



# Automated classification of focal breast lesions according to S-detect: validation and role as a clinical and teaching tool

Mattia Di Segni<sup>1</sup> · Valeria de Soccio<sup>1</sup> · Vito Cantisani<sup>1</sup> · Giacomo Bonito<sup>1</sup> · Antonello Rubini<sup>1</sup> · Gabriele Di Segni<sup>2</sup> · Sveva Lamorte<sup>1</sup> · Valentina Magri<sup>1</sup> · Corrado De Vito<sup>3</sup> · Giuseppe Migliara<sup>3</sup> · Tommaso Vincenzo Bartolotta<sup>4</sup> · Alessio Metere<sup>5</sup> · Laura Giacomelli<sup>5</sup> · Carlo de Felice<sup>1</sup> · Ferdinando D'Ambrosio<sup>1</sup>

Received: 28 January 2018 / Accepted: 17 March 2018 / Published online: 21 April 2018  
© Società Italiana di Ultrasonologia in Medicina e Biologia (SIUMB) 2018

## Abstract

**Purpose** To assess the diagnostic performance and the potential as a teaching tool of S-detect in the assessment of focal breast lesions.

**Methods** 61 patients (age 21–84 years) with benign breast lesions in follow-up or candidate to pathological sampling or with suspicious lesions candidate to biopsy were enrolled. The study was based on a prospective and on a retrospective phase. In the prospective phase, after completion of baseline US by an experienced breast radiologist and S-detect assessment, 5 operators with different experience and dedication to breast radiology performed elastographic exams. In the retrospective phase, the 5 operators performed a retrospective assessment and categorized lesions with BI-RADS 2013 lexicon. Integration of S-detect to in-training operators evaluations was performed by giving priority to S-detect analysis in case of disagreement. 2×2 contingency tables and ROC analysis were used to assess the diagnostic performances; inter-rater agreement was measured with Cohen's *k*; Bonferroni's test was used to compare performances. A significance threshold of  $p=0.05$  was adopted.

**Results** All operators showed sensitivity > 90% and varying specificity (50–75%); S-detect showed sensitivity > 90 and 70.8% specificity, with inter-rater agreement ranging from moderate to good. Lower specificities were improved by the addition of S-detect. The addition of elastography did not lead to any improvement of the diagnostic performance.

**Conclusions** S-detect is a feasible tool for the characterization of breast lesions; it has a potential as a teaching tool for the less experienced operators.

**Keywords** CAD · Breast lesion characterization · Breast tumors · US-elastography · S-detect

## Riassunto

**Obiettivi** Valutare la performance diagnostica ed il potenziale come strumento didattico dell'S-detect nella valutazione delle lesioni mammarie focali.

**Metodi** Sono state arruolate 61 pazienti (età: 21–84 anni) con lesioni mammarie benigne in follow-up o con lesioni sospette per malignità candidate a biopsia. Lo studio è stato basato su una fase prospettica ed una retrospettiva. Nella fase prospettica, dopo il completamento dell'ecografia di base da parte di un senologo esperto, 5 operatori con differente livello di esperienza e differentemente dedicati alla senologia hanno eseguito l'esame elastosonografico. Nella fase retrospettiva, i 5 operatori hanno eseguito una valutazione e categorizzazione delle lesioni con BI-RADS 2013. L'integrazione dell'S-detect con la valutazione degli operatori in formazione è stata eseguita dando priorità all'analisi del software in caso di discordanza. Sono state impiegate le tabelle di contingenza 2×2 e le curve ROC per valutare le performance diagnostiche; la concordanza tra gli operatori è stata misurata con il test *k* di Cohen; il test di Bonferroni è stato impiegato per comparare le performance. È stata adottata una soglia di significatività pari a  $p=0.05$ .

**Risultati** Tutti gli operatori hanno dimostrato una sensibilità > 90% e specificità variabile (50–75%); l'S-detect ha dimostrato una sensibilità > 90% e specificità del 70,8%, con concordanza con gli operatori compresa tra moderata e buona. Le specificità più basse sono state aumentate dall'aggiunta dell'S-detect. L'aggiunta dell'elastosonografia non ha determinato aumento delle performance diagnostiche.

**Conclusioni** L'S-detect è uno strumento impiegabile nella caratterizzazione delle lesioni mammarie ed è un potenziale strumento didattico per gli operatori meno esperti.

## Introduction

US plays a pivotal role in breast imaging as a first-line tool for breast lesion characterization. However, it suffers from operator dependence [1].

The American College of Radiology (ACR) addressed its use with the proposal of the Breast Imaging Report And Data System, a lexicon aimed at easing communication between different specialties about morphology, level of suspicion and suggested management; at present, 5 BI-RADS versions have been published [2]. Compared to the former edition, the latest one includes changes such as the introduction of special cases, changes in the description of surrounding tissues, calcification and vascularity [3–5].

However, despite the extensive application of this lexicon, controversies regarding some topics remain, especially the question of how to apply the subcategorization of the suggested BI-RADS.

Furthermore, other tools, such as US-elastography [6], contrast-enhanced ultrasound (CEUS) [7, 8] and computer-aided diagnosis (CAD) systems, were developed with the aim of improving US performance.

Elastography allows the assessment of lesion stiffness after the application of a stressor force and is classified according to the stressor force, the source of the force and outcomes [9, 10]; it has been proved to be effective in several fields [11–13].

CAD systems work in three phases: image processing, segmentation and feature extraction [14]; they are classified according to the algorithms employed in each phase.

According to some authors, these systems can potentially improve breast lesion classification in terms of performance and operator dependence [14, 15].

Among these systems, S-detect is a software that Samsung Healthcare developed and is based on the deep learning algorithm, which performs lesion segmentation, feature analysis and descriptions according to the BI-RADS 2003 or BI-RADS 2013 lexicon, suggesting dichotomic categorization. In addition to its role as a possible adjunct tool for breast lesion characterization it could be used as a teaching tool.

The primary aim of the present study was to assess the diagnostic performance of S-detect in breast lesion characterization and its potential role as a teaching tool. Its secondary aim was to assess the diagnostic performance of elastography in the differentiation of breast lesions as an adjunct tool to the BI-RADS 2013 lexicon.

## Materials and methods

The Institutional Ethical Committee approved this prospective study. Each patient gave informed consent.

Between July 2016 and June 2017, 65 patients aged between 21 and 84 years (mean age 51 years) underwent US examination using UGEO RS80A machinery (Samsung Healthcare, South Korea) with 3–16 MHz or 3–12 MHz linear array probes—in case of abundant breast volume or deep seated lesions—including elastographic exams and analysis with S-detect software.

Inclusion criteria were as follows:

- Patients with follow-up in progress (i.e., within 2 years of lesion detection) or who previously underwent cytology or biopsy sampling due to likely benign lesions;
- Patients who were candidates for biopsy due to lesions suspicious for malignancy.

Exclusion criteria were as follows:

- Pregnancy;
- Lactation;
- Neoadjuvant chemotherapy in progress or having been completed less than 2 months previously;
- Radiotherapy in progress or having been completed less than 3 months previously;
- Insufficient documentation.

The ultrasound examination was performed in addition to the conventional diagnostic and therapeutic path, with no interference determined on it by the exam itself.

The study was organized into two phases:

- The prospective acquisition of US, elastographic and S-detect images;
- Retrospective image evaluation.

A radiologist with 32 years of experience in breast imaging (CDF) performed the prospective acquisition of baseline images.

Subsequently, a second experienced operator with 18 years of experience in US (CV) and four radiology residents who were in different years of the residency program and had different levels of dedicated breast imaging experience (a 5th-year resident with limited experience in breast imaging, a 2nd-year resident with deeper experience in

breast imaging, a 3rd-year resident with limited experience in breast imaging, a 1st-year resident with deeper experience in breast imaging, D.S.M, B.G, R.A, d.S.V) completed the procedure with elastographic assessment that employed the compression technique and qualitative and semi-quantitative processing with Elastoscanner software (Samsung Healthcare, South Korea). Elastography was performed using the quasi-static technique and free-hand compression with the probe, guided by a quality indicator, and care was taken not to apply pre-compression and not to allow the target lesion to slide away from the field of view. When an adequate compression set was obtained, the lesion was characterized using the subsequent derived data:

- Qualitative elastography according to the BI-RADS 2013 lexicon [2], categorized in the voice Elasticity Assessment and expressed with 3 options: Soft, Intermediate, Hard;
- The qualitative 5-point scale by Tsukuba et al. [16], determined by the lesion's stiffness: score 1—completely soft lesion; score 2—up to 50% of the area of the lesion was stiff; score 3—mostly not deformable; score 4—completely stiff lesion; score 5—stiffness also extended beyond lesion area;
- Semi-quantitative assessment using the strain ratio (SR) between the lesion and breast adipose tissue, with the positioning of two ROIs.

During the same session, the expert operator performed an additional scan, which represented the input image for S-detect software (Samsung Healthcare, South Korea): having selected the image, the operator activated the software and guided the lesion segmentation phase—which the operator guided or which was performed automatically. When the analysis was complete, the software computed its segmentation, description and classification proposals; of these proposals, in the case of multiple options, the one with the most adequate segmentation and characterization was chosen.

The description was automatic for the subsequent parameters encoded in the BI-RADS 2013 lexicon [2]: shape, orientation, margins, pattern, posterior acoustic features; for the other features, manual insertion was requested. When these steps were completed, a structured report was “assigned”.

Between 2 and 6 weeks later, the radiologist with 18 years of experience and the residents retrospectively reviewed images and gave their assessments, blinded to the pathology results.

Nodule features and categories were expressed according to the BI-RADS 2013 lexicon [2]; in the case of BI-RADS 4 assignment, a subcategorization was performed according to the scheme that Jales et al. proposed [17].

Pictures, structured reports, operators' assessments, pathological or follow-up data for the nodules and assessments

were stored in an appositely developed integrated system that comprised a graphic interface and was based on an SQLite encoded relational database, which allowed the filling-in, storage and retrieval of the data.

## Statistical analysis

Data were analyzed, using Stata (Stata v 13.0, StataCorp Lt, College Station, TX, USA) to evaluate:

- The operators' performance with  $2 \times 2$  contingency tables and receiver operator curves (ROC) against the histology results;
- The S-detect performance with  $2 \times 2$  contingency tables and ROC curves against the histology results;
- The differences in performance between S-detect and each operator with the test for the equality of the ROC areas against the histology results;
- The agreement between the operators and S-detect using Cohen's kappa;
- The differences in lesion strain ratio according to the Mann–Whitney  $U$  test;
- Where the strain ratio was concerned, the best cut-off through the ROC-curve analysis of the histology results;
- The performance of the various elastographic outcomes upon integration with conventional US.

Category dichotomization was applied as follows: BI-RADS categories 2 and 3 were considered benign, whereas BI-RADS categories 4a–5 were considered malignant.

To integrate conventional US and elastography, we used the following approaches:

- For strain ratio: if the lesion showed an SR equal to or greater than the cut-off point, the category number was increased by one point;
- For BI-RADS elasticity assessment, hard lesions experienced a category number increase of 1, whereas soft and intermediate lesions' category numbers remained unchanged;
- For Tsukuba maps, lesions with scores of 4 or 5 experienced a change in the assigned category by a factor of 1 (e.g., from BI-RADS 3 to BI-RADS 4a).

The significance of the changes related to the addition of the various elastography modalities was assessed using the test for the equality of ROC areas, with the unmodified performance serving as a gold standard ROC curve against the histology results.

Inter-operator concordance was determined with Cohen's kappa test and interpreted according to the subsequent scale, as Landis and Koch described it [18]:

- Excellent:  $\kappa > 0.8$ ;
- Good:  $\kappa = 0.61\text{--}0.8$ ;
- Moderate:  $\kappa = 0.41\text{--}0.6$ ;
- Fair:  $\kappa = 0.21\text{--}0.4$ ;
- Slight:  $\kappa \leq 0.2$ .

To assess the potential of S-detect as a teaching tool, all the residents' readings that classified lesions as BI-RADS 4a were compared with S-detect assessments. When there was disagreement between the classifications, the S-detect classification replaced the operators' classification. Subsequently, the performance was re-evaluated with  $2 \times 2$  contingency tables and receiver operator curves (ROC) and compared to the baseline performance with the test for the equality of the ROC areas against the histology results.

For all the statistical tests, a value of  $p$  that was less than 0.05 was considered significant.

## Results

Sixty-eight nodules were analyzed in 61 patients. They ranged between 10 and 48 mm in size; of these nodules, 44 were malignant and 24 were benign. Out of 44 malignancies,

37 were infiltrating ductal carcinomas (IDC), 3 were ductal carcinomas in situ (DCIS), 3 were infiltrating lobular carcinomas and 1 was a granular cell tumor. Out of 24 benign lesions, there were 12 fibroadenomas (one of them was characterized through a biopsy and one was characterized through cytology and was being followed up), 1 phyllodes tumor, 2 hamartomas (characterized by their sonographic appearance and followed up), 7 foci of biopsy-proven sclerosing adenosis and/or fibrocystic mastopathy and 2 abscesses (sonographically characterized and followed up).

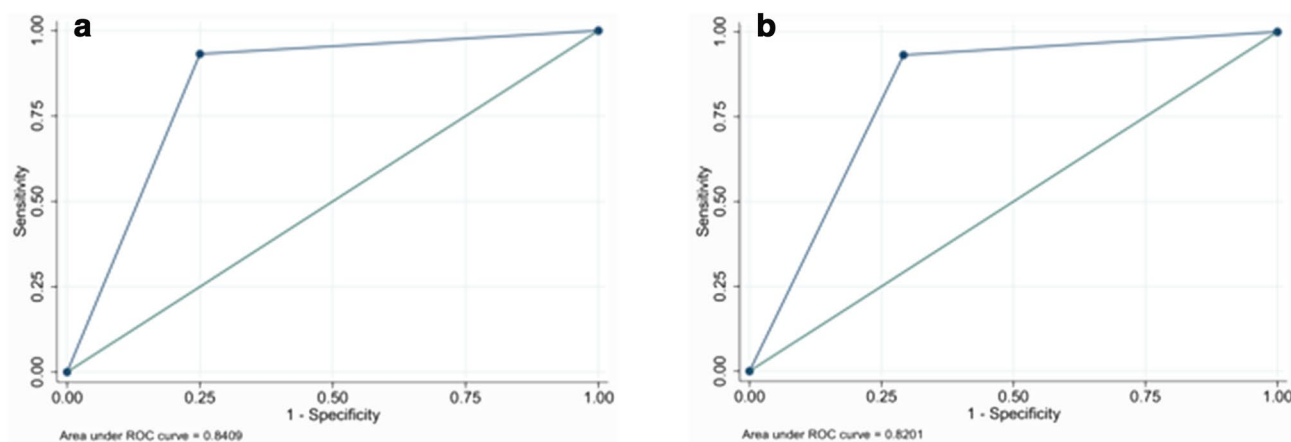
The diagnostic performance of S-detect and that of the expert operator are shown in Table 1.

In the overall assessment of benign vs malignant lesions, a  $k = 0.616$  (95% CI 0.376–0.856) was obtained; this corresponded to substantial agreement.

According to the test for the equality of the ROC areas, the difference in diagnostic performance was not significant (ROC 0.8201 vs 0.8409;  $p = 0.751$ ).

The ROC curves for S-detect and the expert operator are shown (Fig. 1).

S-detect showed similar sensitivity to all of the operators-in-training (ROC curves shown in Fig. 2), whereas its specificity was higher than that of the non-dedicated operators



**Fig. 1** ROC curves for expert operator and for S-detect: **a** expert operator; **b** S-detect

**Table 1** Diagnostic performance of S-detect and the expert operator

Operator	Sensitivity (%)—95% CI	Specificity (%)—95% CI	PLR (%)—95% CI	NLR (%)—95% CI	ROC (%)—95% CI	PPV (%)—95% CI	NPV (%)—95% CI
S-detect	91.1 (78.8–97.5%)	70.8 (48.9–87.4%)	3.12 (1.66–5.87)	0.13 (0.05–0.33)	0.82 (0.71–0.91)	85.4 (72.2–93.9%)	81.0 (58.1–94.6%)
Expert operator	93.2 (81.3–98.6%)	75.0 (53.3–90.2%)	3.73 (1.86–7.49)	0.09 (0.03–0.28)	0.84 (0.74–0.94)	87.2 (74.3–95.2%)	85.7 (63.7–97.0%)

PLR positive likelihood ratio, NLR negative likelihood ratio, ROC receiver operator curve, PPV positive predictive value, NPV negative predictive value

**Table 2** Diagnostic performance of S-detect and in-training operators

Operator	Sensitivity (%)—95% CI	Specificity (%)—95% CI	PLR—95% CI	NLR—95% CI	ROC area—95% CI	PPV (%)—95% CI	NPV (%)—95% CI
S-detect	91.1 (78.8–97.5%)	70.8 (48.9–87.4%)	3.12 (1.66–5.87)	0.13 (0.05–0.33)	0.81(0.71–0.91)	8.54 (72.2–93.9%)	81.0 (58.1–94.6%)
5th year with limited experience <sup>a</sup>	97.7% (88–99.9%)	54.2 (32.8–74.4%)	2.13 (1.38–3.30)	0.04 (0.01–0.30)	0.76 (0.66–0.86)	79.6 (66.5–89.4)	92.9 (66.1–99.8%)
2nd year with deeper experience <sup>a</sup>	95.5 (84.5–99.4%)	70.8 (48.9–87.4%)	3.27 (1.75–6.13)	0.06 (0.02–0.25)	0.83 (0.73–0.93)	85.7 (72.8–94.1)	89.5 (66.9–98.7)
3rd year with limited experience <sup>a</sup>	97.8 (88.2–99.9)	50 (29.1–70.9%)	1.96 (1.31–2.92)	0.04 (0.01–0.32)	0.74 (0.63–0.84)	78.6 (65.6–88.4%)	92.3 (64–99.8%)
1 anno with deeper experience <sup>a</sup>	100 (92–100%)	50 (29.1–70.9%)	2.0 (1.34–2.98)	0.00	0.75 (0.65–0.85)	78.6 (65.6–88.4%)	100 (73.5–100%)

PLR positive likelihood ratio, NLR negative likelihood ratio, ROC receiver operator curve, PPV positive predictive value, NPV negative predictive value

<sup>a</sup>Experience in breast imaging

(see Table 2) but slightly inferior to that of the dedicated expert/operator-in-training.

The difference in ROC area between S-detect and the operators-in-training was not statistically significant when compared to each operator (see Table 3).

The agreement between the operators was not significant, whereas the agreement between S-detect and the operators varied between moderate and good (see Tables 4 and 5) statistical significance.

To measure how S-detect could determine a change in diagnostic performance for operators-in-training, the lesions categorized as BI-RADS 4a were selected and their assessments were compared with the S-detect analyses; where disagreement occurred, the S-detect results replaced the trainees’ assessments and were included in the performance analysis of the whole sample.

A statistically significant increase in diagnostic performance was noted, especially for 2 of the operators with the lowest performance levels (see Table 6 for performance and Table 7 for significance). In comparison, this change was not significant for the operator-in-training with the deepest experience in breast imaging. In addition, a non-significant change was noted for another of the operators. It was related to the change in categorization in only 4 cases (Fig. 3).

Figure 3 shows the ROC curves of the addition and integration of S-detect.

Despite a trend that saw malignancies increase with stiffness (Fig. 4), the addition and integration of elastography did not determine any improvement in diagnostic performance (ROC curves shown in Fig. 5). In fact, the diagnostic performance was jeopardized in some cases (see Tables 8 and 9). This was the case despite the statistically significant distribution of the strain ratio parameter between benign and malignant lesions, as assessed using the Mann–Whitney test ( $z = -3427, p = 0.0006$ ).

The cut-off value for the strain ratio was 1.765, which corresponded to an estimated sensitivity of 0.76 and an estimated specificity of 0.75 (Fig. 4).

The ROC curves of the elastographic indicators are shown in Fig. 5.

**Table 3** Significance of the differences ROC areas between the evaluation of the operators and S-detect readings

Comparison	<i>p</i>
Expert operator vs. S-detect	0.751
5th year with limited experience* vs. S-detect	0.000
2nd year with deeper experience vs. S-detect	0.831
3rd year with limited experience* vs. S-detect	0.151
1st year with deeper experience vs. S-detect	0.206

The addition of the elastographic data did not increase the diagnostic performance of the trainees. Rather, it jeopardized it as it had jeopardized the Strain Ratio (Table 10).

An example case is provided in Fig. 6.

### Discussion

The development and publication of BI-RADS began after the recognition of the need for a globally shared lexicon that would allow the sharing and clear expression of morphology, the operator’s judgment and the strategy considered to be the best advised in the assessment of breast lesions [2].

**Table 4** Agreement between the expert operator and in-training operators

Comparison	Observed agreement	$\kappa$	Standard error	Probability
Expert vs. 2nd year experienced	58.82%	0.0094	0.1210	0.4691
Expert vs. 5th year with limited experience	66.18%	0.1272	0.1169	0.1382
Expert vs. 1st year experienced	60.29%	− 0.0552	0.1135	0.6865
Expert vs. 3rd year with limited experience	58.82%	− 0.0781	0.1154	0.7509

**Table 5** Agreement between S-detect and the operators

Comparison	Observed agreement (%)	$\kappa$	Standard error	Probability
S-detect vs. expert operator	83.82	0.6160	0.1212	0.000
S-detect vs. 5th year with limited experience	85.29	0.6119	0.1179	0.000
S-detect vs. 2nd year experienced	89.71	0.7484	0.1212	0.000
S-detect vs. 3rd year with limited experience	82.61	0.5400	0.1146	0.000
S-detect vs. 1st year experienced	85.29	0.5991	0.1149	0.000

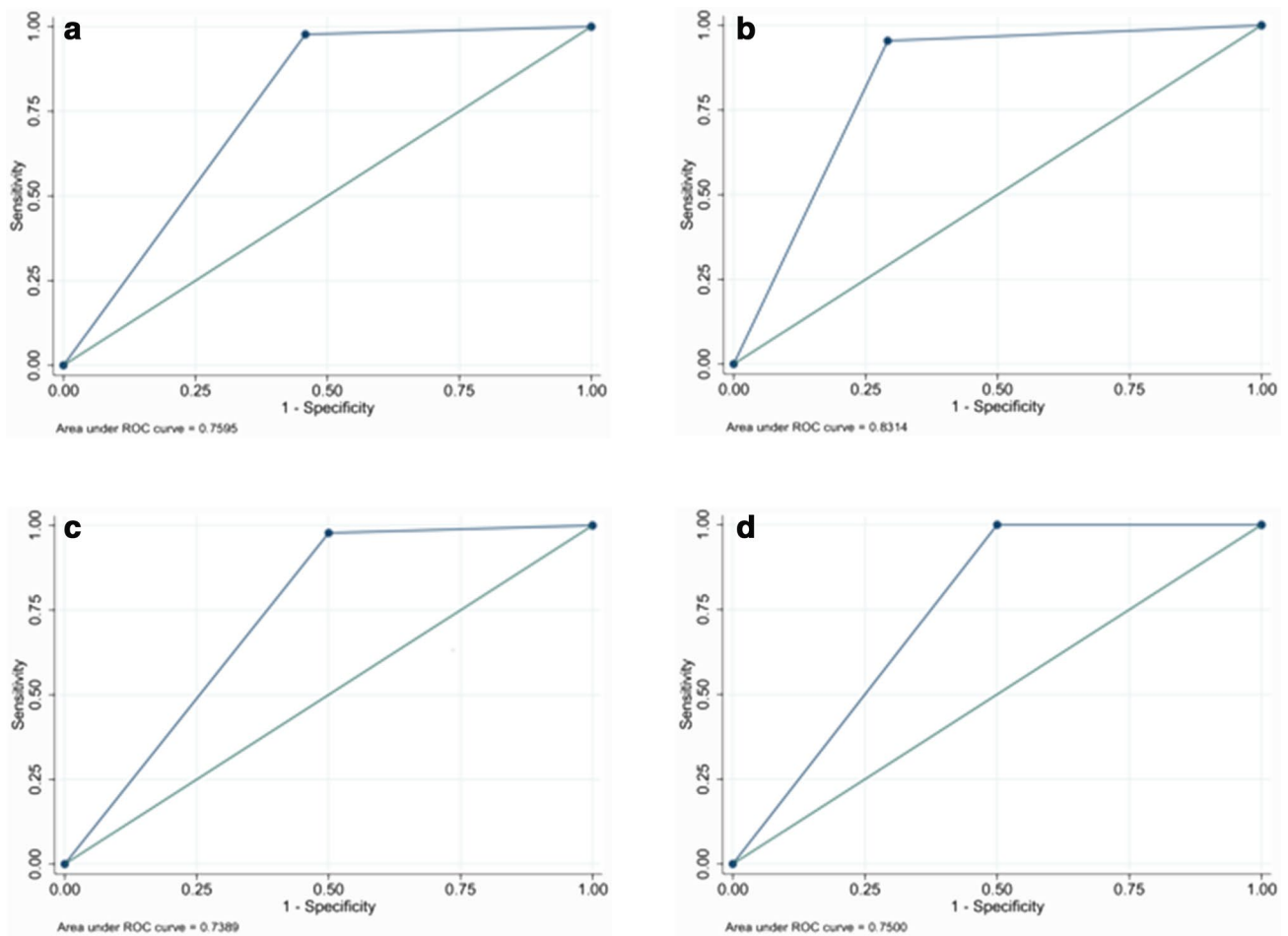
**Table 6** Diagnostic performance of the in-training operators after the integration of S-detect in the assessment of BI-RADS 4a lesions

Operator	Se	Sp	ROC area	PLR	NLR	PPV	NPV
5th year with limited experience	95.5% (84.5–99.4%)	75.0% (53.3–90.2%)	0.85 (0.76–0.95)	3.82 (1.90–7.66)	0.06 (0.02–0.24)	87.5% (74.8–95.3%)	90.0% (68.3–98.8%)
2nd year experienced	93.2% (81.3–98.6%)	79.2% (57.8–92.9%)	0.86 (0.77–0.95)	4.47 (2.04–9.80)	0.09 (0.03–0.26)	89.1% (76.4–96.4%)	86.4% (65.1–97.1%)
3rd year with limited experience	91.1% (78.8–97.5%)	70.8% (48.9–87.4%)	0.81 (0.71–0.91)	3.12 (1.66–5.87)	0.13 (0.05–0.33)	85.4% (72.2–93.9%)	81.0% (58.1–94.6%)
1st year experienced	97.7% (88.0–99.9%)	75.0% (53.3–90.2%)	0.86 (0.77–0.95)	3.91 (1.95–7.83)	0.03 (0.00–0.21)	87.8% (75.2–95.4%)	94.7% (74.0–99.9%)

PLR positive likelihood ratio, NLR negative likelihood ratio, ROC receiver operator curve, PPV positive predictive value, NPV negative predictive value

**Table 7** Significance of the increase in performance for in-training operators

Operator	ROC area—before	ROC area—after	$\chi^2$	p
5th year with limited experience	0.7595	0.8523	4.4813	0.0343
2nd year experienced	0.8314	0.8617	0.9571	0.33
3rd year with limited experience	0.7389	0.8097	2.3377	0.12
1st year experienced	0.7500	0.8636	5.9586	0.0146



**Fig. 2** ROC curve analysis for operators-in-training: **a** 5th-year with limited experience in breast imaging; **b** 2nd-year with deeper experience in breast imaging; **c** 3rd-year with limited experience in breast imaging; **d** 1st-year with deeper experience in breast imaging

Compared to the former edition, the latest includes changes such as the introduction of special cases, changes in the description of surrounding tissues, calcification and vascularity [3–5].

According to Xiao et al. [19], the use of 2013 criteria resulted in 100% sensitivity, 17.4% specificity, 46.8% PPV, 100% NPV and 0.867 ROC.

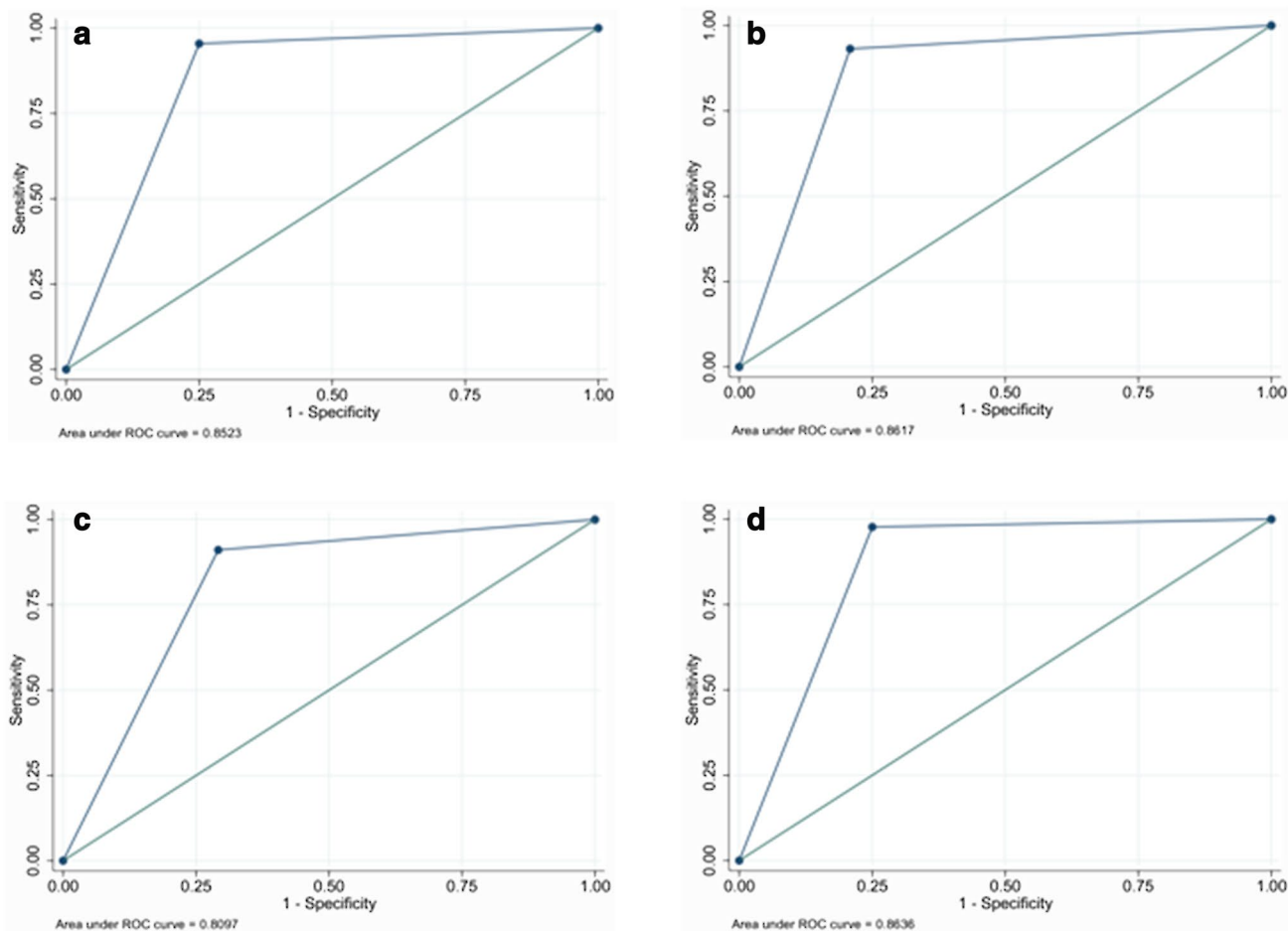
Conversely, according to Fleury et al. [20], using the fifth edition, two radiologists with different experience levels achieved the following results: 94.4% sensitivity, 49.2% specificity, 60.7% PPV, 91.4% NPV, 69.8% accuracy and 0.887 ROC for the more experienced operator and 94.4% sensitivity, 55.4% specificity, 63.8% PPV, 92.3% NPV, 73.1% accuracy and 0.901 ROC for the second operator.

In our experience, despite the limited sample size, when an experienced radiologist operated the US scanner, it showed high performance: 93.2% sensitivity, 75% specificity, 87.2% PPV and 85.7% NPV. This demonstrated agreement with the findings in the literature.

Computer-aided diagnosis and computer-aided classification systems work in three phases: image processing, segmentation and feature extraction [14]; they are classified according to the algorithms employed in each phase.

According to some authors, these systems can improve breast lesion classification in terms of performance and operator dependence [14].

In a previous experiment, using the evaluations of two breast radiologists as a reference and the BI-RADS 2003 lexicon, Moon et al. [21] assessed a 244-neoplasm sample that had been characterized by biopsy to compare the diagnostic abilities of two CAD systems: one conventional and the other with BI-RADS parameter quantification. The latter system considered malignant any lesion with at least one suspicious feature. According to the study, the difference in performance appeared significant: the results included 84% specificity, 87% accuracy, 73% PPV and 97% NPV for sensitivity values set at 95% versus 60% specificity, 71%



**Fig. 3** Performance of the operators-in-training after the integration of S-detect in ambiguous cases (BI-RADS 4a): **a** 5th-year with limited experience; **b** 2nd-year with deeper experience; **c** 3rd-year with limited experience; **d** 1st-year with deeper experience

**Table 8** Comparison of the ROC of morphology assessment before and after the addition of elastographic indicators

Modality	ROC area	Chi square	Probability
Morph-BIRADS	0.8409	0.0401	
BI-RADS + Tsukuba Map	0.8201	0.0511	1.0
BI-RADS + strain ratio	0.7784	0.0540	0.2795
BI-RADS + elasticity assessment	0.7784	0.0540	0.2795

accuracy, 53% PPV and 96% NPV for the conventional CAD system.

In another study [22], 626 US images of pathologically characterized lesions were analyzed using two CAD systems and compared to the retrospective assessments of two radiologists; the results in the subsequent table were obtained with a significant increase in performance over that of the radiologists and were indicative of substantial agreement (Table 11).

In a previous study concerning S-detect by Kim et al. [23], 192 breast lesions were retrospectively assessed by a radiologist and by the software and classified using BI-RADS 2003 criteria. Two cut-offs were applied: one lay between BI-RADS 3 and 4 and the other between BI-RADS 4a and 4b—the latter was only applied to the radiologist’s assessment. In this study, S-detect showed 79.2% sensitivity, 65.8% specificity, 58.3% PPV, 84% NPV, 70.8% accuracy and 0.725 ROC. It was significantly different from the operator’s assessment for all the parameters when a cut-off between BI-RADS 3 and 4 was employed and for specificity, PPV and accuracy when a cut-off between BI-RADS 4a and 4b was used. In addition, the agreement with the operator varied in degree between moderate and scarce.

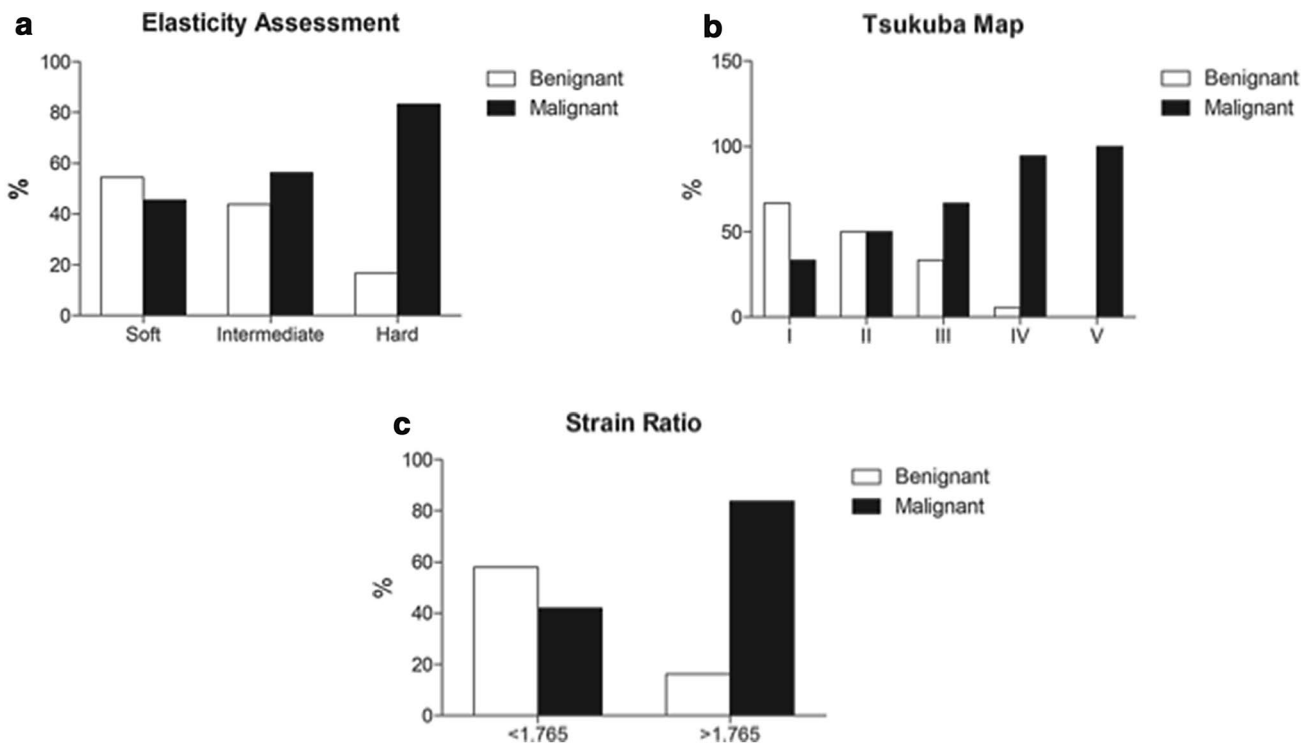
According to Cho et al. [24], when the performance of S-detect was compared to that of 2 radiologists with different levels of experience, the software alone showed lower sensitivity and higher specificity than the more experienced radiologist and the less experienced radiologist—72.2 vs. 94.4% and 94.4% sensitivity and 90.8 vs. 49.2% and 55.4%



**Table 9** Comparison of the performance of the different elastographic indexes in addition to the elastographic evaluation according to the 2×2 contingency table

Modality	Sensitivity (%)—(95% CI)	Specificity (%)—(95% CI)	ROC area	PLR (%)—(95% CI)	NLR (%)—(95% CI)	PPV (%)—(95% CI)	NPV (%)—(95% CI)
BIRADS + elasticity	93.2 (81.3–98.6)	62.5 (40.6–81.2)	0.78 (0.67–0.88)	2.48 (1.47–4.19)	0.11 (0.04–0.34)	82.0 (68.6–91.4)	83.3 (58.6–96.4)
BIRADS + Tsukuba	93.2 (81.3–98.6)	70.8 (48.9–87.4)	0.82 (0.72–0.92)	3.19 (1.70–5.99)	0.10 (0.03–0.30)	85.4 (72.2–93.9)	85.0 (62.1–96.8)
BIRADS + Strain	93.2 (81.3–98.6%)	62.5 (40.6–81.2)	0.78 (0.67–0.88)	2.48 (1.47–4.19)	0.11 (0.04–0.34)	82.0 (68.6–91.4)	83.3 (58.6–96.4)
Morph_BI-RADS	93.2 (81.3–98.6%)	75.0 (53.3–90.2%)	0.84 (0.74–0.94)	3.73 (1.86–7.49)	0.09 (0.03–0.28)	87.2 (74.3–95.2)	85.7 (63.7–97.0)

*PLR* positive likelihood ratio, *NLR* negative likelihood ratio, *ROC* receiver operator curve, *PPV* positive predictive value, *NPV* negative predictive value

**Fig. 4** Distribution of the lesions according to elastographic indicators: **a** elasticity assessment; **b** Tsukuba Map; **c** strain ratio (cut-off 1.765)

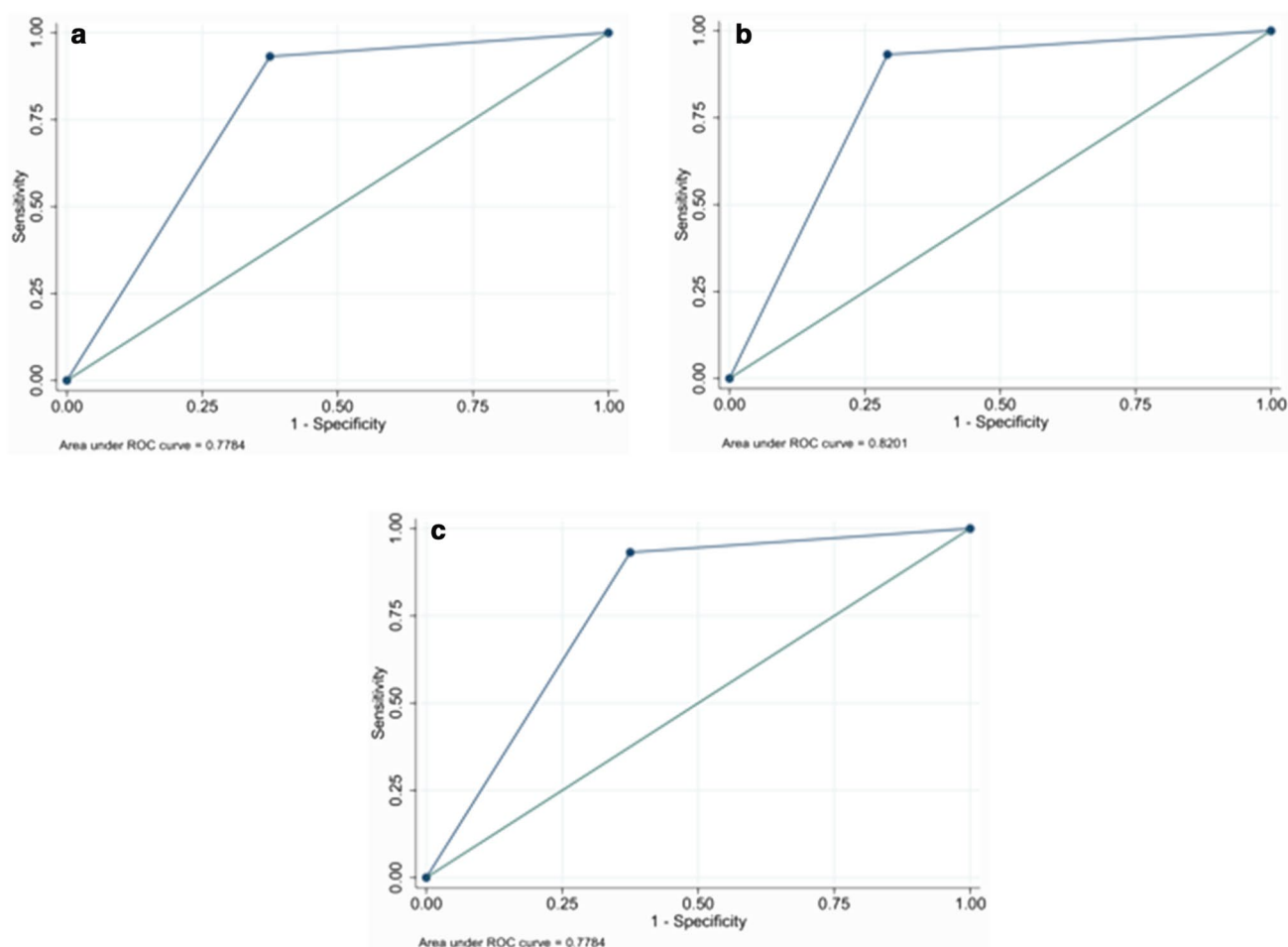
specificity. However, when the operators' readings and the software assessments were combined, a significant increase in specificity and PPV was noted, with no statistically significant detrimental effect on sensitivity. When categorization was analyzed after dichotomization, moderate agreement between S-detect and the operators was observed as well.

Even when we applied different lexicon and categorization modalities from those employed in the study by Kim et al., our experience showed that S-detect had 91.1% sensitivity, 70.8% specificity, 85.4% PPV and 81% NPV, with a 0.81 ROC. In terms of performance and agreement, our study was consistent with that of Kim et al. However, our results partially disagreed with those of Cho et al. in terms

of sensitivity: it remained substantially unaffected in our study. This difference may be related to the composition of our sample.

In addition, inter-observer agreement with the operators, especially the most expert and dedicated—appeared to be good.

When cases categorized as ambiguous (i.e., BI-RADS 4a) were selected, the addition of S-detect assessment in place of the operators' assessment in cases of disagreement led to a significant increase in the diagnostic performance of operators-in-training with less experience and lower performance. This may show a potential role for S-detect as a teaching tool as the program can provide an iconographic



**Fig. 5** ROC curves concerning the integration of various elastographic indicators: **a** elasticity assessment; **b** Tsukuba map; **c** strain ratio

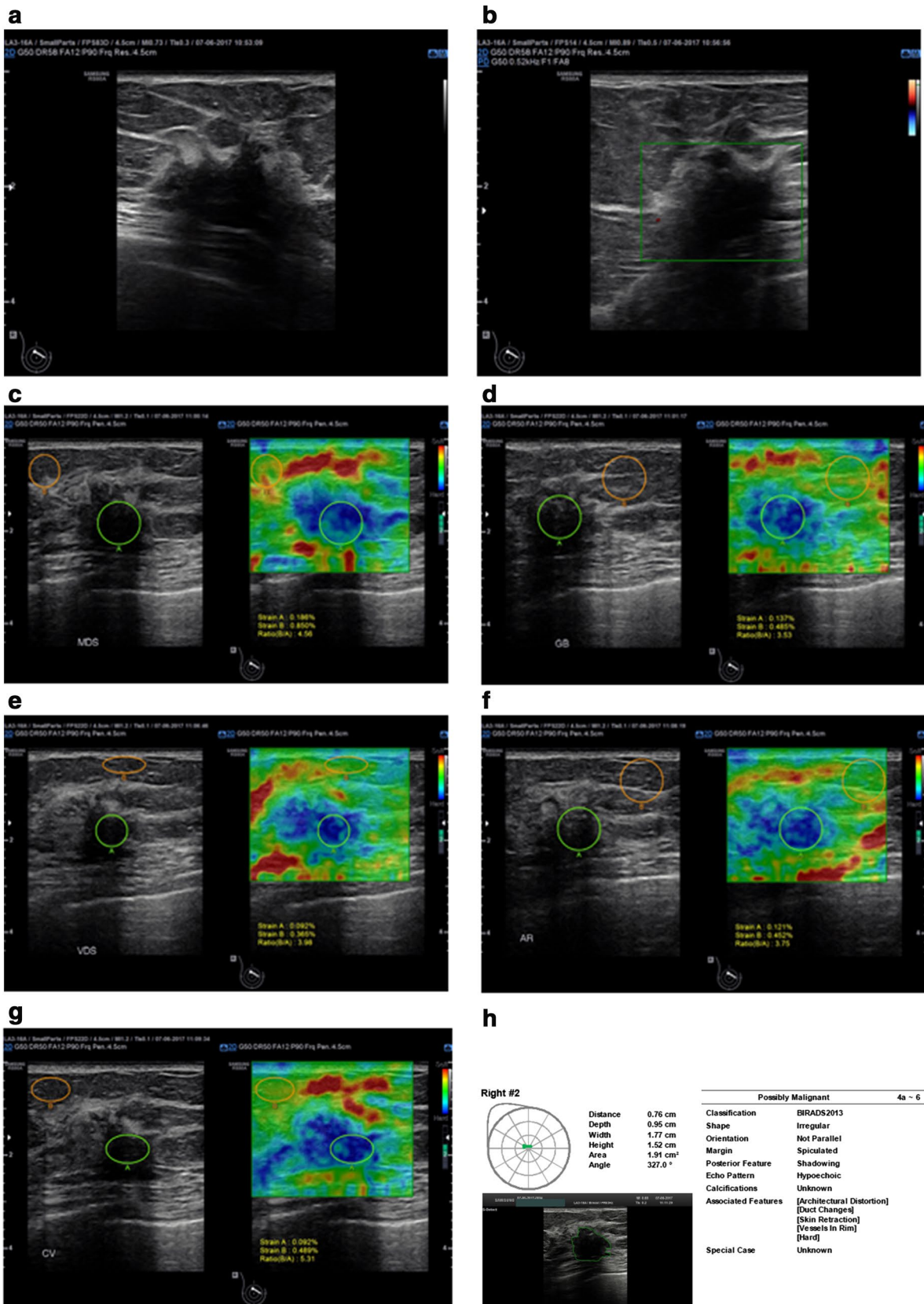
**Table 10** Comparison between the ROC areas of elastography for in-training operators

Operator	BIRADS_morph.	BIRADS + elasticity assessment		BIRADS + Tsukuba map		BIRADS + strain ratio	
	ROC – std err	ROC – std err	$\chi^2 - p$	ROC – std err	$\chi^2 - p$	ROC – std err	$\chi^2 - p$
5th year with limited experience	0.7595–0.0532	0.6970 (0.053)	0.2795	0.7595 0.0532	1.0	0.7386 0.0534	1.0
2nd year experienced	0.8314–0.05	0.8314 (0.050)	1.000	0.8314 (0.050)	1.0	0.7898 (0.053)	0.5927
3rd year with limited experience	0.7389–0.533	0.7389 (0.053)	1.0	0.7389 (0.053)	1.0	0.781	0.5927
1st year experienced	0.7500–0.0521	0.7500 (0.052)	1.0	0.7292 (0.052)	1.0	0.7083 0.0514	0.5927

and written categorization reference for trainees and improve the specificity of less experienced operators, who tend to overestimate lesion features.

Elastography allows the assessment of lesion stiffness. Its role in addressing the diagnosis of breast lesions has been the subject of active discussion as it is considered to be a tool in the effort to overcome US limitations [6, 25]. Even if the fifth edition of BI-RADS includes the qualitative assessment of breast lesions as an additional feature

to morphology [2], its role in breast imaging is still being strongly debated and ACR has made no explicit statement to address the use of the elastographic data. To be specific, according to EFSUMB [26] and WFUMB guidelines [27], elastography can be used to increase the suspicion category of lesions with no suggestive morphological features (i.e., a change from BI-RADS 3—according to the fourth edition—to BI-RADS 4) and to distinguish solid lesions from



**Fig. 6** BI-RADS 5 lesion: **a** a parallel, hypoechoic, irregular lesion that has angular margins and posterior shadowing and is associated with architectural distortion; **b** Scarce vascularity at CDUS; **c-g** Elastographic features with qualitative and semiquantitative stiffness assessment; **h** Structured report according to S-detect

**Table 11** Performance of radiologists and CAC systems according to Shen et al. [26]

Assessment of	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	ROC	Agreement
Radiologists	97.72	50.61	51.57	97.63	0.9431	–
Basic CAC	98.3	53.8	53.47	98.65	0.9435	0.644
Weighted CAC	98.17	53.8	53.47	98.65	0.948	0.644

cysts, whereas its use is not advised in the reduction of the category from 4 to 3.

Previous efforts have described elastography's role in improving sensitivity.

In their evaluation of 12 studies with 2087 masses, Sadigh et al. [28] achieved 88% cumulative sensitivity (95% CI 93–99%), 72% cumulative specificity (95% CI 31–96%), 92% HSROC (95% CI 90–93%), a positive likelihood ratio of 5.38 (95% CI 132–1674) and a negative likelihood ratio of 0.04 (95% CI 004–014). In another paper, Sadigh et al. [29] analyzed 1412 masses from 5 studies and compared elastography with B-mode US. For elastography, they achieved 78% cumulative sensitivity (95% CI 63–88%) and 92% cumulative specificity (95% CI 85–96%), whereas for B-mode US, they achieved 96.6% cumulative sensitivity (95% CI 93–97%) and 65% cumulative specificity (95% CI 42–82%), with no statistically significant difference in modalities reported when lesions were layered according to size.

Also, when applied to BI-RADS fifth edition, improved performance was noted. According to Hao et al. [30], the use of elastography allowed the increase of the ROC from 0.866 to 0.886, leading to 97% sensitivity, 80.6% specificity, 76.7 PPV, 97.7% NPV and 87.1% accuracy. The increase was significant compared to the performance of baseline US alone.

According to Fleury [20], in a retrospective assessment of 929 breast lesions, sensitivity increased to 95.9%, specificity to 80.65%, PPV to 80.65% and NPV to 98.67%, and accuracy reached 91.39%. Of these parameters, specificity increased by 8% and PPV increased by 10%. Accuracy also increased.

According to Xiao et al., [19], in the assessment of lesions < 1 cm, the combined use of US and elastography with Zhi's qualitative score applying a cut-off between 3 and 4 resulted in the achievement of 97.7% sensitivity, 46.3% specificity, 57% PPV and 96.6% NPV. Moreover, there was a 17.4% increase in specificity.

Conversely, in our experience, the additional use of elastography did not show any improvement in the characterization of breast lesions compared to the use of morphology alone.

The limits of our study were related to the limited sample size and the selection of a sample with a high prevalence of malignant lesions. The latter point was very often applicable to previous cases though it was accorded less weight in those cases than it was in our report. An additional limit

was related to the retrospective phase of our study, which characterized the second part of our protocol.

In conclusion, S-detect is a feasible tool for the characterization of breast lesions; it has potential as a teaching tool for less experienced operators.

**Funding** No funding financed this study.

## Compliance with ethical standards

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.


**Conflict of interest** Cantisani V. is lecturer for Bracco and Samsung Healthcare; Bartolotta is lecturer for Samsung Healthcare.

## References

1. Hooley RJ, Scoutt LM, Philpotts LE (2013) Breast ultrasonography: state of the art. *Radiology* 268:642–659. <https://doi.org/10.1148/radiol.13121606>
2. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA (2013) ACR BI-RADS® Atlas, breast imaging reporting and data system, 5th edn. American College of Radiology, Reston
3. Rao AA, Feneis J, Lalonde C, Ojeda-Fournier H (2016) A pictorial review of changes in the BI-RADS fifth edition. *RadioGraphics* 36:623–639. <https://doi.org/10.1148/rg.2016150178>
4. Lee J (2017) Practical and illustrated summary of updated BI-RADS for ultrasonography. *Ultrasonography* 36:71–81. <https://doi.org/10.14366/usg.16034>
5. Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE (2017) BI-RADS® fifth edition: a summary of changes. *Diagn Interv Imaging* 98:179–190. <https://doi.org/10.1016/j.diii.2017.01.001>
6. Goddi A, Bonardi M, Alessi S (2012) Breast elastography: a literature review. *J Ultrasound* 15:192–198. <https://doi.org/10.1016/j.jus.2012.06.009>
7. Drudi F, Giovagnorio F, Carbone A, Ricci P, Petta S, Cantisani V, Ferrari F, Marchetti F, Passariello R (2006) Transrectal colour Doppler contrast sonography in the diagnosis of local recurrence after radical prostatectomy—comparison with MRI. *Ultraschall in der Medizin Eur J Ultrasound* 28:146–151. <https://doi.org/10.1055/s-2006-926583>
8. Cantisani V, Ricci P, Erturk M, Pagliara E, Drudi F, Calliada F, Mortelet K, D'Ambrosio U, Marigliano C, Catalano C (2010) Detection of hepatic metastases from colorectal cancer: prospective evaluation of gray scale US versus SonoVue® low mechanical index real-time enhanced US as compared with multidetector-CT

- or Gd-BOPTA-MRI. *Ultraschall in der Medizin-Eur J Ultrasound* 31:500–505. <https://doi.org/10.1055/s-0028-1109751>
9. Bamber J, Cosgrove D, Dietrich C, Fromageau J, Bojunga J, Calliada F, Cantisani V, Correas J-M, D'Onofrio M, Drakonaki E, Fink M, Friedrich-Rust M, Gilja O, Havre R, Jenssen C, Klauser A, Ohlinger R, Saftoiu A, Schaefer F, Sporea I, Piscaglia F (2013) EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 1: basic principles and technology. *Ultraschall in der Medizin Eur J Ultrasound* 34:169–184. <https://doi.org/10.1055/s-0033-1335205>
  10. Shiina T, Nightingale KR, Palmeri ML, Hall TJ, Bamber JC, Barr RG, Castera L, Choi BI, Chou Y-H, Cosgrove D, Dietrich CF, Ding H, Amy D, Farrokh A, Ferraioli G, Filice C, Friedrich-Rust M, Nakashima K, Schafer F, Sporea I, Suzuki S, Wilson S, Kudo M (2015) WFUMB guidelines and recommendations for clinical use of ultrasound elastography: part 1: basic principles and terminology. *Ultrasound Med Biol* 41:1126–1147. <https://doi.org/10.1016/j.ultrasmedbio.2015.03.009>
  11. Kamble R, Sodhi KS, Thapa BR, Saxena AK, Bhatia A, Dayal D, Khandelwal N (2017) Liver acoustic radiation force impulse (ARFI) in childhood obesity: comparison and correlation with biochemical markers. *J Ultrasound* 20:33–42. <https://doi.org/10.1007/s40477-016-0229-y>
  12. Giannetti A, Biscontri M, Matergi M, Stumpo M, Minacci C (2016) Feasibility of CEUS and strain elastography in one case of ileum Crohn stricture and literature review. *J Ultrasound* 19:231–237. <https://doi.org/10.1007/s40477-016-0212-7>
  13. Ricci P, Marigliano C, Cantisani V, Porfiri A, Marcantonio A, Lodise P, D'Ambrosio U, Labbadia G, Maggini E, Mancuso E, Panzironi G, Di Segni M, Furlan C, Masciangelo R, Taliani G (2013) Ultrasound evaluation of liver fibrosis: preliminary experience with acoustic structure quantification (ASQ) software. *Radiol Med* 118:995–1010. <https://doi.org/10.1007/s11547-013-0940-0>
  14. Jalalian A, Mashohor SBT, Mahmud HR, Saripan MIB, Ramli ARB, Karasfi B (2013) Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clin Imaging* 37:420–426. <https://doi.org/10.1016/j.clinimag.2012.09.024>
  15. Dromain C, Boyer B, Ferré R, Canale S, Delalogue S, Balleyguier C (2013) Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur J Radiol* 82:417–423. <https://doi.org/10.1016/j.ejrad.2012.03.005>
  16. Itoh A, Ueno E, Tohno E, Kamma H, Takahashi H, Shiina T, Yamakawa M, Matsumura T (2006) Breast disease: clinical application of US elastography for diagnosis. *Radiology* 239:341–350. <https://doi.org/10.1148/radiol.2391041676>
  17. Jales RM, Sarian LO, Torresan R, Marussi EF, Álvares BR, Derchain S (2013) Simple rules for ultrasonographic subcategorization of BI-RADS®-US 4 breast masses. *Eur J Radiol* 82:1231–1235. <https://doi.org/10.1016/j.ejrad.2013.02.032>
  18. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
  19. Xiao X, Jiang Q, Wu H, Guan X, Qin W, Luo B (2017) Diagnosis of sub-centimetre breast lesions: combining BI-RADS-US with strain elastography and contrast-enhanced ultrasound—a preliminary study in China. *Eur Radiol* 27:2443–2450. <https://doi.org/10.1007/s00330-016-4628-4>
  20. de Fleury FC (2015) The importance of breast elastography added to the BI-RADS® (5th edition) lexicon classification. *Revista da Associação Médica Brasileira* 61:313–316. <https://doi.org/10.1590/1806-9282.61.04.313>
  21. Moon WK, Lo C-M, Cho N, Chang JM, Huang C-S, Chen J-H, Chang R-F (2013) Computer-aided diagnosis of breast masses using quantified BI-RADS findings. *Comput Methods Programs Biomed* 111:84–92. <https://doi.org/10.1016/j.cmpb.2013.03.017>
  22. Shen W-C, Chang R-F, Moon WK (2007) Computer aided classification system for breast ultrasound based on breast imaging reporting and data system (BI-RADS). *Ultrasound Med Biol* 33:1688–1698. <https://doi.org/10.1016/j.ultrasmedbio.2007.05.016>
  23. Kim K, Song MK, Kim E-K, Yoon JH (2017) Clinical application of S-detect to breast masses on ultrasonography: a study evaluating the diagnostic performance and agreement with a dedicated breast radiologist. *Ultrasonography* 36:3–9. <https://doi.org/10.14366/usg.16012>
  24. Cho E, Kim E-K, Song MK, Yoon JH (2017) Application of computer-aided diagnosis on breast ultrasonography: evaluation of diagnostic performances and agreement of radiologists according to different levels of experience: application of computer-aided diagnosis on breast ultrasonography. *J Ultrasound Med*. <https://doi.org/10.1002/jum.14332>
  25. Botticelli A, Mazzotti E, Di Stefano D, Petrocelli V, Mazzuca F, La Torre M, Ciabatta FR, Giovagnoli RM, Marchetti P, Bonifacino A (2015) Positive impact of elastography in breast cancer diagnosis: an institutional experience. *J Ultrasound* 18:321–327. <https://doi.org/10.1007/s40477-015-0177-y>
  26. Cosgrove D, Piscaglia F, Bamber J, Bojunga J, Correas J-M, Gilja O, Klauser A, Sporea I, Calliada F, Cantisani V, D'Onofrio M, Drakonaki E, Fink M, Friedrich-Rust M, Fromageau J, Havre R, Jenssen C, Ohlinger R, Saftoiu A, Schaefer F, Dietrich C (2013) EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 2: clinical applications. *Ultraschall in der Medizin Eur J Ultrasound* 34:238–253. <https://doi.org/10.1055/s-0033-1335375>
  27. Barr RG, Nakashima K, Amy D, Cosgrove D, Farrokh A, Schafer F, Bamber JC, Castera L, Choi BI, Chou Y-H, Dietrich CF, Ding H, Ferraioli G, Filice C, Friedrich-Rust M, Hall TJ, Nightingale KR, Palmeri ML, Shiina T, Suzuki S, Sporea I, Wilson S, Kudo M (2015) WFUMB guidelines and recommendations for clinical use of ultrasound elastography: part 2: breast. *Ultrasound Med Biol* 41:1148–1160. <https://doi.org/10.1016/j.ultrasmedbio.2015.03.008>
  28. Sadigh G, Carlos RC, Neal CH, Dwamena BA (2012) Accuracy of quantitative ultrasound elastography for differentiation of malignant and benign breast abnormalities: a meta-analysis. *Breast Cancer Res Treat* 134:923–931. <https://doi.org/10.1007/s10549-012-2020-x>
  29. Sadigh G, Carlos RC, Neal CH, Wojcinski S, Dwamena BA (2013) Impact of breast mass size on accuracy of ultrasound elastography vs. conventional B-mode ultrasound: a meta-analysis of individual participants. *Eur Radiol* 23:1006–1014. <https://doi.org/10.1007/s00330-012-2682-0>
  30. Hao S-Y, Jiang Q-C, Zhong W-J, Zhao X-B, Yao J-Y, Li L-J, Luo B-M, Ou B, Zhi H (2016) Ultrasound elastography combined with BI-RADS-US classification system: is it helpful for the diagnostic performance of conventional ultrasonography? *Clin Breast Cancer* 16:e33–e41. <https://doi.org/10.1016/j.clbc.2015.10.003>

## Affiliations

**Mattia Di Segni<sup>1</sup>**  · **Valeria de Soccio<sup>1</sup>** · **Vito Cantisani<sup>1</sup>** · **Giacomo Bonito<sup>1</sup>** · **Antonello Rubini<sup>1</sup>** · **Gabriele Di Segni<sup>2</sup>** · **Sveva Lamorte<sup>1</sup>** · **Valentina Magri<sup>1</sup>** · **Corrado De Vito<sup>3</sup>** · **Giuseppe Migliara<sup>3</sup>** · **Tommaso Vincenzo Bartolotta<sup>4</sup>** · **Alessio Metere<sup>5</sup>** · **Laura Giacomelli<sup>5</sup>** · **Carlo de Felice<sup>1</sup>** · **Ferdinando D'Ambrosio<sup>1</sup>**

✉ Mattia Di Segni  
mattia.di.segni@gmail.com

<sup>1</sup> U.O.C. Diagnostica per Immagini, P. O San Paolo  
- ASL Roma 4, Largo dei Donatori del Sangue 1,  
00053 Civitavecchia (RM), Italy

<sup>2</sup> Arakne S.r.l., Via Edoardo d'Onofrio, 304, 00155 Rome,  
RM, Italy

<sup>3</sup> Department of Public Health and Infectious Diseases,  
Sapienza University of Rome, Piazzale Aldo Moro 5, Rome,  
Italy

<sup>4</sup> Department of Radiology, University of Palermo,  
90127 Palermo, Italy

<sup>5</sup> Department of Surgical Sciences, Sapienza University  
of Rome, viale del Policlinico 155, 00161 Rome, Italy