Prospective evaluation of breast lesions categorisation with S-detect: feasibility and inter-observer variability assessment

Poster No.: C-1228
Congress: ECR 2018
Type: Scientific Exhibit
Authors: V. De Socco, V. Cantisani, G. Bonito, V. Magri, M. Di Segni, A. Rubini, C. De Felice, F. D'Ambrosio, C. Catalano; Rome/IT
Keywords: Breast, Ultrasound, CAD, Computer Applications-Detection, diagnosis
DOI: 10.1594/ecr2018/C-1228

Any information contained in this pdf file is automatically generated from digital material submitted to EPOS by third parties in the form of scientific presentations. References to any names, marks, products, or services of third parties or hypertext links to third-party sites or information are provided solely as a convenience to you and do not in any way constitute or imply ECR's endorsement, sponsorship or recommendation of the third party, information, product or service. ECR is not responsible for the content of these pages and does not make any representations regarding the content or accuracy of material in this file.

As per copyright regulations, any unauthorised use of the material or parts thereof as well as commercial reproduction or multiple distribution by any traditional or electronically based reproduction/publication method is strictly prohibited.

You agree to defend, indemnify, and hold ECR harmless from and against any and all claims, damages, costs, and expenses, including attorneys' fees, arising from or related to your use of these pages.

Please note: Links to movies, ppt slideshows and any other multimedia files are not available in the pdf version of presentations.

www.myESR.org
Aims and objectives

US has a pivotal role in breast imaging especially for breast lesions characterization.

Furthermore, to improve the performance of US, other tools were developed, such as elastography, Computer-Aided Diagnosis (CAD) systems [1] and S-detect.

S-detect is a software based upon the deep learning algorythm, which performs lesion segmentation, feature analysis and describes, according to either BIRADS 2003 or BIRADS 2013, lexicon and suggests a dichotomic categorization. In addition to the role as a possible adjunct tool in breast lesion characterization, another possible application could be as a teaching tool.

The aims of our study were:

• to assess the diagnostic performance of S-detect in breast lesions categorization according to BIRADS 2013 US lexicon;
• to assess the potential role of S-detect as a teaching tool.
Methods and materials

Patient population: The study design was prospective. Between July 2016 and June 2017, 122 female patients aged between 21 and 84 years (mean 51 years) underwent baseline US examination with linear array probes.

The following inclusion criteria were used:

- Patients with follow-up in progress or who previously underwent cytology or biopsy sampling due to likely benign lesions;
- Patients candidate to biopsy due to lesions suspicious for malignancy.

The exclusion criteria employed were:

- Pregnancy;
- Lactation;
- Neo-adjuvant chemotherapy in progress or at less than 2 months since completion;
- Radiotherapy in progress or at less than 3 months since completion;
- Insufficient documentation.

Data acquisition: All US examinations were performed using a UGEO RS80A machinery (Samsung Healthcare, South Korea) with 3-16 MHz or 3-12 MHz (in case of abundant breast volume or deep seated lesions) linear array probes, and performing a scan representing the input image for S-detect software.

Data analysis: The study was organized in two phases:

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

The prospective acquisition of baseline images was performed by a Radiologist with 32 years experience in breast imaging (Fig.1).

Subsequently, patients were assessed by a second experienced operator with 18 years experience in US (Fig.2) and by four Radiology residents with different years of residency program and different dedication level in breast imaging: 5th year with limited experience in breast imaging (Fig.3), 2 years with deeper experience in breast imaging (Fig. 4), 3rd year with limited experience in breast imaging (Fig.5), 1st year with deeper experience in breast imaging (Fig.6).

Once selected the image, the operator activated the software and guided the lesion segmentation phase - performed either automatically or guided by the operator. When
segmentation was complete, the software computed its segmentation, description and classification proposals; of these proposals, in case of multiple options, the one with the most adequate segmentation and characterization was chosen.

The description is automatic for the subsequent parameters encoded in the BI-RADS 2013 lexicon [2]: shape, orientation, margins, pattern, and posterior acoustic features; for the other features, manual insertion was requested. In case of BI-RADS 4 assignment, a subcategorization was performed according to the scheme proposed by Jales et al [3]. When these steps were completed, a structured report was "assigned".

Later on, between 2 and 6 weeks later, both the radiologist with 18 years experience and the residents retrospectively reviewed images and expressed their assessments, blinded to pathology results.

Pictures, structured reports, operators' assessments, pathologic or follow-up data of the nodules and assessments were stored in an appositely developed integrated system made of a graphic interface and based upon a SQLite encoded relational database, which allowed to fill, store and retrieve the data.

**Statistical Analysis:** Data were analyzed with Stata (Stata v 13.0, StataCorp Lt, College Station, TX, USA) to assess:

- Operators' performance with 2x2 contingency tables and ROC curves against the results of histology;
- S-detect performance with 2x2 contingency tables and ROC curves against the results of histology;
- Differences in performance between S-detect and each Operators with the test for equality of ROC areas against the results of histology;
- Agreement between operators and S-detect with Cohen's kappa.

Categories dichotomization was applied as follows: BIRADS 2 and 3 were considered benign, whereas BIRADS categories 4a to 5 were considered malignant.

Inter-operator concordance was determined with Cohen's kappa test and interpreted according to the subsequent scale, as expressed by Landis and Koch [4]:

- excellent: # > 0.8;
- good: # = 0.61-0.8;
- moderate: # = 0.41-0.6;
- fair: # = 0.21-0.4;
To assess the potential of S-detect as a teaching tool, among the residents' assessments, those one classified as BI-RADS 4a were compared with S-detect assessments. In case of disagreement, priority was given to S-detect assessment and their results were applied in the operators' classification, evaluated with 2x2 contingency tables and ROC curves and compared with baseline performance with the test for equality of ROC areas against the results of histology chi-square test.

For all statistical tests a value of P less than 0.05 was considered significant.
**Fig. 1:** Radiologist with 32 years experience in breast imaging

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 2: Radiologist with 18 years experience

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
**Fig. 3:** 5th year with limited experience in breast imaging

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT

**Fig. 4:** 2 years with deeper experience in breast imaging

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 5: 3rd year with limited experience in breast imaging

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 6: 1st year with deeper experience in breast imaging

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Results

Breast lesions: 68 nodules in 61 patients were analyzed, with size between 10 and 48 mm; of these nodules, 44 were malignant and 24 were benign. Among 44 malignancies, 37 were IDC, 3 were DCIS, 3 were Infiltrating Lobular Carcinomas and 1 was a Granular Cell Tumor. Among 24 benign lesions there were: 12 were fibroadenomas (one of them was characterized through a biopsy and one was characterized with cytology and was being followed-up), 1 Phyllodes tumor, 2 hamartomas (characterized through the sonographic appearance and followed-up), 7 foci of biopsy proven sclerosingadenosis and/or fibrocystic mastopathy and 2 abscesses (sonographically characterized and followed-up).

Inter- and intra-observer variability: Diagnostic performance of S-detect and the expert operator is shown in table 1.

In the overall assessment of benign vs malignant lesions, a $k= 0.616$ (95% C.I 0.376-0.856) was obtained, corresponding to a substantial agreement.

Compared to in-training operators, S-detect showed similar sensitivity to all of the operators, whereas specificity was higher compared to that of non-dedicated operators (see Table 2), but slightly inferior compared to that of the dedicated expert in training operator.

Diagnostic performance of S-detect and in-training operators is shown in table 2.

The agreement between the operators was not significant, whereas the agreement with between S-detect and the operators varied between moderate and good (see tables 4 and 5) and statistically significant.

To measure how S-detect could determine an increase in diagnostic performance for in-training operators, all the lesions which were categorized as BIRADS 4a were selected; these were compared with S-detect assessments and in case of disagreement between the two, S-detect agreement was given the priority; then, these evaluations were reinserted in the whole table and underwent the diagnostic test statistics. A statistically significant increase in diagnostic performance was noted, especially for 2 of those operators with the lowest performance (see table 6 for performance and table 7 for significance), whereas for the in-training operator with the deepest experience in breast imaging this change was not significant. In addition, also for another of the operators a non-significant change was noted, related to the change in categorization in 4 cases only.

ROC analysis: According to the test for the equality of ROC areas, the difference in diagnostic performance was not significant (ROC: 0.8201 vs 0.8409; p=0.751).
ROC curves for S-detect and expert operator are shown in figure 7, 8 and 9.

Figures 10, 11, 12 and 13 show the ROC curves of the in-training operators and that of S-detect.

The difference in ROC areas between S-detect and in-training operators was not statistically significant when compared with each operator (see table 3).

Figures 14, 15, 16 and 17 show the ROC curves of the addition and integration of S-detect.
Fig. 7: ROC curves for expert operator and for S-detect: a- expert operator; b- S-detect.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 8: ROC curves for expert operator and for S-detect: a- expert operator; b- S-detect.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
**Fig. 9:** ROC curves for expert operator and for S-detect: a- expert operator; b- S-detect.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 10: ROC curve analysis for in training operators: 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.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 11: ROC curve analysis for in training operators: 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.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 12: ROC curve analysis for in training operators: 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.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 13: ROC curve analysis for in training operators: 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.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 14: Performance of the in-training operators after the integration of S-detect in ambiguous cases (BI-RADS 4°): a- 5th year with limited experience; b- 2nd year with deeper experience; c- 3rd year with limited experience; d- 1st year with deeper experience.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 15: Performance of the in-training operators after the integration of S-detect in ambiguous cases (BI-RADS 4°): a- 5th year with limited experience; b- 2nd year with deeper experience; c- 3rd year with limited experience; d- 1st year with deeper experience.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 16: Performance of the in-training operators after the integration of S-detect in ambiguous cases (BI-RADS 4°): a- 5th year with limited experience; b- 2nd year with deeper experience; c- 3rd year with limited experience; d- 1st year with deeper experience.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Fig. 17: Performance of the in-training operators after the integration of S-detect in ambiguous cases (BI-RADS 4°): a- 5th year with limited experience; b- 2nd year with deeper experience; c- 3rd year with limited experience; d- 1st year with deeper experience.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT

<table>
<thead>
<tr>
<th>Operator</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PLR (%)</th>
<th>NLR (%)</th>
<th>ROC (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-detect</td>
<td>91.1 (78.8% - 97.5%)</td>
<td>70.8% (48.9% - 87.4%)</td>
<td>3.12 (1.66 - 5.87)</td>
<td>0.13 (0.05 - 0.33)</td>
<td>0.81 (0.71 - 0.91)</td>
<td>85.4% (72.2% - 93.9%)</td>
<td>81.0% (58.1% - 94.6%)</td>
</tr>
<tr>
<td>Expert operator</td>
<td>93.2% (81.3% - 98.6%)</td>
<td>75.0% (53.3% - 90.2%)</td>
<td>3.73 (1.86 - 7.49)</td>
<td>0.09 (0.03 - 0.28)</td>
<td>0.84 (0.74 - 0.94)</td>
<td>87.2% (74.3% - 95.2%)</td>
<td>85.7% (63.7% - 97.0%)</td>
</tr>
</tbody>
</table>

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

* Experience in breast imaging

**Table 2**: Diagnostic performance of S-detect and in-training operators.
<table>
<thead>
<tr>
<th>Comparison</th>
<th>Observed agreement</th>
<th>k</th>
<th>Standard Error</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert vs. 2nd year experienced</td>
<td>58.82%</td>
<td>0.0094</td>
<td>0.1210</td>
<td>0.4691</td>
</tr>
<tr>
<td>Expert vs. 5th year with limited experience</td>
<td>66.18%</td>
<td>0.1272</td>
<td>0.1169</td>
<td>0.1382</td>
</tr>
<tr>
<td>Expert vs. 1st year experienced</td>
<td>60.29%</td>
<td>-0.0552</td>
<td>0.1135</td>
<td>0.6865</td>
</tr>
<tr>
<td>Expert vs. 3rd year with limited experience</td>
<td>58.82%</td>
<td>-0.0781</td>
<td>0.1154</td>
<td>0.7509</td>
</tr>
</tbody>
</table>

**Table 3:** Significance of the differences ROC areas between the evaluation of the operators and S-detect readings.
<table>
<thead>
<tr>
<th>Comparison</th>
<th>Observed Agreement</th>
<th>k</th>
<th>Standard Error</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-detect vs. expert operator</td>
<td>83.82%</td>
<td>0.6160</td>
<td>0.1212</td>
<td>0.000</td>
</tr>
<tr>
<td>S-detect vs. 5th year with limited experience</td>
<td>85.29%</td>
<td>0.6119</td>
<td>0.1179</td>
<td>0.000</td>
</tr>
<tr>
<td>S-detect vs. 2nd year experienced</td>
<td>89.71%</td>
<td>0.7484</td>
<td>0.1212</td>
<td>0.000</td>
</tr>
<tr>
<td>S-detect vs. 3rd year with limited experience</td>
<td>82.61%</td>
<td>0.5400</td>
<td>0.1146</td>
<td>0.000</td>
</tr>
<tr>
<td>S-detect vs. 1st year experienced</td>
<td>85.29%</td>
<td>0.5991</td>
<td>0.1149</td>
<td>0.000</td>
</tr>
</tbody>
</table>

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

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

Se: sensitivity; Sp: specificity; PLR: Positive Likelihood Ratio; NLR: Negative Likelihood Ratio; PPV: Positive Predictive Value; NPV: Negative Predictive Value.

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

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Table 6: Diagnostic performance of the in-training operators after the integration of S-detect in the assessment of BI-RADS 4a lesions.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT

<table>
<thead>
<tr>
<th>Operator</th>
<th>ROC area - before</th>
<th>ROC area - after</th>
<th>Chi-square</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>5th year with limited experience</td>
<td>0.7595</td>
<td>0.8523</td>
<td>4.4813</td>
<td>0.0343</td>
</tr>
<tr>
<td>2nd year experienced</td>
<td>0.8314</td>
<td>0.8617</td>
<td>0.9571</td>
<td>0.33</td>
</tr>
<tr>
<td>3rd year with limited experience</td>
<td>0.7389</td>
<td>0.8097</td>
<td>2.3377</td>
<td>0.12</td>
</tr>
<tr>
<td>1st year experienced</td>
<td>0.7500</td>
<td>0.8636</td>
<td>5.9586</td>
<td>0.0146</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert operator vs. S-detect</td>
<td>0.751</td>
</tr>
<tr>
<td>5th year with limited experience* vs. S-detect</td>
<td>0.000</td>
</tr>
<tr>
<td>2 anno with deeper experience vs. S-detect</td>
<td>0.831</td>
</tr>
<tr>
<td>3rd year with limited experience* vs. S-detect</td>
<td>0.151</td>
</tr>
<tr>
<td>1st year with deeper experience vs. S-detect</td>
<td>0.206</td>
</tr>
</tbody>
</table>
Table 7: Significance of the increase in performance for in-training operators.

© Department of Radiological, Oncological and Pathological Sciences, Sapienza, University of Rome - Rome/IT
Conclusion

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

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

According to Xiaoyun Xiao et al [6], the use of 2013 criteria showed 100% sensitivity, 17.4% specificity, 46.8% PPV, 100% NPV and 0.867 ROC.

Conversely, according to Fleury et al [7] their used led to 93.85% sensitivity, 72.07% specificity, 47.16% PPV, 97.78% NPV and 76.64% accuracy.

According to Cho Eun et al. [8], using the fifth edition led for two radiologists with different experience led to 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.

In our experience, even though with a limited sample size, US showed high performance when performed by experienced staff, with 93.2% sensitivity, 75% specificity, 87.2% PPV and 85.7% NPV, in agreement with literature statements.

Compared to the former edition, not every author identify significant differences.

In the paper by Jung Hyun Yoon et al [9], in which the two editions are specifically compared by analyzing 83 lesions undergoing biopsy or follow-up, 2013 criteria showed 96.6% sensitivity, 41.2% specificity, 48.3% PPV, 95.5% NPV, 61.3% accuracy and 0.69 ROC, whereas the former editions showed 96.6% sensitivity, 45.1% specificity, 50% PPV, 95.8% NPV 63.8% accuracy and 70.8% ROC; concordance was defined good (76.3% of cases).

According to Shao-Yun et al [10], employing US only with 2013 criteria showed 96% sensitivity, 77.3% specificity, 73.8% PPV, 97% NPV and 85.3% accuracy compared with 99.3% sensitivity, 48.7% specificity, 56.1% PPV, 99.1% NPV and 68.8% accuracy for 2003 criteria.

In terms of inter-observer agreement, YounJoo Lee [11] showed that for University and Hospital staff appeared good for all criteria, whereas between residents it varied between moderate and scarce grades.
Also according to Cho Eun et al [8], the agreement between operators ranged between substantial and fair for the descriptors and moderate for final assessment.

In addition, BI-RADS criteria are often used as a source for descriptive features employed in the Computer-Aided Diagnosis and Computer-Aided Classifications Systems.

These systems work in three phases: image processing, segmentation and feature extraction [1]; they are classified according to the algorithms employed in each phases.

According to some authors, these systems can potentially improve breast lesions classification, both in terms of performance and operator dependence [12].

In a previous experience, Moon et al. [13], used as a reference the evaluations of two breast radiologists - stated according to the lexicon BI-RADS 2003 - and assessed a 244 neoplasms sample characterized through a biopsy to compare the diagnostic ability of two CAD systems, one conventional and one with BI-RADS parameters quantification; the latter system considered malignant any lesion with at least one suspicious feature. In their study, the difference in performance appeared significant, showing 84% specificity, 87% accuracy, 73% PPV and 97% NPV for sensitivity values set at 95%, versus 60% specificity, 71% accuracy, 53% PPV and 96% NPV.

In another study [14], 626 US images of pathologically characterized lesions were analyzed with two CAD systems and compared with the retrospective assessments made by two radiologists; the results were obtained, with a significant increase in performance compared to that of the radiologists and a substantial agreement.

In a previous study concerning S-detect [15], 192 breast lesions were retrospectively assessed both by a radiologist and by the software and classified with BI-RADS 2003 criteria, applying two cut-offs: between BI-RADS 3 and 4 and between BI-RADS 4a and 4b - the latter was applied only to the radiologist 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 with a significant difference compared to 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 between moderate and scarce.

According to Cho Eun et al [9], when S-detect performance was compared with that of 2 radiologists with different experience, the software alone has a lower sensitivity and a higher specificity compared to both the more experienced and less experienced radiologists - 72.2% vs. 94.4% and 94.4% sensitivity and 90.8% vs. 49.2 and 55.4% specificity. However, when operators' readings and software assessments were combined, a significant increase in specificity and PPV was noted, with no statistically significant detrimental effect on sensitivity. Also, when categorization was analyzed after dichotomization, moderate agreement between S-detect and the operators was observed.
Even if we applied different lexicon and categorization modalities compared to Kim's study, our experience showed that S-detect has 91.1% sensitivity, 70.8% specificity, 85.4% PPV and 81% NPV, with a 0.81 ROC; in addition, inter-observer agreement with the operators, especially the most expert and dedicated - appeared substantial.

When cases categorized as ambiguous (i.e BIRADS 4a) were selected, the addition of S-detect assessment, when given priority in case of disagreement with the operator, led to a significant increase in the diagnostic performance of in training operators with less experience and lower performance, showing a potential use as a teaching tool.

Indeed, in terms of performance and agreement, our study agrees with Ki. et al.

However, our results partially disagree with Cho et al in terms of sensitivity, probably due to our sample composition.

In addition, a study assessed how the accuracy of a CAD system depends from image selection and classification protocol [16]:

- When changing the classification protocol, statistically significant changes were in the ROC only whether the averaging was activated or not, with no difference according to the procedural step the latter was performed;

- When changing the images selection modality, no significant changes in the ROC were noted, both whether the activation of image averaging was performed and not and by changing the procedural step in which the latter was performed.

Limits of our study are related to limited sample size, selected sample with a high prevalence of malignant lesions; the latter point, however, is very often present in previous studies, even though with lesser weight compared to our report. An additional limit is related to the retrospective phase of our study, which characterized the second part of our protocol.

In conclusion S-detect is a feasible tool in the categorization of breast lesions and a good decision-making support, especially for the beginner or non-breast radiologist.
G. Alagna, MD

Università Sapienza di Roma, Policlinico Umberto I, Department of Radiology - Italy.

Viale del Policlinico, 155

00161 Roma RM, Italy

Phone: +39 328 3745751

e-mail: gregorio.alagna@uniroma1.it
References


