Original Investigation

Linear Regression Modeling Based Scoring System to Reduce Benign Breast Biopsies Using Multi-parametric US with Color Doppler and SWE

Burcu Özdemir Demirci M.D., Onur Buğdaycı M.D., Assist. Prof., Gökhan Ertaş, Assoc. Prof., Deniz E.T. Şanlı M.D., Assoc. Prof., Handan Kaya M.D., Prof., Erkin Arıbal M.D., Prof.

Rationale and Objectives: To develop a simple ultrasound (US) based scoring system to reduce benign breast biopsies.

Materials and Methods: Women with BI-RADS 4 or 5 breast lesions underwent shear-wave elastography (SWE) imaging before biopsy. Standard US and color Doppler US (CDUS) parameters were recorded, and the size ratio (Sz_R=longest/shortest diameter) was calculated. Measured/calculated SWE parameters were minimum (SWV_{Min}) and maximum (SWV_{Max}) shear velocity, velocity heterogeneity (SWV_H=SWV_{Max}-SWV_{Min}), velocity ratio (SWV_R=SWV_{Min}/SWV_{Max}), and normalized SWV_R (SWV_{Rn}=(SWV_{Max}-SWV_{Min})/SWV_{Min}). Linear regression analysis was performed by converting continuous parameters into categorical corresponding equivalents using decision tree analyses. Linear regression models were fitted using stepwise regression analysis and optimal coefficients for the predictors in the models were determined. A scoring model was devised from the results and validated using a different data set from another center consisting of 187 cases with BI-RADS 3, 4, and 5 lesions.

Results: A total of 418 lesions (238 benign, 180 malignant) were analyzed. US and CDUS parameters exhibited poor (AUC=0.592-0.696), SWE parameters exhibited poor-good (AUC=0.607-0.816) diagnostic performance in benign/malignant discrimination. Linear regression models of US+CDUS and US+SWE parameters revealed an AUC of 0.819 and 0.882, respectively. The developed scoring system could have avoided biopsy in 37.8% of benign lesions while missing 1.1% of malignant lesions. The scoring system was validated with a 100% NPV rate with a specificity of 74.6%.

Conclusion: The linear regression model using US+SWE parameters performed better than any single parameter alone. The developed scoring method could lead to a significant decrease in benign biopsies.

Key Words: Breast Neoplasms; Decision Trees; Ultrasonography, Mammary; Elasticity Imaging Techniques.

Abbreviations: **Sz_R** lesion size ratio, **SWV_{Min}** minimum shear velocity, **SWV_{Max}** maximum shear velocity, **SWV_H** shear velocity heterogeneity, **SWV_R** velocity ratio, **SWV_{Rn}** normalized velocity ratio, **SS_{US}** US based scoring system, **$SS_{US\ SWE}$** US and SWE based scoring system

INTRODUCTION

Compared to mammography, ultrasound (US) has similar cancer detection rates but with higher false positivity, leading to higher biopsy rates (1). Since malignant masses are usually stiffer than benign masses and normal breast tissue, complementary ultrasound elastography (UE) has been extensively studied, aiming to improve the diagnostic performance of breast US (2). Promising results led to the inclusion of UE into the fifth edition of the BI-RADS US lexicon, and practice guidelines of different societies (3–5).

Among the UE techniques, shear-wave elastography (SWE) has gained interest in improving the differentiation

Acad Radiol 2023; ■:1–11

From the Department of Radiology, Marmara University, Istanbul, Türkiye (B.Ö.D., O.B., E.A.); Department of Biomedical Engineering, Yeditepe University, Istanbul, Türkiye (G.E.); Department of Pathology, Marmara University, Istanbul, Türkiye (H.K.); Department of Radiology, Acibadem Kozyatagi Hospital, Istanbul, Türkiye (D.E.T.Ş.); Department of Radiology, Acıbadem University Medical School, Istanbul, Türkiye (E.A.); Department of Radiology, Gaziantep University, Gaziantep, Türkiye (D.E.T.Ş.). Received November 7, 2022; revised January 15, 2023; accepted January 17, 2023. Address correspondence to: O.B. Fevzi Çakmak Mah, Muhsin Yazıcıoğlu Cd, No:10, 34899 Pendik/İstanbul/Türkiye e-mail: onur.bugdayci@marmara.edu.tr

https://doi.org/10.1016/j.acra.2023.01.024

between benign and malignant masses of the breast (6, 7). SWE images provide an estimation of several quantitative shear-wave velocities and velocity-derived elasticity measures for a lesion, which offers good accuracies (8, 9). Moreover, the combination of SWE with B-mode US has been reported to further improve accuracy (10–12). SWE has also been shown to be highly reproducible and improve the diagnostic performance of inexperienced practitioners (13–15).

Nevertheless, the routine use of UE is still not widely adopted. In a survey study, most breast radiologists from Canada and Québec (68.8% and 82.2%, respectively) reported that they never use UE, most considering it to be an unreliable test (16). Indeed, many studies are advocating different cut-off points, most likely because of the use of different systems or different amounts of precompression (17).

BI-RADS descriptors are used in daily practice to categorize lesions and plan the management of patients (4). However, it does not offer information about which parameters are strongly associated with malignancy, nor does it provide a straightforward decision-making algorithm to categorize patients. This categorization is especially critical in differentiating category 3 and 4A patients, who are either followed-up or biopsied, respectively. As a result, breast US results in many unnecessary (i.e., benign) biopsies and the benchmark for a positive biopsy rate is quite low for breast US (7.4%) (4). Although core needle breast biopsy is a safe procedure, some patients experience considerable anxiety before, and pain during the procedure (18–20). In some women, anxiety can persist for up to 3 years after a benign result and may interfere with adherence to follow-up procedures (20).

In this study, we aim to develop a model from B-mode US findings, Doppler US, and complementary SWE measurements that can be used as a scoring system to reduce benign breast biopsies in clinical practice.

MATERIALS AND METHODS

Patient Population

Between November 2016 and January 2018 patients who underwent B-mode ultrasound imaging (±Doppler US) either after 2D-FFD mammography for any reason (screening or diagnostic), as first-line imaging in symptomatic patients younger than 40 years, or to follow-up of known lesions and who received a Breast Imaging Reporting and Data System (BI-RADS) score of 4 or 5 constituted our study group. The mammography images and ultrasound examinations were evaluated by one of the two radiologists experienced in breast radiology for more than 10 years. Written informed consent was obtained before enrollment. Histopathological results were the reference gold standard. High-risk (B3) lesions were categorized as benign or malignant depending on the histopathological results or 1-year follow-up.

Ultrasound Imaging and Image Evaluation

Patients were examined with shear-wave elastography in the second session before the biopsy. Ultrasound and SWE imaging were performed by a third radiologist with 5 years of experience on an ACUSON S2000 Ultrasound System (Siemens Medical Solutions, Erlangen, Germany) using a 4.0-9.0MHz linear probe (9L4; Siemens Medical Solutions, Erlangen, Germany). This radiologist was blinded to clinical information regarding the patient and examined only the lesion scheduled for biopsy. Biopsy was performed by one of the two radiologists who performed the initial examinations.

The B-mode US images were evaluated according to US BI-RADS guidelines (4). Lesion shape, margin, orientation, vascularity, internal echo structure, and posterior echo features were recorded. The lesion size ratio (Sz_R) (longest-/shortest diameters) was calculated (e.g., for a round lesion, Sz_R would reach its smallest value of 1.0).

SWE imaging was performed using the "virtual touch tissue imaging quantification" technique. During SWE imaging, patients were asked to stop breathing for 5 seconds to minimize motion artifacts. Lesions were imaged in their longest diameter and minimal pressure was applied to the breast. The image was divided into B-mode US and SWE maps during acquisition. The lesion was outlined using the freehand annotation tool on the B-mode image, which automatically copied the selected area to the SWE map. If present, the echogenic halo around the lesion was not included in the measurements. A square-shaped 2-dimensional region of interest (ROI) having an area of 2.25 mm^2 was used. Multiple ROIs were placed in a lesion to minimize sampling error, guided by the color maps created by the scanner, and filling the entire lesion area in most cases. The number of measurements depended on lesion size, with more measurements in larger lesions. There were, however, never less than four measurements. In a case where the imaging system could not provide a velocity measurement because the range of 0.5-10 m/s was exceeded, a fixed value of 10 m/s was considered for the measurement. The two lowest and highest velocity measurements were recorded and averaged to obtain the minimum velocity (SWV_{Min}) and the maximum velocity (SWV_{Max}), respectively. The following parameters were calculated from these values:

$$Velocity\ Heterogeneity(SWV_H) = SWV_{Max} - SWV_{Min} \quad (21)$$

$$Velocity\ Ratio(SWV_R) = SWV_{Min}/SWV_{Max}$$

$$Normalized\ velocity\ ratio(SWV_{Rn}) = (SWV_{Max} - SWV_{Min})/SWV_{Min}$$

Linear Regression Models

Setting Y as the response; βp as the coefficient of the p-th predictor, xp in a linear regression model, the equation expressing the model could be given by (22)

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_p x_p$$

Linear regression models utilizing the features extracted from the US images and the SWE images as the predictors

were designed to distinguish benign from malignant masses as the response. The continuous features were converted into their categorical equivalents using decision tree analyses performed with chi- squared automatic interaction detection for a tree depth of one and a significance level of 0.05 (23). Two separate models were designed: one model utilized US findings only, and the other employed US findings and SWE measures.

Model Fitting and Determination of the Best Fit

The models designed were fitted using bootstrapping incorporated stepwise regression analysis, and the optimal values of the coefficients for the predictors in the models developed were determined. At first, the features and the ground truth diagnoses from all the lesions studied were populated to create the original dataset that was entered into analysis in a stepwise manner and a fitting was done that output the optimal equation having an optimal number of predictors in the model (24). Next, 100 dataset replicates, each with the same number of lesions as in the original dataset, but covering random resamples from the original dataset were generated by performing bootstrapping (25). These datasets were used to perform regression fittings for the optimal equation determined in the first phase. Consequently, a total of one hundred additional model fits were obtained.

The performance of a fit was determined by performing receiver operating characteristic analysis and by calculating the area under the curve given by the analysis (AUC) performed for that fit. The performance of the "initial" fit performed for the original dataset was considered to be reflecting an apparent performance and represented by the apparent AUC. The performances of the subsequent fits for the subsequent dataset replicates and for the original dataset were represented by the bootstrap-sample and the original-sample AUCs. The average of the difference between these performance pairs was assumed to be an estimate of the optimism in the apparent performance and later subtracted from the apparent performance to estimate the internally validated AUC. The fit having an original-sample AUC closest to the internally validated AUC was deemed the best fit (26). The methodology is summarized in Figure 1.

Test Validation

A dataset from a different center, obtained by another operator (D.E.T.S.) on a different scanner (GE Logiq S8, GE Healthcare, Milwaukee, US) using a 4-15MHz matrix linear probe (ML 6-15) for the B-mode US and a 2-8 MHz linear probe (9L-D) for SWE acquisition, was used for validation of the constructed models. This dataset was initially acquired for another study, and the operator was unaware of our research during data acquisition (27). Permission for use of this data was obtained from the local ethics committee (Decision number 2022-09/04) as well as from the publisher. From this dataset, only BI-RADS 3, 4, and 5 lesions were included for validation, which in total consisted of 187 lesions. Patients were between 21 and 84 years of age and lesion size was between 5 and 57 mm. BI-RADS criteria were used for lesion evaluation, and SWE measurements were performed identically to this study. BI-RADS 3 lesions that did not undergo biopsy were followed up for at least 1 year. For a detailed description of the study design, the reader is referred to the study by D.E.T.S. et al. (27). Additional information is presented in Table 1.

Statistical Analysis

Youden's analysis was performed to determine an optimal cut-off for the model with the best fit using the outputs of the model for the original dataset. The cut-off value was next used to assess the sensitivity, specificity, and accuracy metrics for the model in discriminating benign and malignant masses. The sensitivity, specificity, and accuracy were considered very high, high, moderate, low, and very low if their values were 95%-100%, 85%-94.9%, 75%-84.9%, 65%-74.9%, and 0%-64.9%, respectively. Performance according to an AUC attributed excellent, good, fair, poor, and fail if the AUC was 0.90-1.00, 0.80-0.89, 0.70-0.79, 0.60-0.69, and 0.50-0.59, respectively (28). The method proposed by DeLong et al. was used to compare the systematic difference between the two AUCs (29). A p-value of <0.05 was considered for statistical significance. All analyses were performed using our in-house computer software tools developed using MATLAB (v8.2; Mathworks).

Ethical approval for this prospective study was obtained from the local ethics committee (Approval ID: 09.2015.033).

RESULTS

Demographics

Three hundred ninety-seven female patients aged 18-87 years (mean, 46.7 years) were included in the study. There were a total of 418 breast lesions of which 238 (56.9%) were benign and 180 (43.1%) were malignant (Table 1). Lesion size ranged from 3 to 80 mm (mean 20 mm; SD±11.15 mm).

Performance of B-Mode, Doppler US, and SWE Parameters

Evaluation of breast US and SWE images explored the features presented in Table 2. All seven US parameters exhibited poor diagnostic performances in discriminating benign from malignant masses (AUC=0.592-0.696, $p<0.001$), however, Sz_R provided the highest performance (AUC=0.696), showing lower values for malignancy. Except for SWV_{Min}, all SWE parameters performed better than any B-mode US or Doppler US parameter. The best performing parameter was SWV_H with and AUC of 0.816 (Table 2).

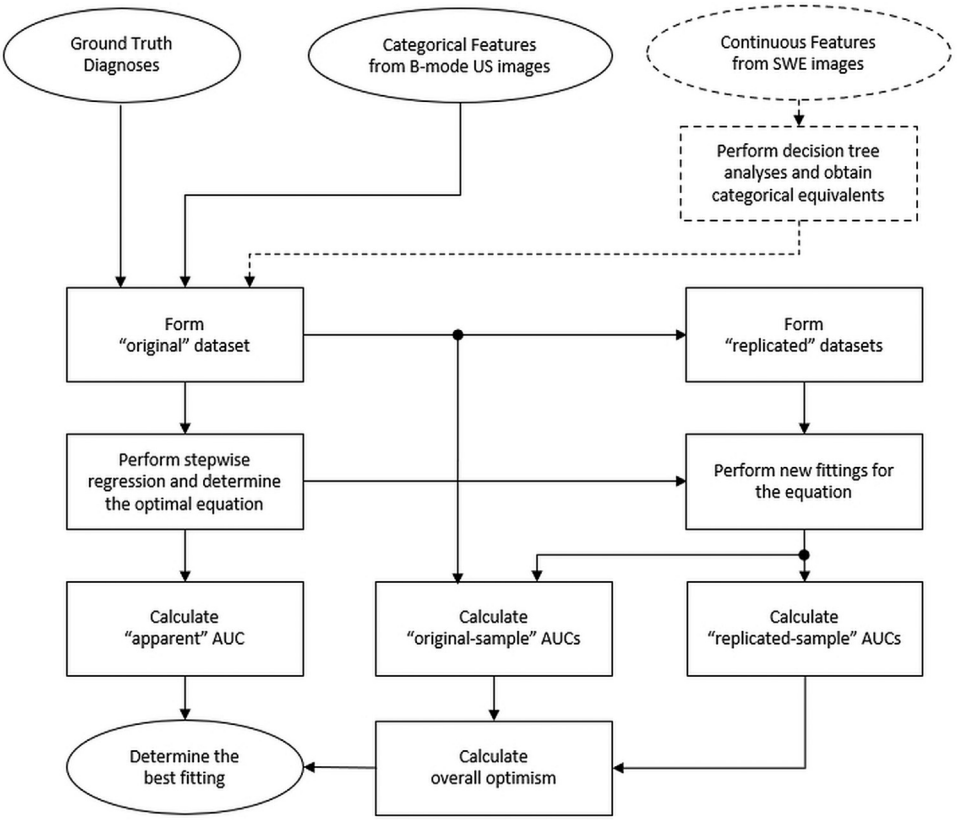

Figure 1. Flowchart depicting the linear regression model, model fitting, and determination of the best fit.

Linear Regression Models

The continuous features depicted the distributions visualized by the box plots are presented in Figure 2. For these features, decision tree analyses output the category assignments and the range for each category, as listed in supplemental Table 1.

The linear regression model that was fitted considering the US features (SS_{US}) output an AUC of 0.819 and removed the shape and the posterior echo features from the model. On the other hand, the linear regression model that was fitted considering both the US and SWE features ($SS_{US,SWE}$) conveyed an AUC of 0.882 and removed the shape, vascularity, posterior echo, SWV_{Min}, SWV_H, and SWV_R features from the model. Table 3 tabulates the statistical attributes of the predictors in each model for the best fit.

Scoring Systems

The scoring systems exhibited good performances in distinguishing benign from malignant breast masses determined from the ROC plots presented in Figure 3. However, $SS_{US, SWE}$ delivered significantly better performance (p=0.001). For the SS_{US}, the optimal cut-off score was found to be 48.3 which led to moderate sensitivity (76.4%), moderate specificity (76.7%), and moderate accuracy (76.6%), while for $SS_{US, SWE}$ the optimal cut-off score of 42.1 provided high sensitivity (84.8%), moderate specificity (75.9%), and moderate accuracy (79.8%). The scoring systems achieved similar specificities, however, higher sensitivity resulting in higher accuracy was provided by $SS_{US, SWE}$.

A simplified scoring system was developed based on the model using the joint evaluation of US and SWE (given in Table 3) to utilize in daily clinical practice. Parameters that were statistically significant ($p<0.05$) and had a positive modifier (i.e., positive β value) in the final score were incorporated into the simplified scoring system given in Table 4. A threshold of <4 had an AUC of 0.856 with a sensitivity of 98.9% and a specificity of 37.8% in discriminating benign from malignant breast lesions, with higher scores indicating malignancy. Ninety benign (37.8% of benign lesions) and two malignant lesions (1.1% of malignant lesions) scored below this cut-off in our dataset (Fig. 4 and 5). Thus, the probability of malignancy below this threshold was 2.17% (2/92 lesions).

Validation

The constructed algorithms and the scoring system were tested with the validation dataset. For SS_{US}/$SS_{US, SWE}$/simplified scoring model the sensitivity was 28% (7/25) / 84% (21/25) / 100% (25/25), the specificity was 99.4% (161/162) / 88.2% (143/162) / 74.6% (121/162), positive predictive value was 87.5% (7/8) / 52.5% (21/40) / 37.8% (25/66), negative predictive value was 89.9% (161/179) / 97.2%

TABLE 1. Distribution of Benign and Malignant Lesions in the Study Group

STUDY GROUP			
		n	%
Diagnosis	Benign	238	56.9
	Malignant	180	43.1
Subtype of Malignant	IDC – NOS	143	79.44
	IDC – Specified	20	11.11
	ILC	10	5.56
	DCIS	5	2.78
	Other	2	1.11
Subtype of Benign	Fibroepithelial Lesion	93	39.07
	Fibrocystic change	38	15.97
	Inflammation*	30	12.61
	Intraductal Papilloma	19	7.98
	Adenosis	17	7.14
	Columnar Cell Change	11	4.57
	Atypia**	11	4.57
	Fat Necrosis	8	3.36
	Apocrine Metaplasia	6	2.47
	Other	6	2.47
BI-RADS Category and the Number of Benign/Malignant Lesions	4A	97	89/8
	4B	145	124/21
	4C	49	16/33
	5	127	9/118

VALIDATION GROUP			
		n	%
Diagnosis	Benign	162	86.6
	Malignant	25	13.4
Type of Diagnosis	Follow-up	129	69
	Biopsy	58	31
Subtype of Malignant	IDC	23	92
	ILC	2	8
Subtype of Benign***	Fibroepithelial Lesion	25	75.8
	Stromal Hyalinization	2	6.08
	Inflammation	30	3.02
	Apocrine Metaplasia	19	3.02
	Adenosis	17	3.02
	Fibrocystic change	11	3.02
	Fibrosis	11	3.02
	Tubular adenoma	8	3.02
Initial BI-RADS Category/Number of biopsied (Number of malignant cases)	3	151	25 (0)
	4A	15	12 (4)
	4C	10	10 (10)
	5	11	11 (11)

DCIS, Ductal Carcinoma in situ; IDC, Invasive Ductal Carcinoma; IDC – NOS, Invasive Ductal Carcinoma - Not Otherwise Specified; ILC, Invasive Lobular Carcinoma.

* Includes acute and chronic inflammation including granulomatous mastitis.

** All cases of atypia included in the benign category were confirmed to be benign by surgery. Malignant cases of atypia were evaluated in the malignant group.

*** Data about biopsied lesions only.

(143/147) / 100% (121/121), and the accuracy was 89.8% / 87.7% / 78.1%. One lesion scored above the thresholds on all three models, but the lesion did not increase in size during follow-up. The $SS_{US, SWE}$ model missed four malignancies (3 BI-RADS 4A and 1 BI-RADS 5 lesions).

All four malignancies out of the 12 biopsy-proven BI-RADS category 4A lesions correctly scored above the threshold in the simplified scoring system, while 54.5% (6/11) of benign BI-RADS 4A lesions (biopsy+follow-up) scored below the threshold. Therefore, the simplified scoring system

TABLE 2. Features Extracted From B-mode US and SWE Images

Feature	Category	Benign*	Malignant*	AUC
Shape Sh	1 Oval	68	10	0.650
	2 Round	29	14	
	3 Irregular	135	154	
Margin Ma	0 Circumscribed	123	46	0.636
	1 Not Circumscribed	109	132	
Vascularity Va	0 None	142	70	0.602
	1 Internal	82	105	
	2 Peripheral	8	3	
Orientation Or	0 Parallel	89	14	0.652
	1 Non-Parallel	143	164	
Internal Echo iE	0 Anechoic	2	0	0.682
	1 Hyperechoic	14	4	
	2 Isoechoic	28	3	
	3 Complex Cystic	23	3	
	4 Heterogenous	65	33	
	5 Hypoechoic	100	135	
Posterior Echo pE	0 No Features	211	130	0.592
	1 Enhancement	4	1	
	2 Shadowing	17	47	
Size Ratio Sz_R	-	1.78±0.62[a] (1.00-4.33)[b]	1.42±0.39[a] (1.00-3.83)[b]	0.696
Minimum Velocity (m/s) SWV_{Min}	-	2.36±0.72[a] (1.18-5.56)[b]	2.70±0.98[a] (0.89-6.52)[b]	0.607
Maximum Velocity (m/s) SWV_{Max}	-	4.32±1.73[a] (1.61-10)[b]	6.92±2.30[a] (2.01-10)[b]	0.808
Velocity Heterogeneity (m/s) SWV_H	-	1.96±1.43[a] (0.15-7.55)[b]	4.22±2.05[a] (0.52-8.26)[b]	0.816
Velocity Ratio SWV_R	-	0.59±0.16[a] (0.24-0.94)[b]	0.42±0.15[a] (0.17-0.85)[b]	0.776
Velocity Ratio Normalized SWV_{Rn}	-	0.86±0.62[a] (0.06-3.25)[b]	1.71±1.02[a] (0.18-4.88)[b]	0.776

* Number of breast masses. ([a]) Mean±Standart Deviation and ([b]) Minimum-Maximum values.

could have reduced 40% (6/15) of biopsies of BI-RADS 4A lesions. Additionally, the system also correctly identified 64% (16/25) of biopsy-proven benign BI-RADS 3 lesions (true negative). On the other hand, 23% (29/126) of BI-RADS 3 lesions that were only followed up scored above the threshold (false positive).

DISCUSSION

In this study, we identified the relative importance of different BI-RADS US and Doppler US descriptors in predicting malignancy. We also evaluated the performance of different SWE parameters in the differentiation of breast lesions. Using regression analysis, we first grouped measured values (or their computed ratios) into categories and calculated an optimal threshold for differentiating benign versus malignant breast lesions using only US and Doppler US in one arm and using US, Doppler US, and SWE in combination in another arm. This analysis was then simplified into an easy-to-use scoring system and validated using an external dataset acquired by a different operator on a different scanner.

The BI-RADS parameters with the highest AUC were hypoechoic composition, nonparallel orientation, and irregular shape. However, the echo structure of the lesion was not found to be a predictor for malignancy in the regression analysis, which revealed that a non-circumscribed margin, internal vascularity, and nonparallel orientation were more significant predictors of malignancy. Other studies have similarly found internal echogenicity not to be associated with malignancy while reporting that non-circumscribed margin and nonparallel orientation demonstrated the highest PPV (30, 31). Internal vascularity was also an important contributor to the assessment of breast lesions (32). However, both descriptive and regression analyses showed that the size ratio (Sz_R) was the most important B-mode US predictor of malignancy. It could be argued that evaluating lesions as "oval" and "non-oval" might increase diagnostic accuracy and warrants further studies.

Shear-wave ultrasound elastography has been shown to increase diagnostic accuracy when used together with B-mode US and its use is now being recommended by different organizations (5, 10–12, 32, 33). In line with previous reports, adding SWE to B-mode US and Doppler US raised

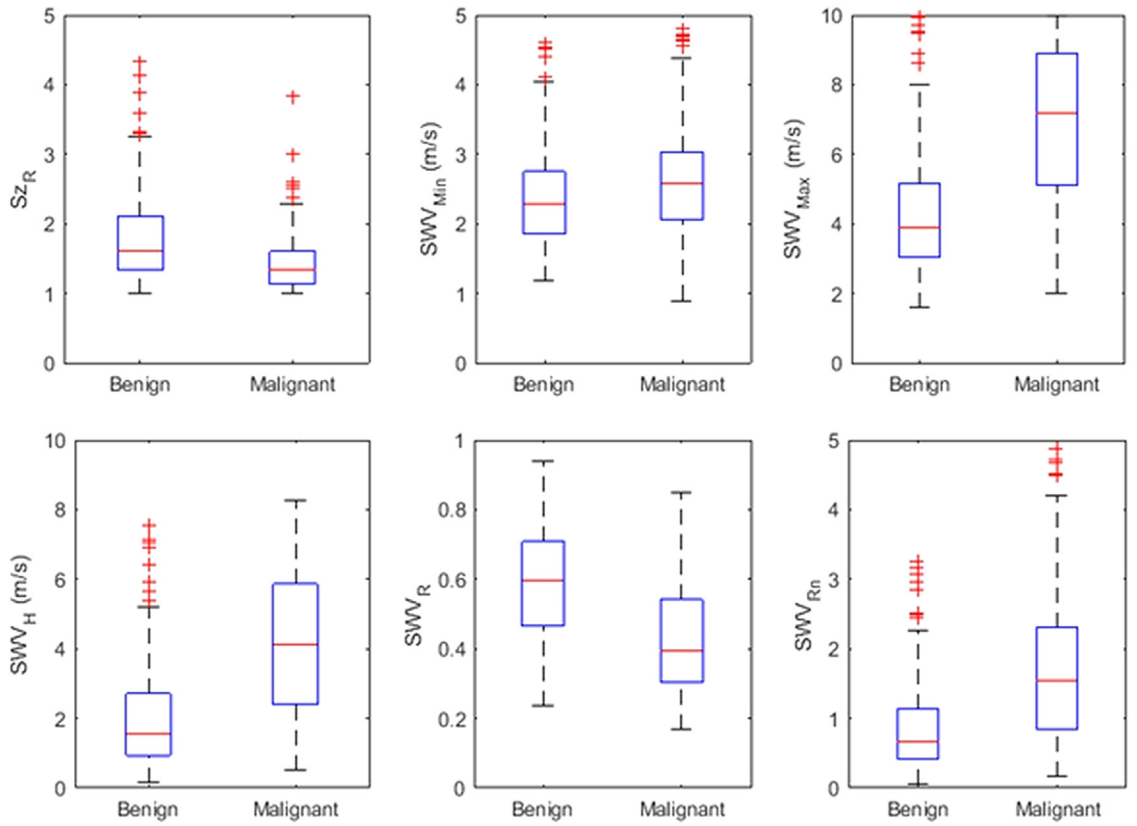

Figure 2. Boxplots for the continuous features from US and SWE images of the benign and malignant breast masses.

the AUC in our study as well, delivering high sensitivity, moderate specificity, and moderate accuracy in the $SS_{US,\ SWE}$ model. Except for minimum shear velocity, all SWE parameters performed better than B-mode US or Doppler US parameters. Velocity heterogeneity had the highest AUC with 0.816, followed by the maximum velocity with 0.808. The normalized velocity ratio (SWV_{Rn}), another parameter of elastographic heterogeneity, was also considered to be very significant. Huang et al. also reported that elastographic heterogeneity (calculated by subtracting the sum of the three

TABLE 3. Linear Regression Model Utilizing Predictors From US and SWE Images

US Only				US and SWE Combined			
Predictor	β	95% CI for β	*p*-Value	Predictor	β	95% CI for β	*p*-Value
Va_1	19.5	11.1, 27.8	<0.001	Or_1	16.9	7.3, 26.4	0.001
Va_2	3.4	-23.2, 30.1	0.800	Ma_1	10.6	2.8, 18.4	0.008
Or_1	18.8	8.0, 29.6	0.001	iE_1	-21.6	-44.3, -1.0	0.061
Ma_1	20.8	12.1, 29.4	<0.001	iE_2	-32.4	-48.5, -16.2	<0.001
iE_1	-15.4	-39.9, -9.1	0.219	iE_3	-30.6	-48.0, -13.1	0.001
iE_2	-23.1	-40.1, -6.1	0.008	iE_4	-36.4	-50.5, -22.3	<0.001
iE_3	-20.9	-38.2, -3.6	0.018	iE_5	-11.2	-24.1, 1.6	0.086
iE_4	-27.3	-41.1, -13.4	<0.001	$Sz_{R,1}$	16.6	7.6, 25.6	<0.001
iE_5	6.1	-6.0, 18.2	0.323	$Sz_{R,2}$	20.4	8.6, 32.3	0.001
$Sz_{R,1}$	17.4	7.0, 27.7	0.001	$SWV_{Max,\ 1}$	2.2	-8.2, 12.6	0.679
$Sz_{R,2}$	25.4	11.9, 38.9	<0.001	$SWV_{Max,\ 2}$	35.0	24.0, 46.0	<0.001
				$SWV_{Max,\ 3}$	48.9	35.6, 62.1	<0.001
				$SWV_{Rn,\ 1}$	18.7	7.9, 29.5	0.001
				$SWV_{Rn,\ 2}$	16.7	3.0, 30.5	0.017

iE, Internal Echogenicty; Ma, Margin; Or, Orientation; Va, Vascularity. Numbers next to these predictors denote categories shown in Table 2.

SWV_{Max}, Maximum Shear Velocity; SWV_{Rn}, Normalized Velocity Ratio; Sz_R, Size Ratio. Numbers next to these predictors denote categories shown in supplemental Table 1.

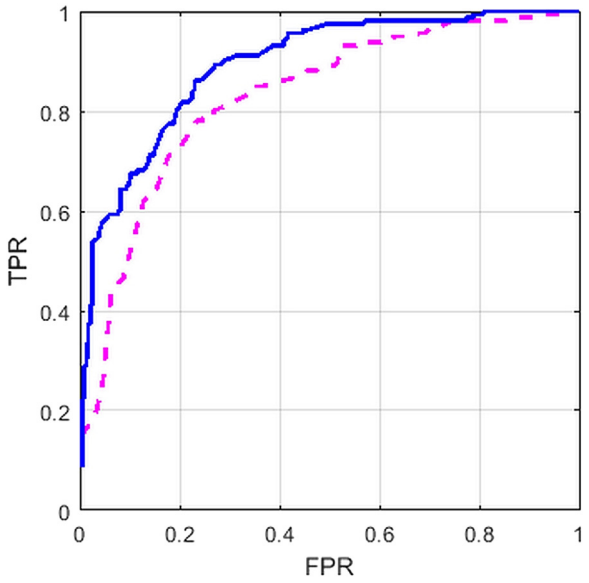

Figure 3. ROC curves of US+CDUS based model (SS_{US} – Dashed line) and US+SWE model ($SS_{US,\ SWE}$ – Solid line). $SS_{US,\ SWE}$ performed significantly better than SS_{US}.

lowest velocities from the sum of the three highest velocities) performed better than other SWE parameters in their study (21).

Regression analysis revealed SWV_{Max} to be the most valuable diagnostic SWE parameter, however, and the most significant indicator of malignancy overall, which is also supported by other studies (34, 35). In addition, SWV_{Max} has also been reported to have almost perfect interobserver agreement (15). Nonetheless, cut-off values differ between studies which has been a major factor in preventing widespread adoption of elastography (17). Some studies found the lesion-to-fat ratio to be more valuable, which was not assessed in our study (36, 37). However, this parameter was also reported to have lower interobserver agreement compared to SWV_{Max} (15). Additionally, the stiffest portion of malignant lesions might be outside the lesion itself, which was also not assessed in our study and could be considered a limitation (38). Some studies have measured the whole lesion within a single ROI and measured mean elasticity and standard deviations of elasticity. However, the diagnostic accuracy was found to be higher when a smaller ROI was placed on the stiffest part of the lesion (39, 40).

TABLE 4. Simplified Scoring System to Reduce Benign Biopsies of BI-RADS 4A Lesions

Parameter	Modifier	Score
SWV_{Max} (m/s)	>7.89	5
	5.67-7.89	3.5
Size Ratio	≤1.18	2
	1.95-1.18	1.5
SWV_{Rn}	0.48-1.49	2
	>1.49	1.5
Orientation	Nonparallel	1.5
Margin	Not circumscribed	1

Our study group consisted of only BI-RADS 4 and 5 lesions, whereas the validation dataset mostly (81%) consisted of BI-RADS 3 lesions. Therefore, the pretest probability of malignancy differed between both populations, with the validation dataset being more in-line with daily practice. The validation of the SS_{US} model revealed an unacceptably low level of sensitivity (28%). Therefore, the high number of missed cancers would make it unsuitable for clinical use. The validation of the $SS_{US,SWE}$ model demonstrated similar sensitivity, higher specificity, and better accuracy compared to the study group, confirming the performance of the model. However, although having a high NPV, it missed three of the four malignant BI-RADS 4A lesions, defying its use as a tool to reduce breast biopsies.

We developed a simplified scoring method for clinical use to aid in the decision-making for breast lesion biopsy (Table 4). The method requires only measuring the maximum and minimum shear velocities besides standard B-mode breast US examination. Using a threshold score of <4, a biopsy could have been avoided in 37.8% (90/238) of benign lesions in our study, with a tradeoff of missing 1.1% (2/180) of malignant lesions. The probability of malignancy below this threshold was 2.17%, which almost perfectly differentiates BI-RADS category 3 from category 4A. Both missed lesions in our study had suspicious features on mammography (Fig 5). There were no missed cancers in the validation group, probably, as mentioned before, because the pretest probability of cancer was lower in the validation group.

The simplified score reached a sensitivity and NPV of 100% in the validation dataset. It correctly identified 54.5% (6/11) of benign BI-RADS 4A lesions and could have reduced 40% (6/15) of biopsies in this category in this dataset. On the other hand, the simplified score output a false positivity of 23% in BI-RADS 3 lesions. When evaluating together the two missed cancers in the study group and the high NPV but also the high false positive rate in the validation group, we believe that this scoring system would be most beneficial when used for downgrading BI-RADS 3 and 4A lesions. Using this system to upgrade BI-RADS 3 lesions would result in higher false positivity, thus counteract its own purpose. On the other hand, downgrading lesions BI-RADS 4B or higher (although a rare occurrence) may result in false negativity, like in our study group.

Evaluating 124 lesions, Kapetas et al. proposed a US-based algorithm in their study (41). They constructed a decision tree algorithm based on the Resistivity Index on Doppler US and maximum velocity on SWE. They reported an AUC of 0.887 with a sensitivity and specificity of 98.46% and 61.02%, respectively. Our study differs in several points. We included established BI-RADS criteria in our evaluation, which are used in daily practice and have been shown to be valuable (42). We also adhered to BI-RADS guidelines in terms of Doppler evaluation since Doppler parameters have a significant overlap between benign and malignant lesions (43).

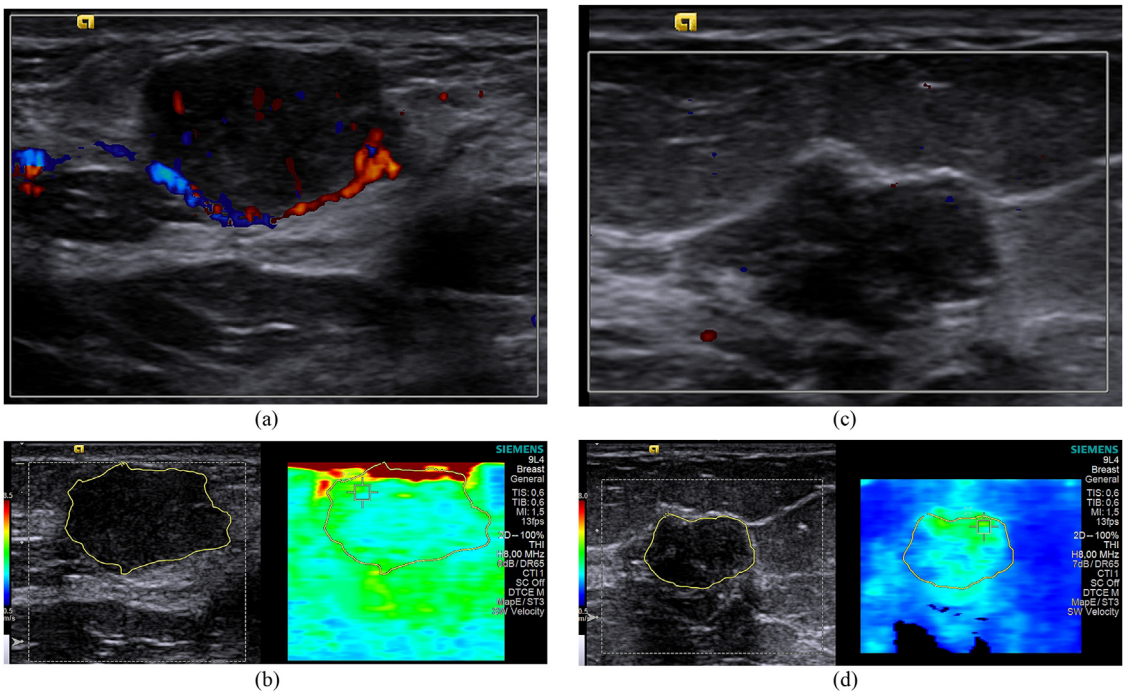

Figure 4. Two cases diagnosed as fibroadenoma on biopsy, both of which received a score of 3. (a, b) 30-year-old woman with a 1.7 cm lesion in her left breast. (a) On CDUS the lesion shows internal and peripheral vascularization. (b) On SWE the lesion had a SWV_{Max} of 3.82 m/s. (c, d) 44-year-old woman with a 1.6 cm lesion with lobulated margins in her right breast. (c) The lesion exhibited no vascularity on CDUS. (d) On SWE the lesion had a SWV_{Max} of 3.73 m/s.

Additionally, we aimed to develop a scoring system, rather than a decision tree. While a decision tree algorithm could discard a malignant lesion based on the absence of a single feature, a scoring system may detect the malignancy by considering additional features the lesion might have.

Lee et al., in their multicenter study, evaluated the benefit of Doppler US and UE (SWE and strain elastography) in a screening setting in women with dense breasts (32). They reported that BI-RADS category 4A lesions could be downgraded to category 3 without missing any malignant lesion if the lesion were soft (no threshold was reported) and did not show any vascularity, potentially avoiding biopsy in 67.7% of benign lesions. However, only 7.9% (19/238) of benign lesions in our study did not show any vascularity and had a SWV_{Max} of ≤

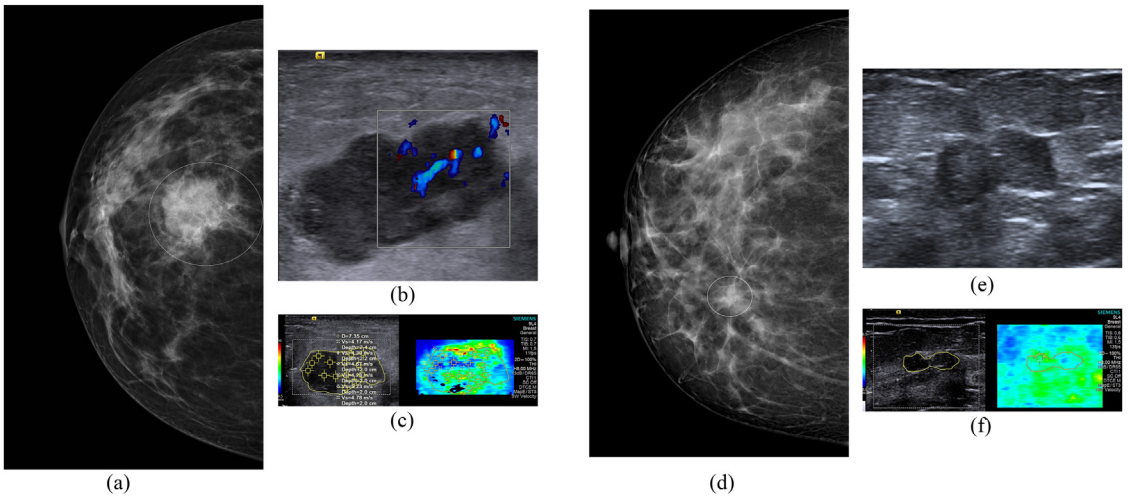

Figure 5. The two cases missed by the scoring system. (a-c) 49-year-old woman with a 3 cm lesion in her left breast with a score of 3. The lesion turned out to be a Her-2 enriched invasive ductal carcinoma. (a) R-CC mammography showing the centrally located lesion with irregular margins. (b) The lesion exhibited internal vascularity on CDUS. (c) It had a SWV_{Max} of 5.23 m/s on SWE. (d-f) 41-year-old woman with a 1.2 cm lesion at the 2 o'clock position in her right breast which was diagnosed as a luminal A invasive ductal carcinoma. The lesion received a score of 2. (d) CC view of the right breast showing an asymmetry in the inner part of the breast. (e) The lesion exhibits small microlobulations on US which was not appreciated during the initial exam and should therefore got a score of 3. (f) On SWE the lesion had a SWV_{Max} of 3.31 m/s.

2.6 m/s, which is considered to be a threshold for benign lesions (3). This discrepancy could be attributed to the study setting (screening only vs. screening and diagnostic combined) and the relatively low number of malignancies (n=68) in their study. The fact that the proportion of soft and avascular benign lesions was higher (42.2%) in our validation dataset, which also has fewer malignancies, supports this assumption. Of note, five malignant lesions (including a 6 mm and a 9 mm lesion) in our study did not demonstrate any vascularity and were soft on SWE (These five malignant lesions were correctly classified by our scoring system).

Most recently, Pfob et al. developed a decision rule to differentiate between BI-RADS category 3 and 4A lesions, studying 1294 women (31). They classified the patient as category 4 if at least one of the following features occurred: calcification, nonparallel orientation, noncircumscribed margin, posterior shadowing, or patient age 50 years or older. They compared their results to three experts and report that they improved sensitivity to 97%. They also report specificity to be 44%, while expert sensitivity and specificity were 79%-95% and 22%-83%, respectively. We deliberately excluded age and lesion size as parameters in our study because this would have introduced significant bias into our analysis by favoring older patients or larger lesions, making the scoring system less useful in younger patients or small lesions.

There are some limitations to the current study. Although our model was validated using data from another center, obtained by another operator using a different scanner, our model fittings were performed on data from a single-center dataset. Other operators using different scanners may obtain dissimilar velocity estimates. The estimates may need to be recalculated using a correction formula (44). Although the placement of a 2D ROI is done under the guidance of the "virtual touch tissue imaging quantification" technique, which has been demonstrated to facilitate accurate feature estimations, manual placement may introduce inter- or intraobserver variability on the feature estimates (45). Our models accept categorical features only, and some extra effort is needed to identify the correct category for a measured value of a continuous feature. Lastly, our validation dataset was relatively small, with few BI-RADS 4A lesions and malignancies.

In conclusion, adding SWE improved the diagnostic accuracy of breast US. Using size ratio, lesion orientation, margin, $SWV_{Max,}$ and $SWV_{Min,}$ we developed a scoring model, which could decrease benign biopsies of BI-RADS 4A lesions considerably. Multi-center studies with larger study groups, different scanners, and operators are needed to verify and, if necessary, modify our results.

DECLARATION OF COMPETING INTEREST

The authors declare no competing interest regarding this work.

ACKNOWLEDGMENTS

This work has been supported by Marmara University Scientific Research Projects Coordination Unit under grant number SAG-E-130515-0155.

REFERENCES

1. Berg WA, Bandos AI, Mendelson EB, Lehrer D, Jong RA, Pisano ED. Ultrasound as the primary screening test for breast cancer: analysis from ACRIN 6666. J Natl Cancer Inst 2016; 108:djv367. doi:10.1093/jnci/djv367.
2. Barr RG, Zhang Z. Effects of precompression on elasticity imaging of the breast: development of a clinically useful semiquantitative method of precompression assessment. J Ultrasound Med: Official J Am Inst Ultrasound Med 2012; 31:895–902. doi:10.7863/jum.2012.31.6.895.
3. Barr RG, Nakashima K, Amy D, et al. WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 2: breast. Ultrasound Med Biol 2015; 41:1148–1160. doi:10.1016/j.ultrasmedbio.2015.03.008.
4. Mendelson E, Böhm-Vélez M, Berg W, Whitman G, Feldman M, Madjar H. ACR BI-RADS ultrasound. ACR BI-RADS® atlas, breast imaging reporting and data system. Reston, VA: American College of Radiology, 2013:334.
5. Săftoiu A, Gilja OH, Sidhu PS, et al. The EFSUMB guidelines and recommendations for the clinical practice of elastography in non-hepatic applications: update 2018. Ultraschall Med 2019; 40:425–453. doi:10.1055/a-0838-9937.
6. Xue Y, Yao S, Li X, Zhang H. Value of shear wave elastography in discriminating malignant and benign breast lesions: a meta-analysis. Medicine 2017; 96:e7412. doi:10.1097/md.0000000000007412.
7. 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; 8:89173–89181. doi:10.18632/oncotarget.21124.
8. Liu B, Zheng Y, Huang G, et al. Breast lesions: quantitative diagnosis using ultrasound shear wave elastography-a systematic review and meta−analysis. Ultrasound Med Biol 2016; 42:835–847. doi:10.1016/j.ultrasmedbio.2015.10.024.
9. Zheng X, Huang Y, Liu Y, et al. Shear-wave elastography of the breast: added value of a quality map in diagnosis and prediction of the biological characteristics of breast cancer. Korean J Radiol 2020; 21:172–180. doi:10.3348/kjr.2019.0453.
10. Berg WA, Cosgrove DO, Doré CJ, et al. Shear-wave elastography improves the specificity of breast US: the BE1 multinational study of 939 masses. Radiology 2012; 262:435–449. doi:10.1148/radiol.11110640.
11. Evans A, Whelehan P, Thomson K, et al. Differentiating benign from malignant solid breast masses: value of shear wave elastography according to lesion stiffness combined with greyscale ultrasound according to BI-RADS classification. Br J Cancer 2012; 107:224–229. doi:10.1038/bjc.2012.253.
12. Luo J, Cao Y, Nian W, et al. Benefit of shear-wave elastography in the differential diagnosis of breast lesion: a diagnostic meta-analysis. Med Ultrason 2018; 1:43–49. doi:10.11152/mu-1209.
13. Aslan H, Pourbagher A, Ozen M. The role of Shear-Wave elastography in the differentiation of benign and malign non-mass lesions of the breast. Ann Ital Chir 2018; 89:385–391.
14. Cosgrove DO, Berg WA, Doré CJ, et al. Shear wave elastography for breast masses is highly reproducible. Eur Radiol 2012; 22:1023–1032. doi:10.1007/s00330-011-2340-y.
15. Hong S, Woo OH, Shin HS, et al. Reproducibility and diagnostic performance of shear wave elastography in evaluating breast solid mass. Clin Imaging 2017; 44:42–45. doi:10.1016/j.clinimag.2017.03.022.
16. Mesurolle B, El Khoury M, Chammings F, et al. Breast sonoelastography: Now and in the future. Diagn Interv Imaging 2019; 100:567–577. doi:10.1016/j.diii.2019.03.009.
17. Barr RG. Future of breast elastography. Ultrasonography 2019; 38:93–105. doi:10.14366/usg.18053.
18. Soo AE, Shelby RA, Miller LS, et al. Predictors of pain experienced by women during percutaneous imaging-guided breast biopsies. J Am Coll Radiol 2014; 11:709–716. doi:10.1016/j.jacr.2014.01.013.

19. Evans A, Trimboli RM, Athanasiou A, et al. Breast ultrasound: recommendations for information to women and referring physicians by the European Society of Breast Imaging. Insights Imaging 2018; 9:449–461. doi:10.1007/s13244-018-0636-z.
20. Miller LS, Shelby RA, Balmadrid MH, et al. Patient anxiety before and immediately after imaging-guided breast biopsy procedures: impact of radiologist-patient communication. J Am Coll Radiol 2016; 13:e62–e71. doi:10.1016/j.jacr.2016.09.034.
21. Huang Y, Li F, Han J, et al. Shear wave elastography of breast lesions: quantitative analysis of elastic heterogeneity improves diagnostic performance. Ultrasound Med Biol 2019; 45:1909–1917. doi:10.1016/j.ultrasmedbio.2019.04.019.
22. Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107:776–782. doi:10.3238/arztebl.2010.0776.
23. Miller B, Fridline M, Liu PY, Marino D. Use of CHAID decision trees to formulate pathways for the early detection of metabolic syndrome in young adults. Comput Math Methods Med 2014; 2014:242717doi:10.1155/2014/242717.
24. Efroymson M. Multiple regression analysis. Mathematical methods for digital computers. New York: John Wiley, 1960:191–203.
25. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15:361–387. doi:10.1002/(sici)1097-0258(19960229)15:4<361::Aid-sim168>3.0.Co;2-4.
26. Musoro JZ, Zwinderman AH, Puhan MA, et al. Validation of prediction models based on lasso regression with multiply imputed data. BMC Med Res Methodol 2014; 14:116. doi:10.1186/1471-2288-14-116.
27. Tekcan Sanli, Yildirim D, Kandemirli S G, Sanli A N, Aribal E. Evaluation of Multiparametric Shear Wave Elastography Indices in Malignant and Benign Breast Lesions. Acad Radiol 2022; 29:S50–S61, doi:10.1016/j.acra.2021.09.015.
28. Li F, He H. Assessing the Accuracy of Diagnostic Tests. Shanghai Arch Psychiatry 2018; 30:207–212. doi:10.11919/j.issn.1002-0829.218052.
29. 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.
30. Gu Y, Tian JW, Ran HT, et al. The utility of the fifth edition of the BI-RADS Ultrasound Lexicon in category 4 breast lesions: a prospective multicenter study in China. *Acad Radiol* 2020; 29:S26–S34. doi:10.1016/j.acra.2020.06.027.
31. Pfob A, Barr RG, Duda V, et al. A new practical decision rule to better differentiate BI-RADS© 3 or 4 breast masses on breast ultrasound. J Ultrasound Med: Official J Am Inst Ultrasound Med 2021; 41:427–436. doi:10.1002/jum.15722.
32. Lee SH, Chung J, Choi HY, et al. Evaluation of screening US-detected breast masses by combined use of elastography and color Doppler US with B-mode US in women with dense breasts: a multicenter prospective study. Radiology 2017; 285:660–669. doi:10.1148/radiol.2017162424.
33. Uematsu T, Nakashima K, Kikuchi M, et al. The Japanese breast cancer society clinical practice guidelines for breast cancer screening and diagnosis, 2018 edition. Breast Cancer 2020; 27:17–24. doi:10.1007/s12282-019-01025-7.
34. Lee EJ, Jung HK, Ko KH, et al. Diagnostic performances of shear wave elastography: which parameter to use in differential diagnosis of solid breast masses? Eur Radiol 2013; 23:1803–1811. doi:10.1007/s00330-013-2782-5.
35. Lin X, Chang C, Wu C, et al. Confirmed value of shear wave elastography for ultrasound characterization of breast masses using a conservative approach in Chinese women: a large-size prospective multicenter trial. Cancer Manag Res 2018; 10:4447–4458. doi:10.2147/cmar.S174690.
36. Au FW, Ghai S, Moshonov H, et al. 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. doi:10.2214/ajr.13.11693.
37. Youk JH, Gweon HM, Son EJ, et al. Diagnostic value of commercially available shear-wave elastography for breast cancers: integration into BI-RADS classification with subcategories of category 4. Eur Radiol 2013; 23:2695–2704. doi:10.1007/s00330-013-2873-3.
38. Xiao Y, Yu Y, Niu L, et al. Quantitative evaluation of peripheral tissue elasticity for ultrasound-detected breast lesions. Clin Radiol 2016; 71:896–904. doi:10.1016/j.crad.2016.06.104.
39. Blank MAB, Antaki JF. Breast lesion elastography region of interest selection and quantitative heterogeneity: a systematic review and meta-analysis. Ultrasound Med Biol 2017; 43:387–397. doi:10.1016/j.ultrasmedbio.2016.09.002.
40. Moon JH, Hwang JY, Park JS, et al. Impact of region of interest (ROI) size on the diagnostic performance of shear wave elastography in differentiating solid breast lesions. Acta radiologica 2018; 59:657–663. doi:10.1177/0284185117732097.
41. Kapetas P, Woitek R, Clauser P, et al. A simple ultrasound based classification algorithm allows differentiation of benign from malignant breast lesions by using only quantitative parameters. Mol Imaging Biol 2018; 20:1053–1060. doi:10.1007/s11307-018-1187-x.
42. Bouzghar G, Levenback BJ, Sultan LR, et al. Bayesian probability of malignancy with BI-RADS sonographic features. J Ultrasound Med: Official J Am Inst Ultrasound Med 2014; 33:641–648. doi:10.7863/ultra.33.4.641.
43. del Cura JL, Elizagaray E, Zabala R, Legórburu A, Grande D. The use of unenhanced Doppler sonography in the evaluation of solid breast lesions. AJR Am J Roentgenol 2005; 184:1788–1794. doi:10.2214/ajr.184.6.01841788.
44. Safonov D, Rykhtik P, Shatokhina I, et al. Shear wave elastography: comparing the accuracy of ultrasound scanners using calibrated phantoms in experiment. Современные технологии в медицине 2017; 9:51.
45. Barr RG. The role of sonoelastography in breast lesions. Semin Ultrasound CT MR 2018; 39:98–105. doi:10.1053/j.sult.2017.05.010.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.acra.2023.01.024.