a larger cohort using proper validation technique demonstrate the generalization of the proposed framework. Furthermore, QUS spectroscopy does not use ionizing radiation and does not need the administration of exogenous contrast agents. These advantages of QUS spectroscopy, along with texture analyses, over other imaging modalities including x-ray mammography, standard B-mode US, and contrast-enhanced MRI make it an ideal tool for rapid and accurate breast cancer diagnosis in clinical settings.

Supporting information

S1 Table. Benign patient characteristics. Lesion size refers to the longest dimension of the tumor.

(DOCX)

S2 Table. Malignant patient characteristics. Lesion size refers to the longest dimension of the tumor. ER is estrogen receptor, PR is progesterone receptor, HER2 is human epithelial growth factor receptor 2. Lesion size refers to the longest dimension of the tumor. IDC stands for invasive ductal carcinoma. ILC stands for invasive lobular carcinoma. DCIS stands for ductal carcinoma *in situ*, IMC stands for invasive mammary carcinoma.

(DOCX)

S3 Table. BI-RADS Distribution among the lesions evaluated.

(DOCX)

Author Contributions

Conceptualization: Laurentius O. Osapoetra, William Chan, William Tran, Michael C. Kolios, Gregory J. Czarnota.

Data curation: Laurentius O. Osapoetra, William Chan.

Formal analysis: Laurentius O. Osapoetra, William Chan, Gregory J. Czarnota.

Funding acquisition: Gregory J. Czarnota.

Investigation: Laurentius O. Osapoetra, William Chan.

Methodology: Laurentius O. Osapoetra, William Chan, Gregory J. Czarnota.

Project administration: Laurentius O. Osapoetra.

Resources: Gregory J. Czarnota.

Software: Laurentius O. Osapoetra, William Chan.

Supervision: Laurentius O. Osapoetra, Gregory J. Czarnota.

Validation: Laurentius O. Osapoetra, Gregory J. Czarnota.

Visualization: Gregory J. Czarnota.

Writing – original draft: Laurentius O. Osapoetra, William Chan.

Writing – review & editing: Laurentius O. Osapoetra, Gregory J. Czarnota.

References

1. Siegel RL, Miller KD, Jemal A, Cancer Statistics. *CA Cancer J. Clin.* 2016; 66: 7–30. <https://doi.org/10.3322/caac.21332> PMID: 26742998
2. Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, et al, on behalf of the ESMO Guidelines Committee. Primary Breast Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment, and Follow-up. *Annals of Oncology.* 2015; 26: v8–v30. <https://doi.org/10.1093/annonc/mdv298> PMID: 26314782
3. <https://gis.cdc.gov/Cancer/USCS/DataViz.html>.
4. Silverstein MJ, et al. Image-Detected Breast Cancer: State-of-the-Art Diagnosis and Treatment. *Journal of the American College of Surgeons.* 2005; 201: 586–597. <https://doi.org/10.1016/j.jamcollsurg.2005.05.032> PMID: 16183499
5. Loughran CF and Keeling CR. Seeding of Tumour Cells following Breast Biopsy: A Literature Review. *The British Journal of Radiology.* 2011; 84: 869–874. <https://doi.org/10.1259/bjr/77245199> PMID: 21933978
6. Oelze ML. Quantitative Ultrasound Techniques and Improvements to Diagnostic Ultrasound Imaging. *IEEE International Ultrasonics Symposium, Dresden.* 2012; 232–239 <https://doi.org/10.1109/ULTSYM.2012.0058>
7. Bhooshan N, Giger ML, Jansen SA, Li H, Lan L, Newstead GM. Cancerous Breast Lesions on Dynamic Contrast-enhanced MR Images: Computerized Characterization for Image-based Prognostic Markers. *Radiology.* 2010; 254: 680–690. <https://doi.org/10.1148/radiol.09090838> PMID: 20123903
8. Sadeghi-Naini A, Suraweera H, Tran WT, Hadizad F, Bruni G, Rastegar RF, et al. Breast-Lesion Characterization using Textural Features of Quantitative Ultrasound Parametric Maps. *Scientific Reports.* 2017; 7:13638. <https://doi.org/10.1038/s41598-017-13977-x> PMID: 29057899
9. Osapoetra LO, Sannachi L, DiCenzo D, Quiaoit K, Fatima K, Czarnota GJ. Breast Lesion Characterization using Quantitative Ultrasound (QUS) and Derivative Texture Methods. *Translational Oncology.* 2020; 13: 100827. <https://doi.org/10.1016/j.tranon.2020.100827> PMID: 32663657
10. Stavros AT, Thickman D, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid Breast Nodules: Use of Sonography to Distinguish between Benign and Malignant Lesions. *Radiology.* 1995; 196: 123–134. <https://doi.org/10.1148/radiology.196.1.7784555> PMID: 7784555
11. Joo S, Yang YS, Moon WK, Kim HC. Computer-aided Diagnosis of Solid Breast Nodules: Use of An Artificial Neural Network Based on Multiple Sonographic Features. *IEEE Trans. On Med. Imaging.* 2004; 23: 1292–1300. <https://doi.org/10.1109/TMI.2004.834617> PMID: 15493696
12. Chen CM, Chou YH, Han KC, Hung GH, Tiu CM, Chiou HJ, et al. Breast Lesions on Sonograms: Computer-aided Diagnosis with Nearly Setting Independent Features and Artificial Neural Networks. *Radiology.* 2003; 226: 504–514. <https://doi.org/10.1148/radiol.2262011843> PMID: 12563146
13. Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A Deep Learning Framework for Supporting the Classification of Breast Lesions in Ultrasound Images. *Physics in Medicine and Biology.* 2017; 62: 7714–7728. <https://doi.org/10.1088/1361-6560/aa82ec> PMID: 28753132
14. Byra M, Galperin M, Ojeda-Fournier H, Olson L, O'Boyle M, Comstock C, et al. Breast Mass Classification in Sonography with Transfer Learning using a Deep Convolutional Neural Network and Color Conversion. *Medical Physics.* 2019; 46: 746:755. <https://doi.org/10.1002/mp.13361> PMID: 30589947
15. Yao LX, Zagzebski JA, Madsen EL. Backscatter Coefficient Measurements using A Reference Phantom to Extract Depth-dependent Instrumentation Factors. *Ultrason. Imaging.* 1990; 12: 58–70. <https://doi.org/10.1177/016173469001200105> PMID: 2184569
16. Mamou J and Oelze ML (eds). *Quantitative ultrasound in soft tissues.* Springer: Dordrecht. 2013.
17. Lizzi FL, Astor M, Feleppa EJ, Shao M, Kalisz A. Statistical framework for ultrasonic spectral parameter imaging. *Ultrasound Med Biol.* 1997; 23: 1371–1382. [https://doi.org/10.1016/s0301-5629\(97\)00200-7](https://doi.org/10.1016/s0301-5629(97)00200-7) PMID: 9428136
18. Lizzi FL, Feleppa EJ, Astor M, Kalisz A. Statistics of ultrasonic spectral parameters for prostate and liver examinations. *IEEE Trans Ultrason, Ferroelec, Freq Contr.* 1997; 44: 935–942. <https://doi.org/10.1109/58.655209>
19. Lizzi FL, Ostromogilsky M, Feleppa EJ, Rorke MC, Yaremko MM. Relationship of ultrasonic spectral parameters to features of tissue microstructure. *IEEE Trans Ultrason, Ferroelect, Freq Contr.* 1986; 33: 319–329. <https://doi.org/10.1109/T-UFFC.1987.26950> PMID: 18291854
20. Sannachi L, Tadayyon H, Sadeghi-Naini A, Tran W, Gandhi S, Wright F, et al. Non-invasive evaluation of breast cancer response to chemotherapy using quantitative ultrasonic backscatter parameters. *Medical Image Analysis.* 2015; 20: 224–236. <https://doi.org/10.1016/j.media.2014.11.009> PMID: 25534283

21. Oelze ML, O'Brien WD, Blue JP, Zachary JF. Differentiation and characterization of rat mammary fibroadenomas and 4T1 mouse carcinomas using quantitative ultrasound imaging. *IEEE Trans. Med. Imaging*. 2004; 23: 764–771. <https://doi.org/10.1109/tmi.2004.826953> PMID: 15191150
22. Oelze ML and Zachary JF. Examination of cancer in mouse models using high-frequency quantitative ultrasound. *Ultrasound Med. Biol.* 2006; 32: 1639–1648. <https://doi.org/10.1016/j.ultrasmedbio.2006.05.006>. PMID: 17112950
23. Sadeghi-Naini A, Falou O, Hudson JM, Bailey C, Burns PM, Yaffe MJ, et al. Imaging Innovations for Cancer Therapy Response Monitoring. *Imaging Med.* 2012; 4: 311–327.
24. Sadeghi-Naini A, Sannachi L, Pritchard K, Trudeau M, Gandhi S, Wright FC, et al. Early Prediction of Therapy Responses and Outcomes in Breast Cancer Patients using Quantitative Ultrasound Spectral Texture. *Oncotarget*. 2014; 5: 3497–3511. <https://doi.org/10.18632/oncotarget.1950> PMID: 24939867
25. Sadeghi-Naini, Papanicolau N, Falou O, Zubovits J, Dent R, Verma S, et al. Quantitative Ultrasound Evaluation of Tumour Cell Death Response in Locally-advanced Breast Cancer Patients Receiving Chemotherapy. *Clin. Cancer Res.* 2013; 19: 2163–2174. <https://doi.org/10.1158/1078-0432.CCR-12-2965> PMID: 23426278
26. Sadeghi-Naini A, Vorauer E, Chin L, Falou O, Tran WT, Wright FC, et al. Early Detection of Chemotherapy-refractory Patients by Monitoring Textural Alterations in Diffuse Optical Spectroscopic Images. *Med. Phys.* 2015; 42: 6130–6146. <https://doi.org/10.1118/1.4931603>. PMID: 26520706
27. Sannachi L, Gangeh M, Tadayyon H, Gandhi S, Wright FC, Slodkowska E, et al. Breast Cancer Treatment Response Monitoring using Quantitative Ultrasound and Texture Analysis: Comparative Analysis of Computational Models. *Translational Oncology*. 2019; 12: 1271–1281. <https://doi.org/10.1016/j.tranon.2019.06.004>. PMID: 31325763
28. Feleppa EJ, Fair WR, Liu T, Kalisz A, Balaji KC, Porter CR, et al. Three-dimensional Ultrasound Analyses of the Prostate. *Mol. Urol.* 2000; 4: 133–139. PMID: 11062367
29. Feleppa EJ, Kalisz A, Sokil-Melgar JB, Lizze FL, Liu T, Rosado AL, et al. Typing of Prostate Tissue by Ultrasonic Spectrum Analysis. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.* 1996; 43: 609–619. <https://doi.org/10.1109/58.503779>
30. Lizzi FL, Astor M, Liu T, Deng C, Coleman DJ, Silverman RH. Ultrasonic spectrum analysis for tissue assays and therapy evaluation. *Int. J. Imaging Syst. Technol.* 1997; 8: 3–10. [https://doi.org/10.1002/\(SICI\)1098-1098\(1997\)8:1<3::AID-IMA2>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1098-1098(1997)8:1<3::AID-IMA2>3.0.CO;2-E).
31. Feleppa EJ, Mamou J, Porter CR, Machi J. Quantitative ultrasound in cancer imaging. *Semin. Oncol.* 2011; 38: 136–150. <https://doi.org/10.1053/j.seminoncol.2010.11.006>. PMID: 21362522
32. Balaji KC, Fair WR, Feleppa EJ, Porter CR, Tsai H, Liu T, et al. Role of advanced 2 and 3-dimensional ultrasound for detecting prostate cancer. *J. Urol.* 2002; 168: 2422–2425. <https://doi.org/10.1097/01.ju.0000036435.13421.57> PMID: 12441931
33. Feleppa EJ. Ultrasonic tissue-type imaging of the prostate: implications for biopsy and treatment guidance. *Cancer biomarkers*. 2008; 4: 201–212. <https://doi.org/10.3233/cbm-2008-44-504> PMID: 18957711
34. Sigel B, Feleppa EJ, Swami V, Justin J, Consigny M, Machi J, et al. Ultrasonic Tissue Characterization of Blood Clots. *Surg. Clin. North Am.* 1990; 70:13–29. [https://doi.org/10.1016/s0039-6109\(16\)45030-9](https://doi.org/10.1016/s0039-6109(16)45030-9) PMID: 2406962
35. Noritomi T, Sigel B, Swami V, Justin J, Gahtan V, Chen X, et al. Carotid Plaque Typing by Multiple-parameter Ultrasonic Tissue Characterization. *Ultrasound Med. Biol.* 1997; 23: 643–650. [https://doi.org/10.1016/s0301-5629\(97\)00013-6](https://doi.org/10.1016/s0301-5629(97)00013-6) PMID: 9253812
36. Konig A and Klauss V. Virtual histology. *Heart*. 2007; 93: 977–982. <http://dx.doi.org/10.1136/hrt.2007.116384>. PMID: 17502324
37. Mamou J, Coron A, Oelze ML, Saegusa-Becroft E, Hata M, Lee P, et al. Three-dimensional high-frequency backscatter and envelope quantification of cancerous human lymph nodes. *Ultrasound Med. Biol.* 2011; 37: 345–357. <https://doi.org/10.1016/j.ultrasmedbio.2010.11.020>. PMID: 21316559
38. Tadayyon H, Sadeghi-Naini A, Wirtzfeld L, Wright FC, Czarnota G. Quantitative Ultrasound Characterization of Locally-advanced Breast Cancer by Estimation of its Scatterer Properties. *Med. Phys.* 2014; 41: 129031–1290312. <http://dx.doi.org/10.1118/1.4852875>. PMID: 24387530
39. O'Connor JPB, Waterton JC, Carano RAD, Parker GJM, Jackson A. Imaging intratumour heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. *Clin. Cancer Res.* 2015; 21: 249–257. <https://doi.org/10.1158/1078-0432.CCR-14-0990> PMID: 25421725
40. Polyak K. Heterogeneity in Breast Cancer. *J. Clin. Invest.* 2011; 121: 3786–3788. <https://doi.org/10.1172/JCI60534>. PMID: 21965334

41. Heindl A, Nawaz S, Yuan Y. Mapping Spatial Heterogeneity in the Tumour Microenvironment: A New Era for Digital Pathology. *Lab. Investig.* 2015; 95: 377–384. <https://doi.org/10.1038/labinvest.2014.155>. PMID: 25599534
42. Sengupta D and Pratz G. Imaging Metabolic Heterogeneity in Cancer. *Mol. Cancer.* 2016; 15: 1–12. <https://doi.org/10.1186/s12943-015-0481-3>. PMID: 26728659
43. Davnall F, Yip CSP, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of Tumour Heterogeneity: An Emerging Imaging Tool for Clinical Practice? *Insights Imaging.* 2012; 3: 573–589. <https://doi.org/10.1007/s13244-012-0196-6> PMID: 23093486
44. Ahmed A, Gibbs P, Pickles M, Turnbull M. Texture Analysis in Assessment and Prediction of Chemotherapy Response in Breast Cancer. *J. Magn. Reson. Imaging.* 2013; 38: 89–101. <https://doi.org/10.1002/jmri.23971>. PMID: 23238914
45. Tan S, Kligerman S, Chen W, Lu M, Kim G, Feigenberg S, et al. Spatial-temporal [18F]FDG-PET Features for Predicting Pathologic Response of Esophageal Cancer to Neoadjuvant Chemoradiation Therapy. *Int. J. Radiat. Oncol. Biol. Phys.* 2013; 85: 1375–1382. <https://doi.org/10.1016/j.ijrobp.2012.10.017>. PMID: 23219566
46. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJR. Quantifying Tumour Heterogeneity in 18F-FDG PET/CT Imaging by Texture Analysis. *Eur. J. Nucl. Med. Mol. Imaging.* 2013; 40: 133–140. <https://doi.org/10.1007/s00259-012-2247-0>. PMID: 23064544
47. Vaidya M, Creach KM, Frye J, Dehdashti F, Bradley JD, El Naqa I. Combined PET/CT Image Characteristics for Radiotherapy Tumour Response in Lung Cancer. *Radiother. Oncol.* 2012; 102: 239–245. <https://doi.org/10.1016/j.radonc.2011.10.014>. PMID: 22098794
48. Goh V, Ganeshan B, Nathan P, Juttla JK, Vinayan A, Miles KA. Assessment of Response to Tyrosine Kinase Inhibitors in Metastatic Renal Cell Cancer: CT Texture as a Predictive Biomarker. *Radiology.* 2011; 261: 165–171. <https://doi.org/10.1148/radiol.11110264>. PMID: 21813743
49. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cybern.* 1973; 6: 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
50. Gomez W, Pereira WCA, Infantosi AFC. Analysis of Co-occurrence Texture Statistics as a Function of Gray-Level Quantization for Classifying Breast Ultrasound. *IEEE Tran. Medical Imaging.* 2012; 31: 1889–1899. <https://doi.org/10.1109/TMI.2012.2206398> PMID: 22759441
51. Liao YY, Tsui PH, Li CH, Chang KJ, Kuo WH, Chang CC, et al. Classification of Scattering Media within Benign and Malignant Breast Tumours based on Ultrasound Texture-feature based and Nakagami-parameter Images. *Med. Phys.* 2011; 38: 2198–2207. <https://doi.org/10.1118/1.3566064>. PMID: 21626954
52. Garra BS, Krasner BH, Horii SC, Ascher S, Mun SK, Zeman RK. Improving the Distinction between Benign and Malignant Breast Lesions: The Value of Sonographic Texture Analysis. *Ultrasonic Imaging.* 1993; 15: 267–285. <https://doi.org/10.1177/016173469301500401>. PMID: 8171752
53. Alvarenga AV, Pereira WCA, Infantosi FC, Azevedo CM. Complexity Curve and Grey Level Co-occurrence Matrix in the Texture Evaluation of Breast Tumour on Ultrasound Images. *Med. Phys.* 2007; 34: 379–387. <https://doi.org/10.1118/1.2401039>. PMID: 17388154
54. Gomez W, Pereira WCA, Infantosi AFC. Analysis of Co-Occurrence Texture Statistics as a Function of Gray-Level Quantization for Classifying Breast Ultrasound. *IEEE Trans. Med. Imaging.* 2012; 31: 1889–1899. <https://doi.org/10.1109/TMI.2012.2206398> PMID: 22759441
55. Bharati MH, Liu JJ, MacGregor JF. Image Texture Analysis: Methods and Comparisons. *Chemometric and Intelligent Laboratory System.* 2004; 72: 57–71. <https://doi.org/10.1016/j.chemolab.2004.02.005>.
56. Galloway MM. Texture Analysis using Gray Level Run Lengths. *Computer Graphics and Image Processing.* 1975; 4: 172–179.
57. Tang X. Texture Information in Run-length Matrices. *IEEE Trans on Image Processing.* 1998; 7: 1602–1609. <https://doi.org/10.1109/83.725367> PMID: 18276225
58. Chu A, Sehgal CM, Greenleaf JF. Use of Gray Value Distribution of Run-lengths for Texture Analysis. *Pattern Recognition Letters.* 1990; 11: 415–420. [https://doi.org/10.1016/0167-8655\(90\)90112-F](https://doi.org/10.1016/0167-8655(90)90112-F).
59. Dasarathy BV and Holder EB. Image Characterizations Based on Joint Gray Level Run-length Distributions. *Pattern Recognition Letters.* 1991; 12: 172–179. [https://doi.org/10.1016/0167-8655\(91\)80014-2](https://doi.org/10.1016/0167-8655(91)80014-2).
60. <https://pyradiomics.readthedocs.io/en/latest/features.html#radiomics.glrIm.RadiomicsGLRLM>.
61. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Shape and Textural Indexes Application to Cell Nuclei Classification. *Int. Journal of Pattern Recognition and Artificial Intelligence.* 2013; 27: 1–23. <https://doi.org/10.1142/S0218001413570024>.
62. Tadayyon H, Sannachi L, Gangeh MJ, Kim C, Ghandi S, Trudeau M, et al. A Priori Prediction of Neoadjuvant Chemotherapy Response and Survival in Breast Cancer Patients using Quantitative Ultrasound. *Nature Scientific Reports.* 2017; 7:45733. <https://doi.org/10.1038/srep45733>. PMID: 28401902

63. Klimonda Z, Karwat P, Dobruch-Sobczak K, Piotrkowska-Wróblewska H, Litniewski J. Breast-lesions Characterization using Quantitative Ultrasound Features of Peritumoural Tissue. *Scientific Reports*. 2019; 9: 7963: 1–9. <https://doi.org/10.1038/s41598-019-44376-z>.
64. Duric N, Littrup P, Babkin A, Chambers D, Azevedo S, Kalinin A, et al. Development of Ultrasound Tomography for Breast Imaging: Technical Assessment. *Med. Phys.* 2005; 32: 1375–1386. <https://doi.org/10.1118/1.1897463>. PMID: 15984689
65. Berger G, Laugier P, Thalabard JC, Perrin J. Global Breast Attenuation: Control Group and Benign Breast Diseases. *Ultrason. Imaging*. 1990; 12: 47–57. <https://doi.org/10.1177/016173469001200104> PMID: 2184568
66. Labyed Y, Bigelow TA, McFarlin BA. Estimate of the Attenuation Coefficient using a Clinical Array Transducer for the Detection of Cervical Ripening in Human Pregnancy. *Ultrasonics*. 2011; 51: 34–39. <https://doi.org/10.1016/j.ultras.2010.05.005>. PMID: 20570308
67. Insana M and Hall TJ. Parametric Ultrasound Imaging from Backscatter Coefficient Measurements: Image Formation and Interpretation. *Ultrason. Imaging*. 1990; 12: 245–267. [https://doi.org/10.1016/0161-7346\(90\)90002-F](https://doi.org/10.1016/0161-7346(90)90002-F).
68. Park SH and Han K. Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiology*. 2018; 286: 800–809. <https://doi.org/10.1148/radiol.2017171920>. PMID: 29309734
69. Bishop CM. *Pattern Recognition and Machine Learning*. Springer: New York, 2006.
70. Cortes C and Vapnik V. Support-vector Networks. *Machine Learning*. 1995; 20: 273–297. <https://doi.org/10.1007/BF00994018>.
71. Roses D. *Breast Cancer*. 2nd Ed. Elsevier; 2005.
72. Lakhani SR, Ellis IO, Schnitt SJ, Tan PH, van de Vijver MJ. *WHO Classification of Tumours of the Breast*. 4th Ed. International Agency for Research on Cancer; 2012.
73. Tsui PH, Yeh CK, Liao YY, Chang CC, Kuo WH, Chang KJ, et al. Ultrasound Nakagami Imaging: A Strategy to Visualize the Scatterers Properties of Benign and Malignant Breast Tumours. *Ultrasound Med. Biol.* 2010; 36: 209–217 <https://doi.org/10.1016/j.ultrasmedbio.2009.10.006>. PMID: 20018436
74. Destrepes F, Trop I, Allard L, Chayer B, Garcia-Duitama J, El Khoury M, et al. Added Value of Quantitative Ultrasound and Machine Learning in BI-RADS 4–5 Assessment of Solid Breast Lesions. *Ultrasound in Medicine & Biology*. 2020; 46: 436–444. <https://doi.org/10.1016/j.ultrasmedbio.2019.10.024>. PMID: 31785840
75. Dobruch-Sobczak K, Piotrkowska-Wróblewska H, Roszkowska-Purska K, Nowicki A, & Jakubowski W. Usefulness of combined BI-RADS analysis and Nakagami statistics of ultrasound echoes in the diagnosis of breast lesions. *Clinical radiology*. 2017; 72: 339.e7–339.e15. <https://doi.org/10.1016/j.crad.2016.11.009>.