The value of S-Detect for the differential diagnosis of breast masses on ultrasound: a systematic review and pooled meta-analysis

Jun Li*, Tian Sang*, Wen-Hui Yu, Meng Jiang, Shu-Yan Huang, Chun-Li Cao, Ming Chen, Yu-Wen Cao, Xin-Wu Cui, Christoph F. Dietrich

* the authors shared the first authorship

1Department of Medical Ultrasound, the First Affiliated Hospital of Medical College, Shihezi University, Shihezi, China, 2Department of Medical Ultrasound, Wuchang Hospital affiliated Wuhan University of Science and Technology, Wuhan, China, 3Department of Medical Ultrasound, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan China, 4Department of Ultrasound, The First People’s Hospital of Huaihua, University of South China, Huaihua, China, 5Department Allgemeine Innere Medizin (DAIM), Kliniken Hirslanden Beau Site, Salem und Permancence, Bern, Switzerland

Abstract

Aim: To evaluate the value of S-Detect (a computer aided diagnosis system using deep learning) in breast ultrasound (US) for discriminating benign and malignant breast masses. Material and methods: A literature search was performed and relevant studies using S-Detect for the differential diagnosis of breast masses were selected. The quality of included studies was assessed using a Quality Assessment of Diagnostic Accuracy Studies (QUADAS) questionnaire. Two review authors independently searched the articles and assessed the eligibility of the reports. Results: A total of ten studies were included in the meta-analysis. The pooled estimates of sensitivity and specificity were 0.82 (95%CI: 0.77-0.87) and 0.86 (95%CI: 0.76-0.92), respectively. In addition, the diagnostic odds ratios, positive likelihood ratio and negative likelihood ratio were 28 (95%CI: 16-49), 5.7 (95%CI: 3.4-9.5), and 0.21 (95%CI: 0.16-0.27), respectively. Area under the curve was 0.89 (95%CI: 0.86-0.92). No significant publication bias was observed. Conclusions: S-Detect exhibited a favourable diagnostic value in assisting physicians discriminating benign and malignant breast masses and it can be considered as a useful complement for conventional US. Keywords: artificial intelligence; ultrasonography; diagnosis; meta-analysis; breast

Introduction

Breast cancer remains a significant scientific, clinical and societal challenge [1,2]. It is the second most common cancer in the world and the most frequently diagnosed cancer in women [3-5]. Approximately 11% of worldwide breast cancers occur in China and the incidence has increased rapidly in recent decades [6]. Some studies have shown that the age, geographic locations, family history, hormone and reproductive history (including nulliparity, late age at first childbirth, early age at menarche and use of oral contraceptives or hormone replacement therapy) are related to the incidence of breast cancer [6-8]. But its pathogenesis is complex and the high-risk factors are difficult to control. The early accurate diagnosis and control of breast cancer are important for the prognosis [9,10].

Breast ultrasound (US) is often used as an adjunct to mammography and magnetic resonance imaging (MRI) to characterize focal breast lesions, thereby improving cancer detection rates [11]. However, US is an operator-
dependent imaging modality that requires experience to depict the breast lesion and diagnose it, in which the reproducibility of the images should also be considered.

Many new imaging techniques have been developed such as ultrasound elastography, contrast-enhanced ultrasound (CEUS), three-dimensional US and automatic breast full-volume scanning imaging system (ABVS) which all have provided more convenience [12,13].

In the recent years, artificial intelligence (AI), from machine learning (ML) to deep learning (DL) algorithms, is gaining extensive attention for its excellent performance in image-recognition tasks [9,14-16]. S-Detect (Samsung Medison Co. Ltd., Seoul, Korea) is a recently developed ultrasound computer assistant diagnosis (CAD) system using deep learning, and has become one of the most increasingly used CAD systems for the diagnosis of breast cancer. It can implement auto-segmentation and interpretation basically of US morphological descriptions, allowing classification of breast lesions in a dichotomic form (possibly benign or possibly malignant) as a reference for radiologists to assist with the final diagnosis. But it is not automatic segmentation in the full sense and needs to be adjusted manually a little if the outline is not good [17]. It is known as a clinically feasible diagnostic tool with a moderate degree of agreement in the final assessment, regardless of the experience of the radiologists specializing in breast imaging [18]. But the diagnostic criteria of the S-Detect classification system are not clearly defined in terms of whether the lesion is calcified, the distribution of blood vessels, changes in surrounding tissues, and the presence of enlarged lymph nodes. It still requires artificial judgment and increases subjectivity.

As a new technique, the application of S-Detect in clinical practice is still controversial and the results vary from different clinical trials. In the results from Jiang et al., the sensitivity was 100% [13], while Gewefel et al showed a sensitivity of just 61.90% [19].

At present, the sensitivity of the results varies greatly among studies and there is no meta-analysis and guidance on this technique for the diagnosis of breast cancer. Therefore, we conducted a meta-analysis to assess the performance of S-Detect in discriminating benign and malignant breast masses.

**Material and method**

**Search strategy**

Literature retrieval was performed to select all relevant studies published before 15th July 2019 from databases including PubMed, Embase, China National Knowledge Infrastructure (CNKI), China Biology Medicine disc, WANFANG databases, Cochrane Library, Google Scholar and Web of Science. The terms used included “computer aided diagnosis”, “breast nodules”, “breast masses”, “breast cancer”, “breast neoplasm”, “breast malignant neoplasm”, “breast tumour”, “diagnosis”, “diagnostic” “S-Detect”, “smart detect”, “ultrasound”, and “artificial intelligence”. The searches were limited to identify the diagnostic studies without language restrictions.

**Study selection**

After searching the candidate studies for inclusion, criteria were set for further identification. The remaining studies were included in the meta-analysis if the following inclusion criteria were satisfied: 1) Studies evaluating the diagnostic accuracy of S-Detect; 2) The diagnosis of malignant and benign breast masses should be confirmed by histopathological examination; 3) Diagnostic accuracy statistics obtained by using this technique alone, including true-positive (TP), false-positive (FP), true-negative (TN), false-negative (FN), should be provided or can be calculated; 4) Clinical studies should have received informed consent from each study subject and obtained approval from the local ethics committees. Furthermore, studies were excluded if: 1) Studies did not provide sufficient data for calculating the TP, FP, TN, and FN parameters or only had obtained t results of the combined diagnosis of the technique and the doctor; 2) Studies incorporated duplicated data; 3) Studies that were not approved by local ethics committees or did not acquire informed consents from all the study subjects. Relevant studies were selected by two researchers independently based on title and abstract or full article. Any discrepancies concerning the study selection were resolved by discussion of the full article.

**Data extraction and quality assessment**

The following information were extracted using a standardized form: the first author’s surname, year of publication, country of the study, number of masses available for analysis, reference standard, mean age of patients, mean size of masses, type of the US system, TP, FN, FP, and TN numbers. The quality of included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADS) questionnaire by the same two observers [20]. The Revman 5.3, special software for Cochrane collaborative network, was used to output the result of QUADS.

**Statistical analysis**

After data extraction, the pooled sensitivity (PSEN), pooled specificity (PSPE), positive likelihood ratio (PLR) and the negative likelihood ratio (NLR) were estimated by the bivariate model and the summary receiver operating characteristic (SROC) [21]. The post-test probabilities were calculated by the PLR and NLR and plotted on
a Fagan nomogram. The SROC curve with 95% confidence and prediction regions was also plotted to illustrate the relationship between sensitivity and specificity. In addition, the publication bias was assessed using the Deeks’ method. All the data analysis and the graphs were made using the STATA version 14.0 software for Windows (Stata Corp, College Station, TX) with the commands MIDAS and METANDI. p<0.05 was regarded as statistically significant. RevMan software, version 5.3 (The Nordic Cochrane Center, The Cochrane Collaboration, Copenhagen, Denmark) was used to make flowcharts and assess the quality of the research included. MetaDisc software, Version 1.4 (Madrid, Spain) was used to test the threshold effect. And the meta-regression analyses were used to deal with the existence of heterogeneity.

Results

Literature searches

Based on the key words, 219 related papers were initially found. After screening the title and abstract, 38 articles were filtered automatically as duplicates, another 167 articles were removed because they were unrelated to the topic or belonged to review. From 14 full-text articles assessed for eligibility, 4 papers were further excluded because of insufficient data or for other reasons. Finally, 10 studies were included in the meta-analysis (fig 1).

Study characteristics

All the 10 studies (English = 5, Chinese = 5) were diagnostic cohort and prospective studies [10,13,16,17,19,22-26]. A total of 2054 breast masses (diameters range from 0.2 cm to 1.32 cm) from 1777 patients (age ranged from 14 to 83 years) were identified and 720 (35%) masses of them were malignant. The detailed characteristics of the included studies were summarized in Table I. The publication dates of the literature were from 2016 to 2019. Nine articles in the analysis evaluated breast masses in Asia [10,13,16,17,22-26], while only 1 article was concerned with Africa [19]. All US examinations were performed with RS80A (Samsung Medison Co., Ltd., Seoul, Korea) which equipped S-Detect technique in US system [10,13,16,17,19,22-26]. Three articles used the fourth edition BI-RADS [13,17,24], while 6 articles used the fifth edition to diagnosis if the lesion was benign or malignant [10,16,19,22,23,26], and one article did not specify which version it used. All studies were based on histology biopsy as the gold standard [10,13,16,17,19,22-26].

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Year</th>
<th>Country</th>
<th>N (Benign vs Malignant)</th>
<th>Age (years±SD)</th>
<th>Size (mm)</th>
<th>BI-RADS edition</th>
<th>Se (%)</th>
<th>Sp (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cho, 2018 [22]</td>
<td>2018</td>
<td>Korea</td>
<td>119(65vs54)</td>
<td>48.5±12.2</td>
<td>16.9±10.7</td>
<td>5th</td>
<td>72.20</td>
<td>90.80</td>
</tr>
<tr>
<td>Du, 2019 [25]</td>
<td>2019</td>
<td>China</td>
<td>202(93vs109)</td>
<td>51.0±15.0</td>
<td>-</td>
<td>-</td>
<td>77.40</td>
<td>78.90</td>
</tr>
<tr>
<td>Gewefel, 2017 [19]</td>
<td>2017</td>
<td>Egypt</td>
<td>45(24vs21)</td>
<td>43.0±7.98</td>
<td>15.00±6.24</td>
<td>5th</td>
<td>61.90</td>
<td>100</td>
</tr>
<tr>
<td>Jiang, 2016 [13]</td>
<td>2016</td>
<td>China</td>
<td>48(42vs6)</td>
<td>46.6±13.3</td>
<td>14.7±9.7</td>
<td>4th</td>
<td>100</td>
<td>90.50</td>
</tr>
<tr>
<td>Kim, 2017 [17]</td>
<td>2017</td>
<td>Korea</td>
<td>192(120vs72)</td>
<td>44.02±12.13</td>
<td>19.26±9.25</td>
<td>5th</td>
<td>90.62</td>
<td>65.01</td>
</tr>
<tr>
<td>Li, 2019 [16]</td>
<td>2019</td>
<td>China</td>
<td>563(403vs160)</td>
<td>42.0±12.0</td>
<td>-</td>
<td>-</td>
<td>92.30</td>
<td>90.60</td>
</tr>
<tr>
<td>Wang, 2019 [26]</td>
<td>2019</td>
<td>China</td>
<td>220(181vs39)</td>
<td>42.6±12.9</td>
<td>-</td>
<td>-</td>
<td>87.60</td>
<td>81.30</td>
</tr>
<tr>
<td>Wu, 2019 [23]</td>
<td>2019</td>
<td>China</td>
<td>338(209vs129)</td>
<td>42.6±12.9</td>
<td>-</td>
<td>-</td>
<td>87.07</td>
<td>72.27</td>
</tr>
<tr>
<td>Zhao, 2019 [10]</td>
<td>2019</td>
<td>China</td>
<td>266(161vs105)</td>
<td>42.66±16.42</td>
<td>17.21±8.20</td>
<td>4th</td>
<td>80.60</td>
<td>96.00</td>
</tr>
<tr>
<td>Zhou, 2017 [24]</td>
<td>2017</td>
<td>China</td>
<td>61(36vs25)</td>
<td>42.66±16.42</td>
<td>17.21±8.20</td>
<td>4th</td>
<td>80.60</td>
<td>96.00</td>
</tr>
</tbody>
</table>

SD: standard deviation; -: not mentioned; N: number of masses; BI-RADS: Breast Imaging Reporting and Data System; US: ultrasound; Se: sensitivity; Sp: specificity
Methodology quality assessment

According to the methodological assessment by the QUADAS-2 checklist, the quality of the included studies was judged to be high (fig 2). Most of the quality assessment items of the included studies had a low risk of bias.

Accuracy of S-Detect in distinguishing benign and malignant breast masses

As shown, the pooled sensitivity (PSEN) and pooled specificity (PSPE) of S-Detect technique in distinguish benign and malignant breast masses were 0.82 (95%CI: 0.77-0.87) and 0.86 (95%CI: 0.76-0.92), respectively (fig 3). Significant heterogeneity in pooling the sensitivity ($I^2=77.04\%$, $p<0.01$) and specificity ($I^2=91.61\%$, $p<0.01$) was detected, so we pooled the sensitivity and specificity in the random effects model (fig 4). The DOR, PLR and NLR were 28 (95%CI: 16-49), 5.7 (95%CI: 3.4-9.5) and 0.21 (95%CI: 0.16-0.27), respectively. The area under the curve (AUC) was 0.89 (95%CI: 86-92) (fig 5). The Spearman correlation coefficient ($r=0.333$, $p=0.347$) indicated that there was no significant threshold effect ($p>0.05$), which also showed that other factors may cause the heterogeneity. All the above statistical analyses results were acceptable.

Publication bias

The Deeks’ funnel plot for testing publication bias showed that the studies were distributed symmetrically with a $p$-value of 0.10 ($p>0.05$), indicating no clear evidence of publication bias (fig 6).

Heterogeneity detection

The threshold effect test suggested other factors besides the threshold effect might result in heterogeneity between included studies [10,13,16,17,19,22-26]. The meta-regression was adopted. The results suggested that the source of cases (China or other group), the number of masses (number ≥200 or <200) and the edition of BI-RADS (the fourth or the fifth) had something to do with heterogeneity (Table II). Among these parameters and categories, it was found that only the variable of the number of the masses has statistical significance for the sensitivity ($p=0.04$). The studies based on more than 200 masses displayed a higher sensitivity of 0.86 (95%CI: 0.83-0.90) [10,16,23]; others based on less than 200 nodules showed a little lower summary sensitivity of 0.75 (95%CI: 0.67-0.83) [13,17,19,22,24-26]. As for the summary specificity, the level of the number of the masses
also has statistical significance (p=0.01). The studies based on more than 200 masses evidenced a low specificity of 0.79 (95%CI: 0.67-0.90), while others offered a higher specificity of 0.91 (95%CI: 0.83-0.98). The difference of sensitivity and specificity were obvious after the comparison of the two subgroups of data.

**Sensitivity analysis**

To find out if there was any research affecting the stability regarding PSEN and PSPE, we removed the studies one by one and the results of the sensitivity and specificity analysis are shown in Table III. The results showed that, excluded each single study, the PSEN and PSPE were not significantly influenced and the Higgins $I^2$ also did not change significantly.

**Fagan plot analysis**

The analysis of the Fagan plot showed that S-Detect could give some useful information about the patients’ illness. The PLR and NLR values were calculated by setting the prior probabilities of 25%, 50% and 75% (fig 7). As shown in the figure, when the pre-test probability was 50%, there was 85% probability of the correct diagnosis of breast cancer with S-Detect following the “positive” measurement and lowering the probability of breast cancer with S-Detect to 17% following the “negative” measurement (fig 7b). When the pre-test probabilities were 25% and 75%, the positive post-test probabilities were 66% and 94% while the negative post-test probabilities were 6% and 38%, respectively (fig 7a,c).

**Discussions**

To our best knowledge, this is the first meta-analysis that comprehensively investigates the diagnostic performance of S-Detect technique (one of the most increasingly used ultrasound CAD systems) in the differentiating of benign and malignant breast masses. All selected studies were of high quality, so it is likely that the quality of the studies did not have a significant impact on the results of the meta-analysis. In addition, it is not expected that the meta-analysis overestimates the effect of S-Detect evaluation, since there is no evidence of publication bias.

### Table II. Analysis regarding the S-Detect technology in diagnosis application of the breast masses, results of meta-regression

<table>
<thead>
<tr>
<th>Parameter and category</th>
<th>N</th>
<th>Se(95% CI)</th>
<th>p</th>
<th>Sp(95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>7</td>
<td>0.86(0.81,0.90)</td>
<td>0.16</td>
<td>0.84(0.75,0.93)</td>
<td>0.21</td>
</tr>
<tr>
<td>other</td>
<td>3</td>
<td>0.73(0.63,0.82)</td>
<td>0.88</td>
<td>0.76,1.00</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥200</td>
<td>5</td>
<td>0.86(0.83,0.90)</td>
<td>0.04</td>
<td>0.79(0.67,0.90)</td>
<td>0.01</td>
</tr>
<tr>
<td>&lt;200</td>
<td>5</td>
<td>0.75(0.67,0.83)</td>
<td>0.91(0.83,0.98)</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Edition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.84(0.78,0.89)</td>
<td>0.13</td>
<td>0.87(0.77,0.97)</td>
<td>0.49</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.78(0.66,0.89)</td>
<td>0.86(0.71,1.00)</td>
<td>0.49</td>
<td></td>
</tr>
</tbody>
</table>

N: Number of included studies; CI: confidence interval; Total: total number of breast masses; Se: sensitivity; Sp: specificity
The value of S-Detect for the differential diagnosis of breast masses on ultrasound

In this study, the PSEN and PSPE of S-Detect in the differentiation of breast masses were 0.82 (95%CI: 0.75-0.87) and 0.86 (95%CI: 0.73-0.93), respectively. In a study conducted by Badu-Peprah et al using US alone to diagnose breast masses [27], the sensitivity was 100% (95%CI: 93.2%-100%) higher than S-Detect, but the specificity of 80.4% (95%CI: 66.9%-90.2%) was lower compared to S-Detect. However, in comparison with the study of Kolb et al also using US alone [28], they showed completely different results where the sensitivity and specificity were 75.3% and 96.8% respectively.

Apart from US, there are also some other useful imaging techniques that can be used to screen and diagnose breast cancer such as mammography and MRI which are all anatomy- and morphology-based imaging techniques [29,30]. Mammography is well known to be a powerful screening tool in the detection of early breast cancer and could reduce breast cancer related death, but it is imperfect, particularly for women with dense breasts [31,32]. MRI has become more prevalent in the last few years [33]. Previous meta-analysis showed that MRI combined with CAD system can reduce the number of misdiagno-

<table>
<thead>
<tr>
<th>Delete article</th>
<th>Se (95% CI)</th>
<th>F (95% CI), %</th>
<th>p</th>
<th>Sp (95% CI)</th>
<th>F (95% CI), %</th>
<th>p</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cho [22]</td>
<td>0.84(0.78,0.88)</td>
<td>77.33(62.76,91.89)</td>
<td>&lt;0.01</td>
<td>0.85(0.74,0.92)</td>
<td>91.90(87.98,95.82)</td>
<td>&lt;0.01</td>
<td>0.90(0.87,0.92)</td>
</tr>
<tr>
<td>Du [25]</td>
<td>0.83(0.77,0.88)</td>
<td>78.38(64.66,92.10)</td>
<td>&lt;0.01</td>
<td>0.87(0.76,0.93)</td>
<td>92.80(89.44,96.17)</td>
<td>&lt;0.01</td>
<td>0.90(0.87,0.92)</td>
</tr>
<tr>
<td>Gewefel [19]</td>
<td>0.84(0.79,0.88)</td>
<td>74.87(58.31,91.44)</td>
<td>&lt;0.01</td>
<td>0.83(0.74,0.89)</td>
<td>91.92(88.02,95.83)</td>
<td>&lt;0.01</td>
<td>0.90(0.87,0.92)</td>
</tr>
<tr>
<td>Jiang [13]</td>
<td>0.82(0.76,0.87)</td>
<td>76.18(60.69,91.67)</td>
<td>&lt;0.01</td>
<td>0.85(0.74,0.92)</td>
<td>91.50(87.34,95.67)</td>
<td>&lt;0.01</td>
<td>0.89(0.86,0.91)</td>
</tr>
<tr>
<td>Kim [17]</td>
<td>0.82(0.76,0.87)</td>
<td>79.67(66.98,92.37)</td>
<td>&lt;0.01</td>
<td>0.87(0.78,0.93)</td>
<td>92.31(88.64,95.97)</td>
<td>&lt;0.01</td>
<td>0.90(0.87,0.92)</td>
</tr>
<tr>
<td>Li [16]</td>
<td>0.81(0.75,0.86)</td>
<td>68.01(45.67,90.35)</td>
<td>&lt;0.01</td>
<td>0.87(0.78,0.93)</td>
<td>86.74(79.35,94.12)</td>
<td>&lt;0.01</td>
<td>0.89(0.86,0.91)</td>
</tr>
<tr>
<td>Wang [26]</td>
<td>0.81(0.75,0.86)</td>
<td>73.71(56.19,91.23)</td>
<td>&lt;0.01</td>
<td>0.85(0.74,0.91)</td>
<td>89.07(83.31,94.84)</td>
<td>&lt;0.01</td>
<td>0.88(0.85,0.91)</td>
</tr>
<tr>
<td>Wu [23]</td>
<td>0.81(0.75,0.87)</td>
<td>79.62(66.88,92.36)</td>
<td>&lt;0.01</td>
<td>0.86(0.75,0.93)</td>
<td>92.44(88.85,96.02)</td>
<td>&lt;0.01</td>
<td>0.89(0.86,0.91)</td>
</tr>
<tr>
<td>Zhao [10]</td>
<td>0.82(0.75,0.87)</td>
<td>79.51(66.69,92.34)</td>
<td>&lt;0.01</td>
<td>0.87(0.77,0.93)</td>
<td>92.74(89.33,96.14)</td>
<td>&lt;0.01</td>
<td>0.90(0.87,0.92)</td>
</tr>
<tr>
<td>Zhou [24]</td>
<td>0.83(0.76,0.88)</td>
<td>78.22(64.37,92.07)</td>
<td>&lt;0.01</td>
<td>0.83(0.74,0.90)</td>
<td>91.19(86.82,95.56)</td>
<td>&lt;0.01</td>
<td>0.89(0.86,0.92)</td>
</tr>
</tbody>
</table>

CI: confidence interval; AUC: area under the curve; Se: sensitivity; Sp: specificity
ses and shorten the time required to interpret the image, thus improving the accuracy of diagnosis. However, the limitations of MRI are its high cost, as well as modest specificity resulting in false positive examinations [34].

The earliest study in this meta-analysis on S-Detect technique was published in 2016 [13]. A study by Jeong et al revealed a high diagnostic sensitivity (88.6%) and accuracy (86.0%) of S-Detect, which was not statistically different from an experienced radiologist (84.1% and 91.0%, respectively) but the specificity was significantly lower than the experienced radiologist (83.9 vs. 96.4%) [35]. However, operator dependency needs to be improved in this study.

All the 10 included articles use the BI-RADS to make the final assessment, but they did not use the same cut-off value. BI-RADS 4a was clearly described as the cut-off value in 7 [10,19,22-26] but not in 3 studies [13,16,17]. The 10 studies used different indications for breast masses and there was a wide variation in the number and pathological type of focal lesions selected [10,13,16,17,19,22-26]. The methods for obtaining pathological results are also inconsistent. So, they may have a high heterogeneity inevitably. Therefore, a random-effects model was used to interpret the results with more caution. First, meta-regression was selected to explain the high heterogeneity. It is obvious from the results that the total number of the breast masses may be an important reason for the heterogeneity. 5 included articles had more than 200 masses to differentiate benign or malignant masses [10,16,23,25,26], while other 5 articles had less than 200 masses [13,17,19,22,24]. The p-value of the sensitivity and specificity were 0.04 and 0.01, respectively, which indicated that the result have certain statistical significance. It is demonstrated that the total number of the breast masses in studies may be a cause of heterogeneity. As can be seen from the results, articles which had more than 200 masses [10,16,23,25,26] may have a higher sensitivity than those with less than 200 masses [13,17,19,22,24] (0.86 vs. 0.79) and have a low specificity than those with less than 200 masses [13,17,19,22,24] (0.79 vs. 0.91). Perhaps later studies can further assess the impact of sample size on results by increasing the masses number included. Other possible causes of high heterogeneity have also been analysed. Among the articles included in the analysis, 7 were from China [10,13,16,23-26], the remaining 3 were from other countries including Korea [17,22] and Egypt [19]. The results showed no significant difference in specificity and sensitivity. S-Detect has emerged as a useful tool in the field of medical US over the past few years [9]. But so far, only four countries have conducted clinical trials on this technique including China [9,10,13,16,23,24], South Korea [17,18,22], Egypt [19] and Italy [36,37].

US BI-RADS is considered as a standardized system of reporting the possibility of cancer of a breast lesion imaged on US [38]. In this analysis, 6 of the included studies used the fifth version of the reporting system [10,16,19,22,23,26] while the other 3 studies used the fourth [13,17,24]. The regression analysis showed that there was no significant difference in the results of the study (p=0.05). Still, we could not ignore the impact of version differences on the results because the fifth version of BI-RADS expanded and normalized the contents of the fourth edition, which might lead to differences in the diagnosis of masses [38,39]. And both the sensitivity and specificity of the fifth edition are higher than the fourth edition (0.84 vs. 0.78 and 0.87 vs. 0.86).

What is more, the Fagan plot analysis also has been adopted to explore the clinical practical application of the S-Detect technique. Results indicated that S-Detect had the potential to identify breast malignant masses. Considering a patient who is estimated clinically to have a 50% probability of breast cancer after the initial assessment, the likelihood that this patient has breast cancer if S-Detect is positive increases from 50 to 85%. This high probability would be considered good performance. By contrast, if the result of S-Detect was negative, then the probability that this patient has breast cancer is 17%, which is good to rule out breast cancer with confidence. In actual clinical practice, masses with a pre-test probability of malignancy of 25% would undergo biopsy, regardless of the results of S-detect. But the Fagan plot analysis suggests that S-detect could be a useful aid in providing doctors with diagnostic advice.

Five of the included studies compared the accuracy of diagnosis of breast masses among radiologists with different work experience and S-Detect, concluding that S-Detect is especially helpful for younger radiologists [10,13,16,22,24].

In this meta-analysis, we tried to collect all published studies concerning the S-Detect technique by using extensive and comprehensive search terms and multiple databases, then carried out rigorous evaluation and analysis one by one, and further statistically processed the data through quantitative synthesis to reach comprehensive conclusions. Meanwhile, a relatively objective evaluation of S-Detect technique for breast cancer can be performed by analyzing the heterogeneity of different studies.

This diagnostic meta-analysis has several limitations. S-Detect technique has also been applied in Italy [36,37], but these two articles were excluded for their own reasons. The study done by Bartolotta et al has no data for assessing the diagnostic performance of S-Detect alone.
but this article showed that both inter- and intra-observer agreement of two residents improved after the use of S-Detect, with k values ranging from 0.7 to 0.81 (from good to very good, without statistical significance) [36]. Moreover, the study done by Di Segni et al only selected lesions classified as BI-RADS 4a for further inspection and comparison using S-Detect technique [37]. It also proved that S-Detect is a feasible tool for the characterization of breast lesions and has a potential as a teaching tool for the less experienced operators. Second, the present analysis included a relatively small number of studies (n=10). These factors above may influence the assessment of diagnostic performance of S-Detect. Finally, all articles included were single-centre studies, the research results may have monotonicity and limitations. To further evaluate the role of S-Detect, large, prospective, international, multicentre studies in various regions are necessary.

**Conclusion**

Based on the above, we have obtained reliable evidence that artificial intelligence technology applied to ultrasound has developed rapidly and the S-Detect technique is becoming a useful supplement to traditional ultrasound in clinical practice, especially for inexperienced doctors. Also, this analysis can be used as a basis for further investigation on the S-Detect technique.

**Acknowledgements:** We acknowledge the support of the Corps Science and Technology Key Project (Grant No. 2019DB012) and the National Natural Science Foundation of China (Grant No. 81860498).

**Conflict of interest:** none

**References**