



ELSEVIER

<https://doi.org/10.1016/j.ultrasmedbio.2019.10.024>

● *Original Contribution*

ADDED VALUE OF QUANTITATIVE ULTRASOUND AND MACHINE LEARNING IN BI-RADS 4–5 ASSESSMENT OF SOLID BREAST LESIONS

FRANÇOIS DESTREMPES,* ISABELLE TROP,^{†,‡} LOUISE ALLARD,* BORIS CHAYER,*
 JULIAN GARCIA-DUITAMA,* MONA EL KHOURY,^{†,‡} LUCIE LALONDE,^{†,‡} and GUY CLOUTIER*^{†,§}

* Laboratory of Biorheology and Medical Ultrasonics, University of Montreal Hospital Research Center (CRCHUM), Montréal, Québec, Canada; [†] Department of Radiology, Breast Imaging Center, University of Montreal Hospital (CHUM), Montréal, Québec, Canada; [‡] Department of Radiology, Radio-Oncology and Nuclear Medicine, University of Montreal, Montréal, Québec, Canada; and [§] Institute of Biomedical Engineering, University of Montreal, Montréal, Québec, Canada

(Received 24 May 2019; revised 17 September 2019; in final from 25 October 2019)

Abstract—The purpose of this study was to evaluate various combinations of 13 features based on shear wave elasticity (SWE), statistical and spectral backscatter properties of tissues, along with the Breast Imaging Reporting and Data System (BI-RADS), for classification of solid breast lesions at ultrasonography by means of random forests. One hundred and three women with 103 suspicious solid breast lesions (BI-RADS categories 4–5) were enrolled. Before biopsy, additional SWE images and a cine sequence of ultrasound images were obtained. The contours of lesions were delineated, and parametric maps of the homodyned-*K* distribution were computed on three regions: intra-tumoral, supra-tumoral and infra-tumoral zones. Maximum elasticity and total attenuation coefficient were also extracted. Random forests yielded receiver operating characteristic (ROC) curves for various combinations of features. Adding BI-RADS category improved the classification performance of other features. The best result was an area under the ROC curve of 0.97, with 75.9% specificity at 98% sensitivity. (E-mail: guy.cloutier@umontreal.ca) © 2019 World Federation for Ultrasound in Medicine & Biology. All rights reserved.

Key Words: Ultrasonography, Ultrasound imaging, Breast tumors, Elasticity imaging techniques, Machine learning.

INTRODUCTION

Mammography has long been established as an efficient screening method for breast cancer but its performance in classifying abnormalities is limited, requiring additional tests for lesion characterization. Once a suspicious lesion is detected, additional ultrasound (US) and/or mammographic views are obtained to classify lesions into one of five Breast Imaging Reporting and Data System (BI-RADS) categories (Madjar and Mendelson 2008). By adding US for analysis of mammographically identified breast lesions, Zonderland et al. (1999) obtained a combined sensitivity of 91% and specificity of 98% for breast cancer detection on a database comprising 70% of breast lesions suspected as benign or probably benign (*i.e.*, BI-RADS

categories 2 and 3). Lesions classified as BI-RADS 4 or 5 (suspicious for malignancy or highly suggestive of malignancy) undergo biopsy, but many biopsies still confirm the absence of cancer (Mitka 2007). Surveillance is acceptable for women in the BI-RADS 3 category (probably benign) because of the very low likelihood of cancer (defined as <2%).

Ultrasound is an excellent modality for lesion evaluation and follow-up considering availability, cost and absence of radiation, but improvements in specificity are still important targets to reach. Adding US shear wave elasticity (SWE) to standard B-mode imaging features has shown promise in improving classification of lesions (Athanasίου et al. 2010). A recent meta-analysis by Xue et al. (2017) of 25 articles revealed a sensitivity and specificity for SWE overall of 0.88 (0.84–0.91) and 0.87 (0.84–0.89), respectively. In particular, combining SWE parameters with BI-RADS assessment yielded an improvement on lesion classification, thus potentially avoiding unnecessary biopsies (Berg et al. 2012; Evans

Address correspondence to: Guy Cloutier, Laboratory of Biorheology and Medical Ultrasonics, Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), 900 St-Denis, Suite R11.720, Montréal, Québec, Canada, H2X 0A9. E-mail: guy.cloutier@umontreal.ca

et al. 2012a; Au *et al.* 2014; Klotz *et al.* 2014). These results did not only consider suspicious for malignancy or highly suggestive of malignancy lesions (*i.e.*, BI-RADS 4 and 5 lesions).

Quantitative US (QUS) has also been shown to be of interest for breast lesion characterization. One family of QUS parameters is based on statistical analysis of backscatter US echoes. The spatial organization of cells—randomly positioned or spatially organized, their number per unit volume, the contrast between the compressibility and mass density of their nuclei and those of the ambient cytoplasm—are factors that leave a signature on the distribution of the US echo envelope (Destrempe and Cloutier 2010, 2013). In a study on 98 solid breast lesions (Trop *et al.* 2015), one of four women with a later confirmed benign lesion could have avoided biopsy, without any false negatives, based on homodyned- K modeling of the echo envelope (Destrempe and Cloutier 2010, 2013). Dobruch-Sobczak *et al.* (2017) subsequently reported that classification based on BI-RADS grade alone improved substantially when combined with features based on Nakagami imaging, another statistical modeling of US echoes.

Another family of QUS parameters is based on spectral analysis of US echoes. Among these parameters, the total attenuation coefficient has been studied to measure losses in US energy caused by absorption in tissues along propagation (Bigelow and Labyed 2013). As found in a retrospective study (spanning 1998–2010) on 509 invasive lobular carcinomas by Jones *et al.* (2013), attenuation effects resulting in posterior shadowing are common in malignant breast lesions.

In this study, we postulated that SWE, homodyned- K analysis and attenuation estimation provide complementary assessment of tumors, the first paradigm based on mechanical properties of breast lesions and their surrounding tissues, the second on their cellular characteristics and the third on losses of signal energy caused by intervening tissues. Specifically, the goal was to implement this combined SWE–QUS strategy with machine learning random forest classifiers using as input elastography and statistical and spectral backscatter features of tissues, together with the usual BI-RADS category, for cancer detection in suspicious-for-malignancy or highly-suggestive-of-malignancy solid breast lesions at US.

METHODS

This study was approved by our institutional review board, and all patients gave written informed consent.

Patient selection

All consecutive women evaluated at the breast imaging center and for whom a recommendation for biopsy of

a solid mass (BI-RADS 4 or 5 lesion) was made were invited to participate in the study. Women with implants, breastfeeding women and women who had undergone prior surgery in the same breast quadrant as the lesion of interest or radiotherapy were excluded; prior biopsy in the same breast was not a criterion for exclusion. When multiple biopsies were recommended for one patient, only one lesion was selected for inclusion into this study, chosen at the discretion of the radiologist. Radiologists favored masses over other types of lesion morphologies. If multiple masses were present, the most suspicious one (based on BI-RADS assessment) was selected. When lesions of similar BI-RADS assessment were present, the larger one was selected for study inclusion.

Information on menopausal status, hormone replacement therapy and family history of breast or ovarian cancer was obtained directly from the patient. Breast density was assessed subjectively from digital 2-D mammography and classified as per BI-RADS into one of four categories: type A (almost entirely fatty), type B (scattered areas of fibroglandular density), type C (heterogeneously dense) and type D (extremely dense). Lesion depth was characterized on US as superficial third, middle third or posterior third, by dividing breast thickness at the level of the lesion of interest in equal thirds.

US imaging and histopathology

Ultrasound evaluation of breast lesions was performed as per standard practice, with a high-frequency linear transducer, either on a GE Logiq 9 system (General Electric Healthcare, Chicago, IL, USA) or a Toshiba Aplio 500 (Canon Medical Systems, Tustin, CA, USA), by one of three fellowship-trained breast imagers (15–24 y of experience). Each lesion was assigned a BI-RADS category assessment based on integration of clinical, mammographic and sonographic features.

For women who gave written consent, three SWE images were immediately obtained with an Aixplorer US system (Supersonic Imagine, Aix-en-Provence, France) using a SL15-4 probe. For each lesion, SWE parameters were adjusted to obtain an elastogram that covered the lesion entirely and the probe was kept immobile for 3–5 s to reach stability of elastograms. Then, for the same lesion, a 1-s cine loop (100 frames/s) of radiofrequency (RF) images was obtained under B-mode guidance with the same system and probe, by changing the acquisition mode. A fixed setting was used for all acquisitions: nominal center frequency = 7.0 MHz, focused at 2 cm, bandwidth = 68%, depth = 4 cm, lateral and axial discretizations = 0.20 and 0.00639 mm/pixel, respectively.

After acquisition, the radiologist performed percutaneous biopsy as per standard procedure, under local anesthesia, with 14G or 16G automatic needles, retrieving four to eight samples per lesion, at the radiologist's

discretion. Breast pathologists performed histopathologic analyses. Final diagnoses were categorized as benign or malignant. Malignancy was defined as identification of an infiltrating carcinoma or ductal carcinoma *in situ*; all other diagnoses were considered benign. The label “fibrocystic changes” was applied to all lesions with pathologic diagnoses that included sclerosing adenosis, apocrine metaplasia, ductal hyperplasia and sclerocystic changes. For women in whom a high-risk diagnosis was obtained at percutaneous biopsy, the final diagnosis recorded was the one based on surgical pathology. In the absence of surgery, stability at imaging was considered indicative of no associated malignancy.

Patient management decisions were made without consideration of the data obtained from elastography or QUS.

Lesion analysis

The data acquired with the Aixplorer scanner were transferred to a workstation for analysis; lesions were anonymized so that neither patient identity nor lesion BI-RADS classification was revealed.

Uncompressed B-mode images were computed as the echo envelope of RF data (Kallel et al. 1994) as in Destrempes et al. (2013), without application of filters or log compression except for display. The contour of each lesion was manually delineated (segmented) on the first image of the cine loop of RF data and on the first of three corresponding B-mode images matched with SWE maps by a breast radiologist, with an in-house software written in MATLAB (2010a, The MathWorks, Natick, MA, USA) that required little training for using.

Contours were then propagated along remaining images using an algorithm that compensates for motion (Destrempes et al. 2011). In addition to the segmented intra-tumoral zone, a 3-mm-thick supra-tumoral zone above the lesion was defined to take into account spiculations and formation of an echogenic rim in peritumoral tissues caused by reaction of surrounding parenchyma to malignant cells. A 5-mm-thick infra-tumoral zone was also defined to capture posterior features, if present, associated with the lesion. Supra- and infra-tumoral zones were automatically computed from the intra-tumoral zone segmentation.

The median (over three acquisitions) of the maximum elasticity E_{\max} on each of these three regions of interest (ROIs) was computed from SWE images directly obtained from the Aixplorer system, for a total of three elastographic features (one per ROI). For each of the three ROIs on B-mode images, a local sliding estimation window was swept across the ROI, and only pixels with the same statistical properties as the center pixel were considered for estimation, as described in Destrempes et al. (2016). At each location of the window, four

parameters with a physical interpretation were derived from the postulated homodyned K -distribution model of the echo envelope (Destrempes et al. 2013), from which the following were retained based on Trop et al. (2015): (i) infra-tumoral total signal power μ_n normalized by the maximal intensity in the ROI (Trop et al. 2015); (ii) intra-tumoral reciprocal $1/\alpha$ of the scattering clustering parameter α (Dutt and Greenleaf 1994; Destrempes et al. 2013); (iii) supra-tumoral coherent-to-diffuse signal ratio k (Dutt and Greenleaf 1994; Destrempes et al. 2013); and (iv) supra-tumoral diffuse-to-total signal power ratio $1/(\kappa + 1)$ (Destrempes et al. 2016). This process yielded four homodyned- K parametric maps. The mean value and interquartile range (IQR) of each of these maps were extracted as QUS features, for a total of eight biomarkers. Note that any other statistic (e.g., median, standard deviation, skewness) could have been considered as features.

The total attenuation coefficient's slope (under linear dependency with respect to frequency) α_{att} (dB/MHz/cm) on intra- and infra-tumoral zones was estimated based on the spectral fit algorithm (Bigelow and O'Brien 2005) that requires use of a RMI 403 GS reference phantom (Gammex, Middleton, WI, USA), thus yielding two additional QUS features. The acquisition settings were fixed throughout the study, and the acquisition on the reference phantom was performed after data acquisition using the same settings.

In addition to these 13 quantitative features, the BI-RADS category, as assessed and recorded by a breast radiologist, was considered as a clinical feature for classification. Lesion analyses were carried out using computer programs in C++, and programs written with MATLAB. Manual segmentations were based on a program written with MATLAB. Examples of segmented zones and quantitative parametric maps are provided in Figures 1–4.

Data analysis based on machine learning

Random forests (Breiman 2001) were used as classifier model as they are liable to avoid overfitting in the case of relatively small data, as opposed to neural networks. A random forest consists of independent decision trees (1000 in this work, except for feature selection) from which classification is reached by considering the most frequent decision among all decisions output by each tree. Note that, based on Theorem 1.2 of Breiman (2001), addition of more trees does not cause overfitting in random forest classifiers. Classification efficiency of various combinations of features was assessed by estimating the area under the receiver operating characteristic (ROC) curve (AUC). A bootstrap cross-validation method was adopted to estimate true- and false-positive rates from which ROC curves were computed.

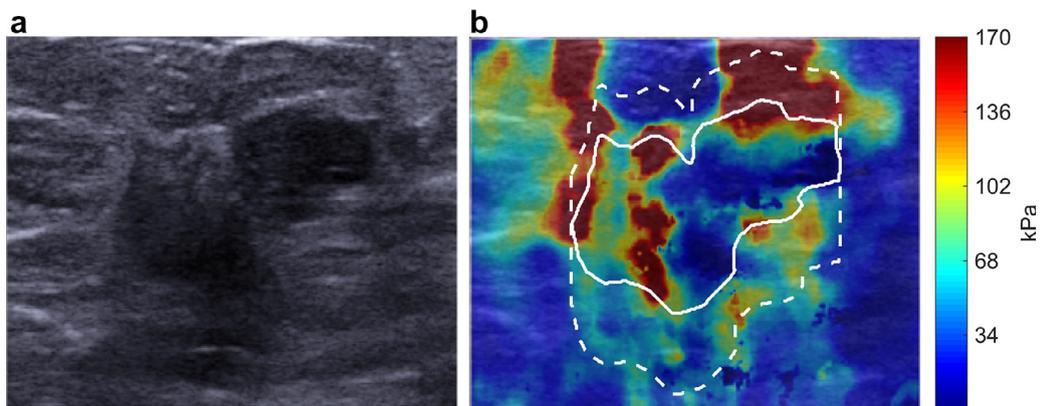


Fig. 1. Example of a malignant lesion. A 71-y-old woman without prior breast problems presented with a new palpable mass in the left upper outer breast. Mammography and ultrasound revealed a 2.5-cm irregular highly suspicious mass, classified as BI-RADS category 5. Biopsy confirmed a luminal A (ER+PR+Her2-) grade 2 infiltrating ductal carcinoma. (a) Echograph B-mode image within Q-box. (b) Shear wave elasticity (SWE) map (kPa) displayed within Q-box together with contours delineated on the corresponding echograph B-mode image. BI-RADS = Breast Imaging Reporting and Data System.

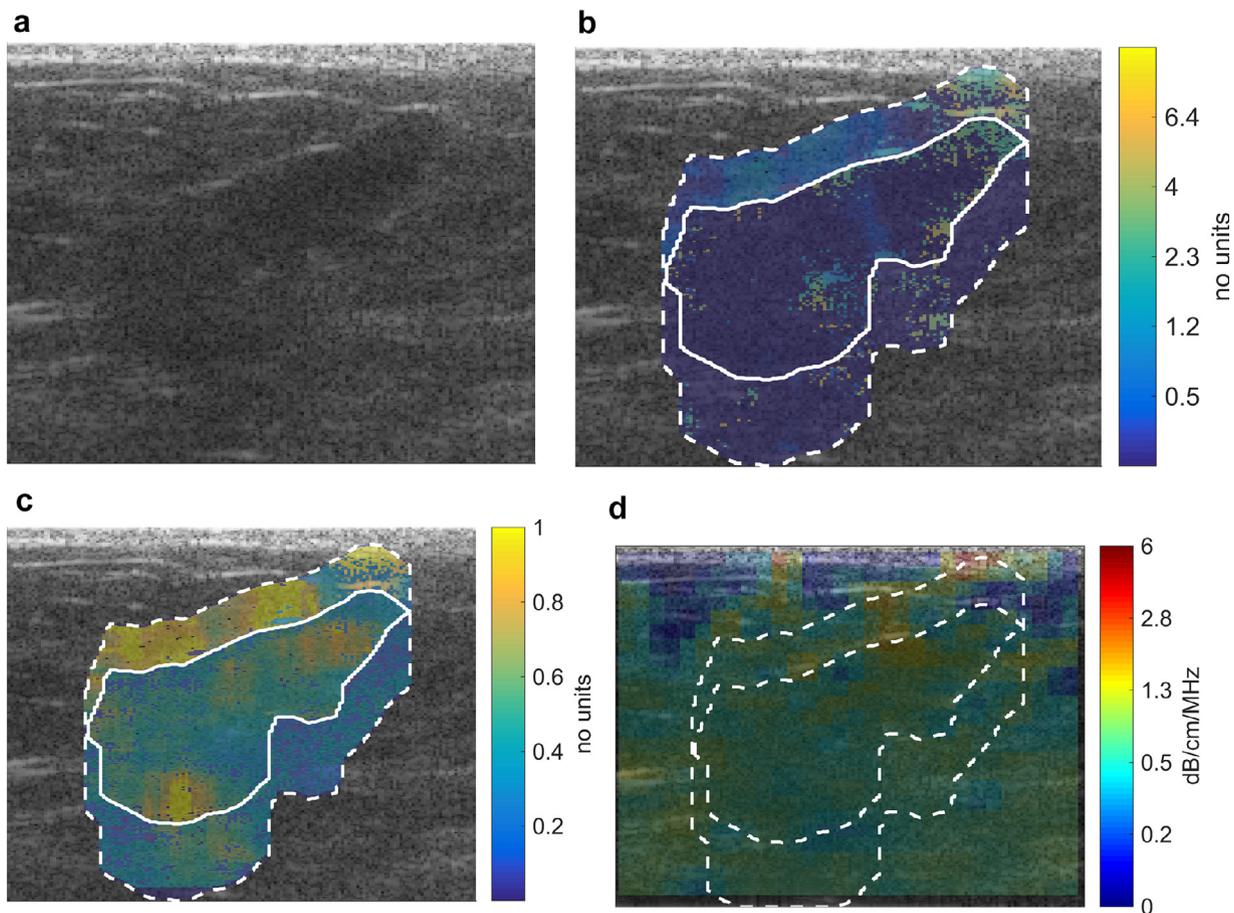


Fig. 2. Example of a malignant lesion (same lesion as in Fig. 1). (a) Echo envelope of radiofrequency (RF) data within 5 mm of the lesion. (b) Homodyned- K reciprocal of the scatterer clustering parameter $1/\alpha$ (no units)—in log scale for display—is superimposed on B-mode (echo envelope of RF data) together with contours delineated on the echo envelope. (c) Homodyned- K diffuse-to-total signal power ratio $1/(\kappa + 1)$ (no units) is superimposed on B-mode (echo envelope of RF data). (d) Total attenuation coefficient's slope map α_{att} (dB/cm/MHz)—in log scale for display—is superimposed on B-mode (echo envelope of RF data).

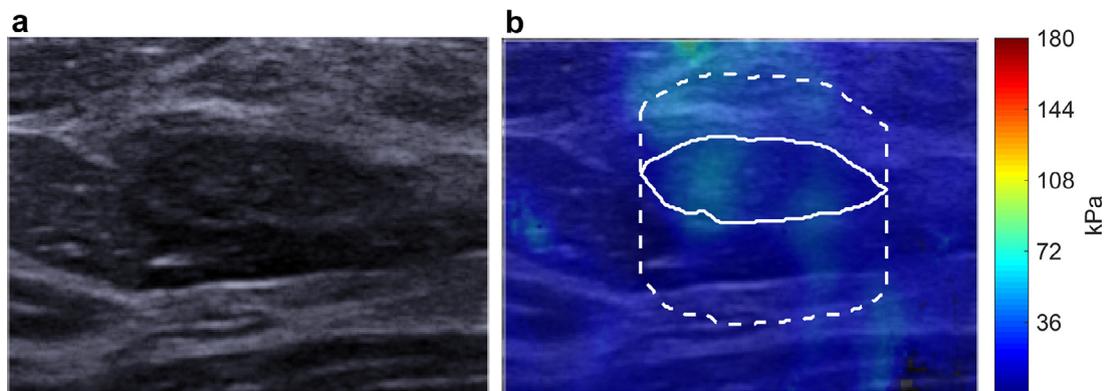


Fig. 3. Example of a benign lesion. A 44 y-old woman, whose mother had breast cancer at 65, presented with a new lesion in the upper inner left breast detected at mammography. Ultrasound characterization revealed a 16-mm, oval, well-circumscribed mass, classified as BI-RADS category 4A. Biopsy confirmed a benign fibroadenoma. (a) Echograph B-mode image within Q-box. (b) A shear wave elasticity map (kPa) is displayed within the Q-box together with contours delineated on the corresponding echograph B-mode image. BI-RADS = Breast Imaging Reporting and Data System.

In view of the heavy computational load inherent to bootstrap methods, and because the number of combinations of features grows exponentially as a function of the number of features, a rapid feature selection procedure was applied on all combinations of at most four features including the BI-RADS category. For this purpose, a

single random forest comprising 3000 trees was trained on the entire database for each such combination (Genuer et al. 2010). The 40 combinations of features with the highest G -mean values (He and Garcia 2009) and remaining ones with same G -mean were then selected for further assessment.

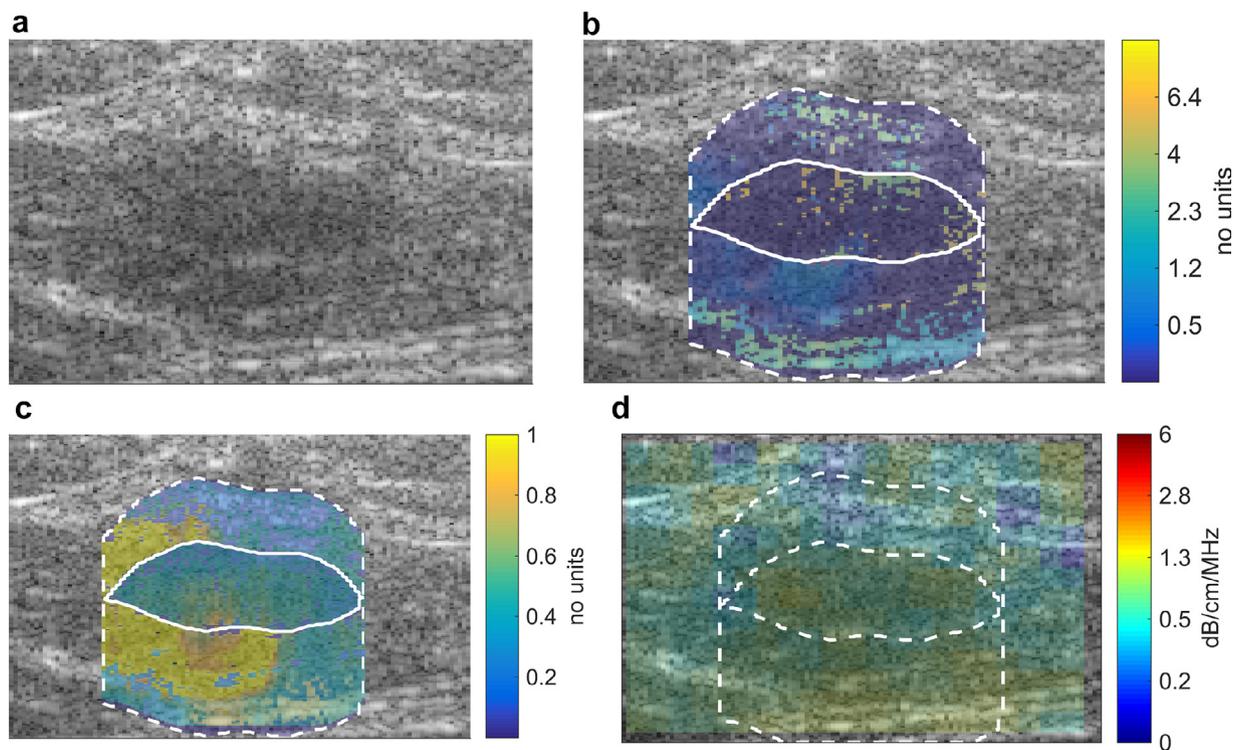


Fig. 4. Example of a benign lesion (same lesion as in Fig. 3). (a) Echo envelope of RF data within 5 mm of the lesion. (b) Homodyned- K reciprocal of the scatterer clustering parameter $1/\alpha$ (no units)—in log-scale for display—is superimposed on B-mode (echo envelope of RF data) together with contours delineated on the echo envelope. (c) Homodyned- K diffuse-to-total signal power ratio $1/(\kappa + 1)$ (no units) is superimposed on B-mode (echo envelope of RF data). (d) Total attenuation coefficient's slope map α_{att} (dB/cm/MHz)—in log-scale for display—is superimposed on B-mode (echo envelope of RF data).

For each of the selected combinations of features, the 0.632+ bootstrap method (Efron and Tibshirani 1997), with 1000 bootstraps, was applied to estimate the true- and false-positive rates. To obtain a ROC curve, stratified resampling with replacement was performed, with strata based on a proportion of positives varying from 0–1 (in steps of 1 of 40). The number of true and false positives was computed only on data not belonging to a given bootstrap sample.

For each of the 40 selected combinations of features, the area under the ROC curve was estimated, and the combination of features yielding the highest value was considered as the best combination. The jackknife technique (DeLong *et al.* 1988) was then used to provide a sample of 103 AUCs (one AUC per lesion), from which a 95% confidence interval for the AUC corresponding to the best combination of features was estimated based on percentiles.

Thus, combinations of 3 QUS or SWE features among 13 (one per ROI) and the BI-RADS category were assessed. For comparison's sake, other combinations of features were tested: combinations of 4 features (at least 1 per ROI) other than the BI-RADS category and the BI-RADS category alone. We also tested the three SWE features combined with the BI-RADS category, and combinations of 3 QUS features among 10 (one per ROI), along with the BI-RADS category. When testing the BI-RADS category alone, the random forest classifier was used as for the other combinations, albeit with a single feature, and results of biopsy were the gold standard as for the other combinations. These analyses were carried out using R statistical software (R Foundation, Vienna, Austria) and package “RandomForest” Version 4.6-12 for R.

RESULTS

Study population

Between January 2015 and September 2017, 112 women were recruited. There were 9 subsequent exclusions: 4 retrieved consent; 4 acquisitions were incomplete (missing RF or SWE data); and 1 woman was found to have had surgery in the area of interest. The final study population consisted of 103 women for whom 103 lesions were considered. Mean patient age was 53.3 y (range: 15–92 y). Fifty-two women had undergone menopause (50.5%, 52 of 103) and 6 (6 of 52, 11.5%) were taking hormone replacement therapy. Seven women (7 of 103, 6.8%) had been treated for breast cancer in the contralateral breast, one had been treated for ovarian cancer and one had undergone radiotherapy for mediastinal lymphoma. Four women had previously undergone surgery for benign lesions in the same breast but in a different quadrant from the study lesion (from 4–30 y prior) and 3

women had had percutaneous benign biopsies in the ipsilateral breast (from 2 mo to 4 y prior). Fifteen women had first- and second-degree family histories of breast cancer (15 of 103, 14.6%), and one woman's grandmother had been treated for ovarian cancer.

Breast lesions

Thirty-seven of 103 lesions (35.9%) were located in the left breast, and 66 (64.1%) in the right breast. Lesion diameter ranged from 4.4–43.5 mm (mean = 17.1 mm, median = 13.5 mm). Lesion depth within the breast was classified as the deepest third for 27 lesions (26%), the middle third for 56 lesions (54%) and the superficial third for 20 lesions (20%). Mammographic evaluation was available for 98 lesions (95%): parenchymal density was considered fatty in 13 (13%) (type A); 33 (34%) were classified as type B and 37 (38%) as type C; and 15 (15%) were extremely dense (type D). Sixty-three lesions (61.2%) were classified as moderately suspicious (BI-RADS 4): 35 lesions (34.0%) as BI-RADS 4A, 17 (16.5%) as BI-RADS 4B, 11 (10.7%) as BI-RADS 4C. Forty lesions (38.8%) were classified as highly suspicious (BI-RADS 5).

Histopathology analysis confirmed 39 of the 40 BI-RADS 5 lesions (97.5%) as malignant. One of thirty-five BI-RADS 4A lesions (2.9%) was classified as malignant, as were 2 of 17 BI-RADS 4B lesions (11.8%) and 7 of 11 BI-RADS 4C lesions (63.6%). Forty-nine of 103 lesions were malignant (47.6%) (Table 1). All malignant cases were invasive cancers, the majority invasive ductal carcinomas (37 of 49, 75.5%), of which 8 were associated with ductal carcinoma *in situ* (8 of 37, 21.6%). Other invasive cancer subtypes encountered were invasive lobular carcinomas (5 of 49, 10.2%), 3 mucinous carcinomas (3 of 49, 6.1%), 2 papillary carcinomas (4.1%), 1 tubular carcinoma and 1 adenocarcinoma.

Fifty-four non-cancerous lesions were diagnosed. The majority were fibroadenomas (34 of 54, 63.0%); 6 corresponded to fibrocystic changes (6 of 54, 11.1%); 5 were papillomas, one of which harbored atypia (5 of 54; 9.3%); and 5 were fibrosis (9.3%). The remaining lesions corresponded to one radial scar, one fat necrosis, one

Table 1. Demographic distribution of BI-RADS categories and malignancy within the study population

BI-RADS assessment	Total No. of lesions	Malignant	Benign
4A	35 (34.0%)	1 (2.9%)	34 (97.1%)
4B	17 (16.5%)	2 (11.8%)	15 (88.2%)
4C	11 (10.7%)	7 (63.6%)	4 (36.4%)
5	40 (38.8%)	39 (97.5%)	1 (2.5%)
Total	103	49 (47.6%)	54 (52.4%)

BI-RADS = Breast Imaging Reporting and Data System.

pseudoangiomatous stromal hyperplasia and one borderline phyllodes tumor.

Machine learning model analysis performance

The area under the ROC curve (AUC) for each type of combinations of features can be found in Table 2. On BI-RADS 4 and 5 category lesions, the highest AUC obtained by considering all types of features that is, 0.97 (0.968–0.972), was obtained with the supra-tumoral $1/(\kappa + 1)$ and intra- and infra-tumoral α_{att} in addition to the BI-RADS category. These features yielded on the ROC curve a specificity of 84.8%, 75.9% or 34.6% at a sensitivity of 97%, 98% or 99%, respectively. When retaining the 40 combinations with the best *G*-means among QUS features together with the BI-RADS category (at the feature selection step), an additional combination was found to yield an AUC of 0.97: the supra-tumoral *k* IQR and intra- and infra-tumoral α_{att} along with the BI-RADS category. This combination of features yielded a specificity of 81.5%, 68.5% or 37.4% at a sensitivity of 97%, 98% or 99%, respectively. Combining the 3 SWE features with BI-RADS category yielded an AUC of 0.96, with a specificity of 57.7%, 43.8% or 21.9% at a sensitivity of 97%, 98% or 99%, respectively. In comparison, considering all features except BI-RADS category yielded an AUC of 0.74 for the best combination consisting of supra-tumoral $1/(\kappa + 1)$ IQR, intra-tumoral E_{max} , intra- and infra-tumoral α_{att} , which gave a specificity of 18.5%, 14.1% or 0.08% for a sensitivity of 97%, 98% or 99%, respectively. In turn, BI-RADS assessment alone provided an AUC of 0.95, with a specificity of 47.5%, 21.6% or 10.8% at a sensitivity of 97%, 98% or 99%, respectively. Corresponding ROC curves are provided in Figure 5.

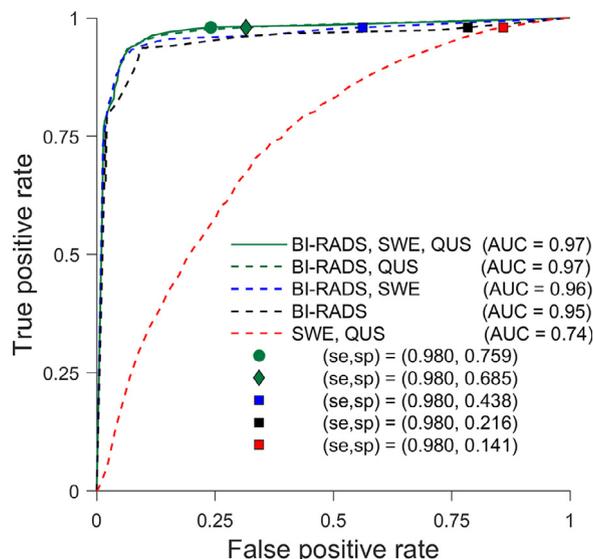


Fig. 5. Receiver operating characteristic (ROC) curves. ROC curves were obtained from the best combination of at most four features chosen from the different types of features: BI-RADS category, shear wave elasticity (SWE), and quantitative ultrasound (QUS), comprising features extracted from homodyned-*K* and total attenuation coefficient’s slope maps, as tested on lesions in BI-RADS categories 4 and 5. The interpolated specificity (1 – false positive rate) corresponding to a sensitivity of 98% is also presented on each curve. BI-RADS = Breast Imaging Reporting and Data System.

DISCUSSION

We propose an ancillary tool to sonographic characterization of breast lesions before biopsy, based on quantitative measurements and estimation, to improve clinical assessment. The model is based on random forests classifiers and

Table 2. Areas under the ROC curves for various combinations of features, as estimated with the 0.632+ bootstrap method*

Type of features	Best found combination of features	AUC [95% CI] BI-RADS 4–5	Specificity [†]			Youden’s index
			se = 97%	se = 98%	se = 99%	
BI-RADS category	BI-RADS category	0.95 [0.947–0.952]	47.5%	21.6%	10.8%	0.84
SWE, QUS	Supra: $1/(\kappa + 1)$ IQR, Intra: E_{max} , Infra: α_{att}	0.74 [0.727–0.755]	18.5%	14.1%	0.08%	0.37
BI-RADS category, SWE	BI-RADS category, Supra: E_{max} , Intra: E_{max} , Infra: E_{max}	0.96 [0.959–0.967]	57.7%	43.8%	21.9%	0.86
BI-RADS category, QUS	BI-RADS category, Supra: <i>k</i> IQR, Intra: α_{att} , Infra: α_{att}	0.97 [0.969–0.970]	81.5%	68.5%	37.4%	0.87
BI-RADS category, SWE, QUS	BI-RADS category, Supra: $1/(\kappa + 1)$, Intra: α_{att} , Infra: α_{att}	0.97 [0.968–0.972]	84.8%	75.9%	34.6%	0.87

AUC = area under the receiver operating curve; CI = confidence interval; ROC = receiver operating characteristic curve; BI-RADS = Breast Imaging Reporting and Data System; QUS = quantitative ultrasound; SWE = shear wave elasticity; IQR = interquartile range; E_{max} = maximum elasticity; Supra = supratumoral; Intra = intratumoral; Infra = infratumoral; se = sensitivity.

Specificity and Youden’s index values in bold are maximum values.

* BI-RADS category; SWE; E_{max} ; QUS, including mean and IQR of homodyned-*K* parametric maps μ_m , $1/\alpha$, *k* and $1/(\kappa + 1)$; and total attenuation coefficient’s slope α_{att} . Youden’s index (Schisterman et al. 2005) is also reported for each ROC curve.

† The specificity of interpolated points on the ROC curves corresponding to a high sensitivity (se) are reported.

uses, in addition to the BI-RADS category, elastography features; homodyned- K features estimated from uncompress, unfiltered echo envelopes of RF signals; and the total attenuation coefficient's slope estimated with spectral analysis of RF signals. When considering all types of features for 103 lesions, the AUC was 0.97, and a specificity of 55.9% was reached (representing 30 of 54 lesions) at 98.4% sensitivity.

In the context of US characterization before biopsy, it is aimed at determining the portion of ROC curves where sensitivity is close to 100%. Under this condition, BI-RADS category taken as a single feature performed poorly (sensitivity = 97.5%, specificity = 27.5%), and quantitative features alone performed even more poorly (sensitivity = 97.6%, specificity = 16.3%). Nonetheless, combination of BI-RADS category with elasticity and QUS features raised the performance substantially, reaching a specificity of 55.9% at a sensitivity of 98.4%. These findings suggest that BI-RADS assessment and quantitative features would benefit of being used jointly for improvement of breast classification efficiency to avoid unnecessary biopsies. This is consistent with table 3 of [Dobruch-Sobczak et al. \(2017\)](#).

Results obtained can be explained from a biological perspective. Malignant masses typically feature spiculated margins, yielding an expected decrease in parameter $1/(\kappa + 1)$, which is equivalent to the finding of [Trop et al. \(2015\)](#); and posterior shadowing, resulting in an expected higher difference between infra- and intra-tumoral α_{att} (because of signal propagation in the posterior zone), which is consistent with [Jones et al. \(2013\)](#). Indeed, these trends were observed: mean values of 0.80 (malignant lesions) versus 0.86 (benign lesions) for $1/(\kappa + 1)$ (no units), and 0.066 versus 0.049 for infra α_{att} (dB/cm/MHz) minus intra α_{att} .

No SWE feature was retained in the best combination of biomarkers, but it is worth noting that when considering SWE and QUS alone, the intra-tumoral maximum elasticity (E_{max}) was retained in the best selected combination of features. Note that in [Berg et al. \(2012\)](#) and [Evans et al. \(2012b\)](#), E_{max} or E_{mean} was computed on the lesion and adjacent tissues presenting stromal stiffness, which justified considering E_{max} in the tumoral and two peritumoral zones. The relevance of elasticity in shells surrounding lesions was also studied by [Huang et al. \(2019\)](#).

As per meta-analysis by [Xue et al. \(2017\)](#), the overall AUC under the ROC curve was 0.93 (0.90–0.95) with SWE parameters. In contrast, the proposed SWE–QUS features combined with BI-RADS grade yielded an AUC of 0.97 (0.968–0.972). Note that some studies included BI-RADS 2 or 3 category lesions in patients' recruitment (*e.g.*, [Zonderland et al. 1999](#); [Berg et al. 2012](#); [Dobruch-Sobczak et al. 2017](#)), thus increasing the overall specificity because of a lower rate of

malignancy. See [Davis and Goadrich \(2006\)](#) on this issue. Recently, machine learning classification strategies based on SWE features were proposed for breast cancer diagnosis ([Zhang et al. 2016](#); [Zhou et al. 2018](#)). Comparison of performance with these studies is difficult because BI-RADS categories of selected women were not specified.

Limitations

Possibilities of subanalyses were limited in this work because of the small sample size of some subcategories. For instance, it would have been useful to restrict subanalyses on BI-RADS categories 4A and 4B, those most likely to benefit from improved lesion characterization after which recommendation for biopsy may be modified to short-term surveillance only, that is downgraded to BI-RADS 3 category, but there were only 3 malignant lesions in these two categories. For the same reason, subanalyses on malignant lesions (*e.g.*, invasive ductal carcinomas vs. invasive lobular carcinomas) were not performed. Thus, studies on representative databases of larger size would have to be performed to further assess and refine the findings of this study.

Based on the meta-analysis by [Xue et al. \(2017\)](#) and the multicentric study performed by [Berg et al. \(2012\)](#), we restricted our choice of SWE features to the maximum elasticity within three zones of interest. Likewise, QUS features were restricted to a few possible ones, based on our previous study ([Trop et al. 2015](#)) and work by [Jones et al. \(2013\)](#). It would be worthwhile pursuing breast lesion classification based on further features, including other SWE features based on the full elasticity map, as was done in [Zhang et al. \(2016\)](#) and [Zhou et al. \(2018\)](#), albeit without including BI-RADS category as a feature.

In this work, classification of breast lesions required the pre-processing steps of localization and manual segmentation of lesions. It is conceivable that these two steps could be performed automatically in the foreseeable future, possibly using deep learning, which would make the proposed method for SWE and QUS analyses more convenient in the context of clinical routine. Note that in our previous study ([Trop et al. 2015](#)), it was found that the performed K -homodyned feature analysis was robust to manually drawn initial lesion contour perturbations within 0.5–1.0 mm.

CONCLUSIONS

The proposed classifier based on BI-RADS category combined with elastography and QUS features performed best on BI-RADS categories 4 and 5. Although it would be expected that further analysis on a substantially larger study population may result in a best combination

of features different from that obtained in the present study, this work illustrates the potential of combining features of distinct type, such as clinical (BI-RADS category), SWE (based on biomechanical properties of biological tissues) and QUS (based on acoustical properties of tissues). The reported analysis with random forests trained classifiers suggests the potential of avoiding biopsy in up to 75.9% of BI-RADS 4 and 5 category lesions later confirmed to be benign, while reaching a sensitivity of 98%.

Acknowledgments—This work was supported by the Cancer Research Society of Canada and the Quebec Breast Research Foundation (Grant 19075), and by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 503381-16). We thank the anonymous reviewers for their helpful comments.

Conflict of interest disclosure—The authors declare no competing interests.

REFERENCES

- Athanasiou A, Tardivon A, Tanter M, Sigal-Zafrani B, Bercoff J, Defieux G, Gennisson JL, Fink M, Neuschwander S. Breast lesions: Quantitative elastography with supersonic shear imaging—Preliminary results. *Radiology* 2010;256:297–303.
- Au FW, Ghai S, Moshonov H, Kahn H, Brennan C, Dua H, Crystal P. Diagnostic performance of quantitative shear wave elastography in the evaluation of solid breast masses: Determination of the most discriminatory parameter. *AJR Am J Roentgenol* 2014;203:W328–W336.
- Berg WA, Cosgrove DO, Doré CJ, Schäfer FKW, Svensson WE, Hookey RJ, Ohlinger R, Mendelson EB, Balu-Maestro C, Locatelli M, Tourasse C, Cavanaugh BC, Juhan V, Stavros AT, Tardivon A, Gay J, Henry JP, Cohen-Bacrie C. Shear-wave elastography improves the specificity of breast US: The BE1 multinational study of 939 masses. *Radiology* 2012;262:435–449.
- Bigelow TA, O'Brien WD, Jr. Evaluation of the spectral fit algorithm as functions of frequency range and $\Delta k a_{\text{eff}}$. *IEEE Trans Ultrason Ferroelectr Freq Control* 2005;52:2003–2010.
- Bigelow TA, Labyed Y. Attenuation compensation and estimation. In: Mamou J, Oelze ML, (eds). *Quantitative ultrasound in soft tissues*. Heidelberg/New York: Springer; 2013. p. 71–93.
- Breiman L. Random forests. *Machine Learn* 2001;45:5–32.
- Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. Pittsburgh, PA. ACM; 2006. p. 233–240.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 1988;44:837–845.
- Destrempes F, Cloutier G. A critical review and uniformized representation of statistical distributions modeling the ultrasound echo envelope. *Ultrasound Med Biol* 2010;36:1037–1051.
- Destrempes F, Cloutier G. Review of envelope statistics models for quantitative ultrasound imaging and tissue characterization. In: Mamou J, Oelze ML, (eds). *Quantitative ultrasound in soft tissues*. Heidelberg/New York: Springer; 2013. p. 219–274.
- Destrempes F, Meunier J, Giroux MF, Soulez G, Cloutier G. Segmentation of plaques in sequences of ultrasonic B-mode images of carotid arteries based on motion estimation and a Bayesian model. *IEEE Trans Biomed Eng* 2011;58:2202–2211.
- Destrempes F, Porée J, Cloutier G. Estimation method of the homodyned K -distribution based on the mean intensity and two log-moments. *SIAM J Imaging Sci* 2013;6:1499–1530.
- Destrempes F, Franceschini E, Yu FTH, Cloutier G. Unifying concepts of statistical and spectral quantitative ultrasound techniques. *IEEE Trans Med Imaging* 2016;35:488–500.
- Dobruć-Sobczak K, Piotrkowska-Wroblewska H, Roszkowska-Purka K, Nowicki A, Jakubowski W. Usefulness of combined BI-RADS analysis and Nakagami statistics of ultrasound echoes in the diagnosis of breast lesions. *Clin Radiol* 2017;72:339.e7–339.e15.
- Dutt V, Greenleaf JF. Ultrasound echo envelope analysis using a homodyned K distribution signal model. *Ultrasonic Imaging* 1994;16:265–287.
- Efron B, Tibshirani R. Improvements on cross-validation: The 632+ bootstrap method. *J Am Stat Assoc* 1997;92:548–560.
- Evans A, Whelehan P, Thomson K, Brauer K, Jordan L, Purdie C, McLean D, Baker L, Vinnicombe S, Thompson A. Differentiating benign from malignant solid breast masses: Value of shear wave elastography according to lesion stiffness combined with greyscale elastogram according to BI-RADS classification. *Br J Cancer* 2012a;107:224–229.
- Evans A, Whelehan P, Thomson K, McLean D, Brauer K, Purdie C, Baker L, Jordan Rauchhaus P, Thompson A. Invasive breast cancer: Relationship between shear-wave elastographic findings and histologic prognostic factors. *Radiology* 2012b;263:673–677.
- Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests. *Pattern Recog Lett* 2010;31:2225–2236.
- He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowledge Data Eng* 2009;21:1263–1284.
- Huang L, Ma M, Du Z, Liu Z, Gong X. Quantitative evaluation of tissues stiffness around lesion by sound touch elastography in the diagnosis of benign and malignant breast lesions. *PLoS One* 2019;14:e0209219943.
- Jones KN, Magut M, Henrichsen TL, Boughey JC, Reynolds C, Glazebrook KN. Pure lobular carcinoma of the breast presenting as a heterogeneous mass: Incidence and imaging characteristics. *AJR Am J Roentgenol* 2013;201:W765–W769.
- Kallel F, Bertrand M, Meunier J. Speckle motion artifact under tissue rotation. *IEEE Trans Ultrason Ferroelectr Freq Control* 1994;41:105–122.
- Klotz T, Boussion V, Kwiatkowski F, Dieu-de Fraissinette V, Bailly-Glatre A, Lemery S, Boyer L. Shear wave elastography contribution in ultrasound diagnosis management of breast lesions. *Diagn Interv Imaging* 2014;95:813–824.
- Madjar H, Mendelson EB. *The practice of breast ultrasound: Techniques, findings, differential diagnosis*. 2nd ed. Stuttgart/New York: Thieme; 2008.
- Mitka M. New ultrasound “elasticity” technique may reduce need for breast biopsies. *JAMA* 2007;297:455.
- Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden index to discriminate individuals using pooled blood samples. *Epidemiology* 2005;16:73–81.
- Trop I, Destrempes F, El Khoury M, Robidoux A, Gaboury L, Allard L, Chayer B, Cloutier G. The added value of statistical modeling of backscatter properties in the management of breast lesions at ultrasound. *Radiology* 2015;275:666–674.
- Xue Y, Yao S, Li X, Zhang H. Benign and malignant breast lesions identification through the values derived from shear wave elastography: Evidence for the meta-analysis. *Oncotarget* 2017;8:89172–89181.
- Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, Zheng H. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics* 2016;72:150–157.
- Zhou Y, Xu J, Liu Q, Li C, Liu Z, Wang M, Zheng H, Wang S. A radiomics approach with CNN for shear-wave elastography breast tumor classification. *IEEE Trans Biomed Eng* 2018;65:1935–1942.
- Zonderland HM, Coerkamp EG, van de Vijver MJ, van Voorthuisen AD. Diagnosis of breast cancer: Contribution of US as an adjunct to mammography. *Radiology* 1999;213:413–422.