Original papers

Med Ultrason 2020, *Vol. 22, no. 4, 415-423*
DOI: 10.11152/mu-2501

The value of S-Detect in improving the diagnostic performance of radiologists for the differential diagnosis of thyroid nodules

Qi Wei[1], Shu-E Zeng[2], Li-Ping Wang[1], Yu-Jing Yan[1], Ting Wang[1], Jian-Wei Xu[3], Meng-Yi Zhang[1], Wen-Zhi Lv[4], Xin-Wu Cui[1], Christoph F. Dietrich[5]

[1]Department of Medical Ultrasound, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Medical Ultrasound, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3]Department of Medical Ultrasound, The First Affliated Hospital of Zhengzhou University, Zhengzhou, China, [4]Department of Artifcial Intelligence, Julei Technology, Wuhan, China, [5]Department Allgemeine Innere Medizin (DAIM), Kliniken Hirslanden Beau Site, Salem und Permancence, Bern, Switzerland

Abstract

Aims: To compare the diagnostic value of S-Detect (a computer aided diagnosis system using deep learning) in differentiating thyroid nodules in radiologists with different experience and to assess if S-Detect can improve the diagnostic performance of radiologists. **Materials and methods:** Between February 2018 and October 2019, 204 thyroid nodules in 181 patients were included. An experienced radiologist performed ultrasound for thyroid nodules and obtained the result of S-Detect. Four radiologists with different experience on thyroid ultrasound (Radiologist 1, 2, 3, 4 with 1, 4, 9, 20 years, respectively) analyzed the conventional ultrasound images of each thyroid nodule and made a diagnosis of "benign" or "malignant" based on the TI-RADS category. After referring to S-Detect results, they re-evaluated the diagnoses. The diagnostic performance of radiologists was analyzed before and after referring to the results of S-Detect. **Results:** The accuracy, sensitivity, specificity, positive predictive value and negative predictive value of S-Detect were 77.0, 91.3, 65.2, 68.3 and 90.1%, respectively. In comparison with the less experienced radiologists (radiologist 1 and 2), S-Detect had a higher area under receiver operating characteristic curve (AUC), accuracy and specificity ($p < 0.05$). In comparison with the most experienced radiologist, the diagnostic accuracy and AUC were lower ($p<0.05$). In the less experienced radiologists, the diagnostic accuracy, specificity and AUC were significantly improved when combined with S-Detect ($p<0.05$), but not for experienced radiologists (radiologist 3 and 4) ($p>0.05$). **Conclusions:** S-Detect may become an additional diagnostic method for the diagnosis of thyroid nodules and improve the diagnostic performance of less experienced radiologists.

Keywords: thyroid nodule; ultrasound; computer-aided diagnosis; diagnosis

Introduction

Focal thyroid disease is the most common thyroid and endocrine disease [1]. Since 1975, the incidence of thyroid cancer has been rising [2]. It is reported that it was newly found in 567,000 cases worldwide in 2018, and the global incidence in women is three times higher than that of men [3]. Researchers found that up to 30% of the people suffered from clinically occult thyroid cancer through autopsy [4].

The imaging modalities used for thyroid evaluation include ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), scintigraphy [5] and so on. Among them, US is the primary imaging technique for focal and diffuse thyroid disease with the advantage of convenient, radiation-free and high-resolution [6-8]. Since 2009, several different Thyroid Imaging Reporting and Data Systems (TI-RADS) have been proposed to formulate standardized terminology for thyroid ultrasound reporting. The latest TI-RADS was proposed by the American College of Radiology (ACR) in 2017 [9].

Received 10.03.2020 Accepted 22.05.2020
Med Ultrason
2020, Vol. 22, No 4, 415-423
Corresponding author: Prof. Dr. med. Xin-Wu Cui
Department of Medical Ultrasound,
Tongji Hospital, Tongji Medical College,
Huazhong University of Science and
Technology, No. 1095, Jiefang Avenue,
Wuhan 430030, Hubei Province, China
Phone: +86-15927103161
Fax: +86-27-83662640
E-mail: cuixinwu@live.cn

TI-RADS is a useful but not the best reporting tool for describing thyroid disease [10-12]. The size of thyroid nodules will affect the diagnostic performance of TI-RADS [10,13]. Some studies have shown better inter-observer agreement of TI-RADS [14,15], while others are quite the opposite [16-18].

S-Detect (Samsung Ultrasound RS80A, Samsung Medison Co. Ltd., Seoul, South Korea) is a computer-aided diagnosis (CAD) software for ultrasound to identify benign and malignant thyroid nodules and breast lesions. At present, only a few studies have evaluated the diagnostic performance of S-detect for the differential diagnosis of breast masses, and it is believed that S-detect can be used as an additional tool for the diagnosis of breast cancer [19-23]. However, only a few articles reported the diagnostic performance of S-Detect for diagnosing thyroid cancer, and most of them were published by Korean researchers [24-29]. In order to further study the diagnostic value of S-detect in thyroid ultrasound, more validation sets from different countries are required. The diagnostic performance of S-Detect in the differentiation of thyroid nodules may be higher than that of less experienced radiologists and has different added value to radiologists with different experience.

Therefore, the purpose of this study is to compare the diagnostic value of S-Detect with radiologists of different experience in diagnosing benign and malignant thyroid nodules and assessing how helpful S-Detect is for different level radiologists based on the Chinese population.

Materials and methods

This research was approved by the Institutional Review Board of Tongji Medical College of Huazhong University of Science & Technology. Informed consent was achieved from all patients. The diagnostic performance study of S-Detect was of a prospective design.

Patients

Between February 2018 and October 2019, a total of 222 thyroid nodules (≥0.5 cm) from 199 patients who were scheduled for thyroid US examinations and US-guided fine needle aspiration (FNA) were enrolled in this study. Inclusion criteria were: 1) patients at least 18 years old; 2) patients who had available histopathologic results. Exclusion criteria were: 1) patients with a thyroid nodule <5 mm; 2) patients without surgery, FNA and core needle aspiration for thyroid nodules; 3) patients with known pathological findings of thyroid nodules before US examination.

Nonconclusive FNA results were excluded based on the Bethesda system [30]. Finally, a benign thyroid nodule was diagnosed when the benign status was evident in the surgical specimen or FNA. A malignant thyroid nodule was diagnosed when malignancy was evident in the surgical specimen.

US examination

A 3-12 MHz linear transducer (Samsung Ultrasound RS80A, Samsung Medison Co. Ltd., Seoul, South Korea) was used in this study. The EU-TIRADS proposed by Russ et al in 2017 was set as the preferred semi-automatic thyroid assessment in this study [31]. S-Detect was performed by an experienced radiologist with 10 years of clinical experience in the performance of thyroid US to give a diagnostic suggestion of "possibly benign" or "possibly malignant" for the thyroid nodule of interest (fig 1).

First of all, conventional US was performed to scan the whole thyroid gland and the representative images

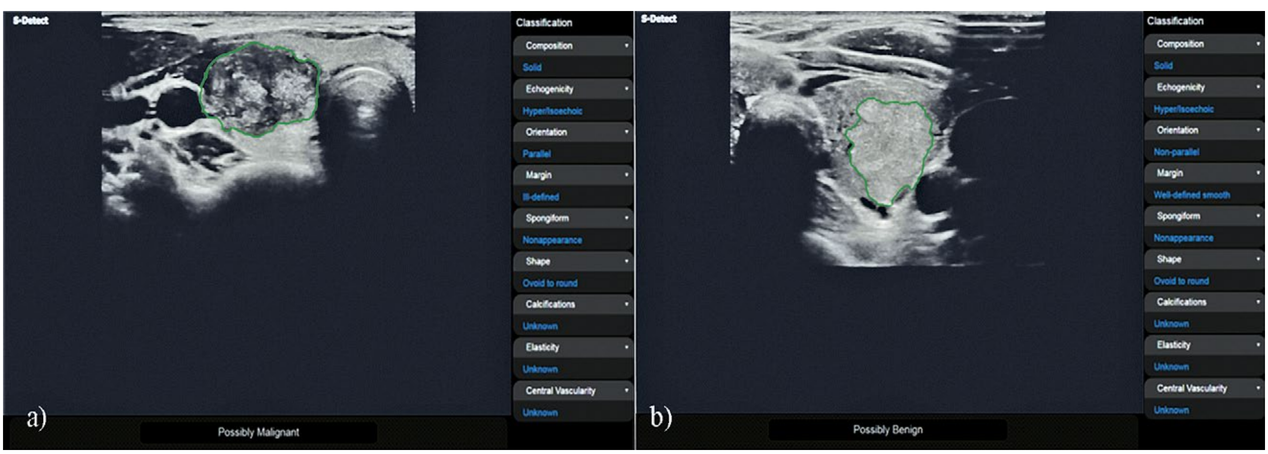

Fig 1. a) A thyroid nodule in a 22-year-old woman with pathological findings suggesting papillary thyroid carcinoma: S-Detect analyzed the characteristics of the lesion and the result was "Possibly Malignant"; b) A thyroid nodule in a 26-year-old woman with pathological findings suggesting nodular goiter: S-Detect judged the lesion as "Possibly Benign".

were selected based on transverse scans. After freezing the US image of the thyroid nodule, a region of interest around the lesion was set manually, then S-Detect was performed on the lesions. Then, the S-Detect software calculated the mass contours automatically and gave several different mass contours options for the radiologist to choose. The radiologist could edit the margin manually of the mass if needed. Meanwhile, US characteristics of the mass, including composition, orientation, shape, margins, echogenicity and spongiform status, were evaluated by the software. Finally, the nodule was diagnosed as “possibly benign” or “possibly malignant” by S-Detect and the size of the nodule was recorded.

Analysis by radiologists

Four other radiologists with different experience were selected to analyze the images, Radiologist 1, 2, 3, 4 had 1, 4, 9 and 20 years of experience in thyroid US, respectively. All radiologists retrospectively analyzed B-mode ultrasound images and evaluated the following features: composition, echogenicity, shape, margin and echogenic foci based on ACR TI-RADS categories [9]. A recent study showed there were no malignant diagnoses of thyroid in TR2 and TR3 (Bethesda, Category VI) [32]. TR3 is of low malignancy, therefore the cut-off between benign and malignant thyroid nodules was set at TR4 in this study. After they finished the analysis, all results suggested by S-Detect were given to them for reference. They re-analyzed all the B-mode images and gave a new category to judge the benign and malignancy of each nodule. The primary outcome was the added value of S-Detect in the differential diagnosis of benign and malignant thyroid nodules by radiologists with different experience. The secondary outcomes included the diagnostic performance of S-Detect in differentiating benign and malignant thyroid nodules compared with radiologists with different experience and intra-observer differences in different vocabulary items.

Statistical analysis

Accuracy, specificity, sensitivity, positive predictive value (PPV) and negative predictive value (NPV) were calculated to show the diagnostic performance of S-Detect, the individual radiologists and the integration of S-Detect with each radiologist in differentiating benign and malignant thyroid nodules. The diagnostic performances between S-Detect and radiologists were compared using the McNemar test (accuracy, sensitivity and specificity) and the generalized score statistical method (PPV and NPV) [33]. The difference in the description of the characteristics of ultrasound between radiologists was compared by the chi-square test. The area under the receiver operating characteristic curve (AUC) for the S-Detect and four radiologists were compared by the Delong method [34]. The ROC curve was plotted by R software pROC, version 1.15.3 (https://CRAN.R-project.org/package=pROC). All statistical analyses were done with R software, version 3.6.0, (http://www.R-project.org). For all tests mentioned above, a p value of <0.05 was considered significant.

Results

Characteristics of the thyroid nodules

A total of 204 thyroid nodules were evaluated in 181 patients (35 males and 146 females), including 112 (54.9%) benign lesions and 92 (45.1%) malignant lesions (fig 2). The mean age of the patients was 46±12 years (range, 18 to 74 years). The mean nodule diameter was 1.5 ± 1.1cm (range 0.5–5.2 cm). All malignant diagnoses were confirmed after surgical resection, and included 90 papillary thyroid carcinomas (PTCs), 1 follicular thyroid carcinoma and 1 B-cell non-Hodgkin lymphoma. The 68 surgically confirmed benign nodules included 9 follicular adenoma, 43 mutilnodular goiter (MNG), 12 Hashimoto’s thyroiditis (HT), 3 subacute thyroiditis (SAT) and 1 Hürthle cell adenoma (HCA).

The diagnostic performance of S-Detect and radiologists

The diagnostic performance of four radiologists and S-Detect for thyroid cancer were demonstrated in Table I. The corresponding ROC curves were shown in fig 3. As the experience of radiologists increased, their diagnostic accuracy increased accordingly.

In comparison with the most experienced radiologist (radiologist 4), S-Detect had lower accuracy (84.8% vs 77.0%, p=0.010), PPV (76.1% vs 68.3%, p=0.016), NPV (96.6% vs 90.1%, p=0.040) and AUC value (0.859 vs 0.782, p=0.005). The sensitivity and specificity of S-Detect showed no statistical significance between S-Detect

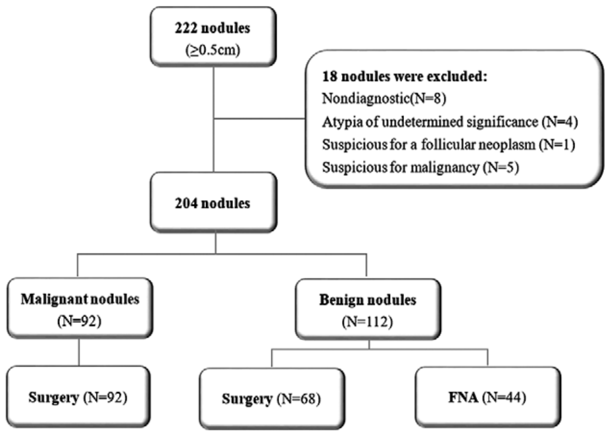

Fig 2. Flowchart of the study

Table I. The comparison of the diagnostic performance of S-Detect and four radiologists.

Diagnostic value	S-Detect	Radiologist 1	p[1]	Radiologist 2	p[2]	Radiologist 3	p[3]	Radiologist 4	p[4]
Accuracy,%	77.0 (70.7-82.2)	63.7 (56.9-70.0)	<0.001	65.2 (58.4-71.4)	0.001	72.5 (66.1-78.2)	0.176	84.8 (79.2-89.1)	0.010
Sensitivity,%	91.3 (83.7-95.5)	95.7 (89.4-98.3)	0.289	84.8 (76.1-90.7)	0.180	83.7 (74.8-89.9)	0.092	96.7 (90.9-98.9)	0.125
Specificity,%	65.2 (56.0-73.4)	37.5 (29.1-46.7)	<0.001	49.1 (40.0-58.2)	0.005	63.4 (54.2-71.7)	0.832	75.0 (66.2-82.1)	0.052
PPV, %	68.3 (59.6-75.9)	55.7 (47.9-63.2)	<0.001	57.8 (49.3-65.8)	0.001	65.3 (56.3-73.2)	0.279	76.1 (67.6-82.9)	0.016
NPV, %	90.1 (81.7-94.9)	91.3 (79.7-96.6)	0.779	79.7 (68.8-87.5)	0.029	82.6 (73.2-89.1)	0.048	96.6 (90.4-98.9)	0.040
AUC	0.782 (0.718-0.847)	0.666 (0.592-0.739)	<0.001	0.669 (0.595-0.744)	0.001	0.7354 (0.666-0.805)	0.100	0.859 (0.805-0.913)	0.005

p[1]: Comparison between S-Detect and radiologist 1; p[2]: Comparison between S-Detect and radiologist 2; p[3]: Comparison between S-Detect and radiologist 3; p[4]: Comparison between S-Detect and radiologist 4; PPV: positive predictive value; NPV: negative predictive value; AUC: area under receiver operating characteristic curve.

and the most experienced radiologist (91.3% vs 96.7%, p=0.125; 65.2% vs 75.0%, p=0.052, respectively).

Compared to the less experienced radiologists (radiologists 1 and 2), S-Detect had higher accuracy, specificity, PPV, AUC value (all p<0.05, respectively) and a similar sensitivity (91.3% vs 95.7%, p=0.289; 91.3% vs 84.8%, p=0.180, respectively). However, the NPV of S-Detect was higher than that of radiologist 2 (90.1% vs 79.7%, p=0.029), and there was no significant difference compared with radiologist 1 (90.1% vs 91.3%, p=0.779).

The NPV of S-Detect was higher than that of radiologist 3 (90.1% vs 82.6%, p=0.048), but the accuracy, sensitivity, specificity, PPV and AUC value did not show significant differences to radiologist 3 (all p>0.05, respectively).

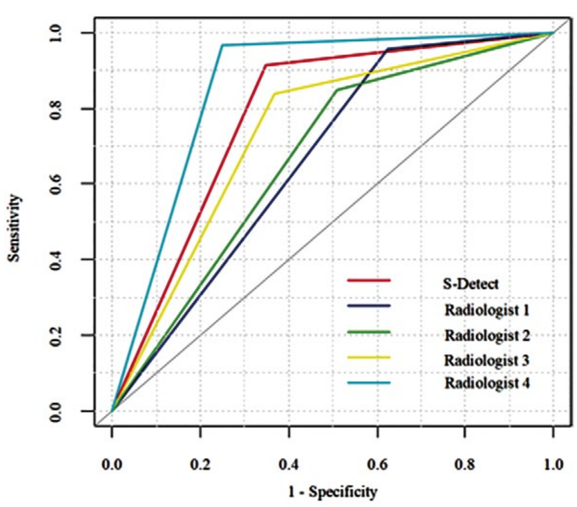

Fig 3. ROC curves of S-Detect and radiologists.

The diagnostic performance of radiologists after referring S-Detect

The four radiologists were asked to make a new judgement of each thyroid nodule referring to the results of S-Detect, and the changes they made were shown in Table II. The least experienced radiologist (radiologist 1) made the most changes, while the most experienced radiologist (radiologist 4) made the lowest number of changes. The diagnostic performance of the four radiologists after referring to the results of S-Detect is shown in Table III.

When S-Detect was used to assist the less experienced radiologists (radiologists 1 and 2), the diagnostic accuracy (63.7% vs 75.0%, p<0.001; 65.2% vs 74.5%, p<0.001, respectively), specificity (37.5% vs 58.9%, p<0.001; 49.1% vs 59.8%, p=0.002, respectively), PPV (55.7% vs 65.4%, p<0.001; 57.8% vs 65.4%, p<0.001, respectively) and AUC values (0.666 vs 0.767, p<0.001; 0.669 vs 0.761, p<0.001, respectively) significantly improved, whereas the sensitivity did not show significant differences (p>0.999 and *p*=0.065, respectively). The NPV showed significant improvement in radiologist 2 (79.7% vs 90.5%, p=0.012), but there was no significant

Table II. Changes made by four radiologists after referencing S-Detect.

Radiologist	Number of changes	Correct number	Inaccurate number
Radiologist 1	31	27	4
Radiologist 2	25	22	3
Radiologist 3	9	7	2
Radiologist 4	5	3	2

Table III. The diagnostic performance of four radiologists after referencing S-Detect.

Diagnostic value	Radiologist 1 +S-Detect	p[a]	Radiologist 2 + S-Detect	p[b]	Radiologist 3 + S-Detect	p[c]	Radiologist 4 + S-Detect	p[d]
Accuracy, %	75.0 (68.6-80.4)	<0.001	74.5 (68.1-80.0)	<0.001	75.0 (68.6-80.4)	0.182	85.3 (79.8-89.5)	>0.999
Sensitivity, %	94.6 (87.9-97.7)	>0.999	92.4 (85.1-96.3)	0.065	88.0 (79.9-93.2)	0.125	97.8 (92.4-99.4)	>0.999
Specificity, %	58.9 (49.7-67.6)	<0.001	59.8 (50.6-68.4)	0.002	64.3 (55.1-72.6)	>0.999	75.0 (66.2-82.1)	>0.999
PPV, %	65.4 (57.0-73.0)	<0.001	65.4 (56.9-73.0)	<0.001	66.9 (58.2-74.7)	0.221	76.3 (67.8-83.0)	0.877
NPV, %	93.0 (84.6-97.0)	0.592	90.5 (81.7-95.3)	0.012	86.7 (77.8-92.4)	0.042	97.7 (91.9-99.4)	0.318
AUC	0.767 (0.702-0.833)	<0.001	0.761 (0.695-0.828)	<0.001	0.762 (0.695-0.829)	0.074	0.864 (0.811-0.917)	0.604

p[a]: Comparison between radiologist 1 and radiologist 1 combined with S-Detect; p[b]: Comparison between radiologist 2 and radiologist 2 combined with S-Detect; p[c]: Comparison between radiologist 3 and radiologist 3 combined with S-Detect; p[d]: Comparison between radiologist 4 and radiologist 4 combined with S-Detect; PPV: positive predictive value; NPV: negative predictive value; AUC: area under receiver operating characteristic curve.

Table IV. Intra-observer variability of US characteristics and the difference between the most experienced radiologist and the least experienced radiologist.

Characteristics	Intra-observer difference, % of agreement between 2 examinations				Difference (%) (95% CI)	p
	Radiologist 1 (%)	Radiologist 2 (%)	Radiologist 3 (%)	Radiologist 4 (%)		
Composition	98.5	99.0	99.5	100.0	1.5 (-0.181 to 3.122)	0.082
Echogenicity	98.5	98.0	99.5	100.0	1.5 (-0.181 to 3.122)	0.082
Shape	99.0	99.5	100.0	100.0	1.0 (-0.372 to 2.333)	0.156
Margin	91.7	94.6	97.1	97.5	8.3 (1.536 to 10.228)	0.009
Echogenic foci	97.1	96.6	99.5	100.0	2.9 (0.623 to 5.260)	0.014

p value of radiologist 1 (the least experienced radiologist) versus radiologist 4 (the most experienced radiologist); comparison of proportions for radiologist 1 and radiologist 4.

difference compared with radiologist 1 (91.3% vs 93.0%, p=0.592) with the assistance of S-Detect.

In the experienced radiologists (radiologists 3 and 4), the diagnostic performance including accuracy, sensitivity, specificity, PPV and the AUC value were not significantly improved when combined with S-Detect (all p>0.05, respectively). The ROC curve between the re-diagnosis results of four radiologists and S-Detect is shown in fig 4.

Intra-observer variability in the recording of US characteristics before and after referring to the S-Detect result

The intra-observer variability of US characteristics and the difference between the most experienced radiologist and the least experienced radiologist are shown in Table IV. Compared with the most experienced radiologist, the margin is the most difficult to assess for the least experienced radiologist.

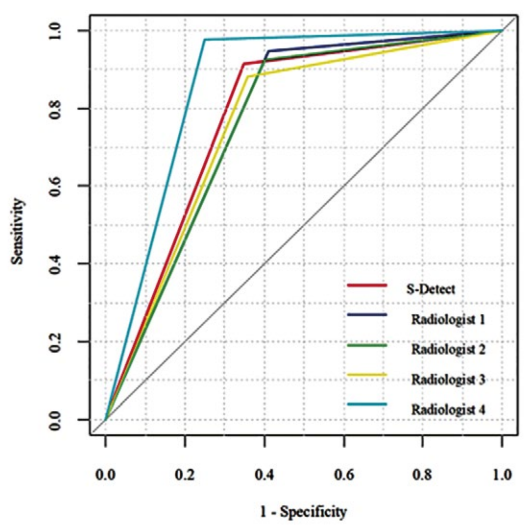

Fig 4. The ROC curve of S-Detect and radiologists after referring to S-Detect.

Discussion

This study evaluated the usefulness of S-Detect in the differential diagnosis of benign and malignant thyroid nodules, especially its value in less experienced and experienced thyroid radiologists. The diagnostic performance of S-Detect in thyroid nodules was equivalent to a radiologist with 9 years of work experience. The computer-aided (CAD) system is significantly helpful for less experienced radiologists (radiologists 1 and 2) to improve their diagnostic performance, but not for experienced radiologists.

FNA is a safe, simple, and the most cost-effective diagnostic method for evaluating thyroid nodules, which plays a critical role in the selection of surgical, interventional or conservative management [35,36]. In most cases, FNA can clearly determine benign or malignant thyroid diseases, but FNA cannot reliably exclude cancers in 20% to 30% of thyroid nodules [37]. Moreover, there has been a problem with over-diagnosis and overtreatment because of the slow growth and lower invasiveness of thyroid cancers [38]. The CAD system is expected to have high specificity and accuracy to avoid unnecessary FNA. Some studies reported that the CAD system showed a similar sensitivity but lower accuracy and specificity compared to the experienced radiologist [27,28]. The CAD system still showed similar sensitivity and lower specificity when an experienced radiologist stratified the risk of malignancy according to TI-RADS and American Thyroid Association (ATA) guidelines [39]. These studies show that the diagnostic sensitivity of CAD system is high, but the diagnostic specificity of CAD system needs to be improved. Different from these studies, our study not only shows similar sensitivity to radiologists, but also shows similar specificity. In our study, the diagnostic specificity and accuracy of S-Detect were comparable to the experienced radiologists and significantly higher than that of less experienced radiologists. When combined with the results of S-Detect, the specificity and accuracy of less experienced radiologists were significantly improved. On the other hand, studies have reported the high sensitivity of S-Detect and concluded that S-Detect may help to rule out thyroid malignancies and ultimately avoid unnecessary FNA [25,27]. Therefore, we can conclude that S-Detect may be used to reduce the number of unnecessary FNA of benign thyroid nodules.

The diagnostic performance of S-Detect was good (AUC=0.782), which was significantly higher than the less experienced radiologists (radiologists 1 and 2) and lower than the most experienced radiologist with 20 years of work experience. Therefore, the results of S-Detect are more valuable when less experienced radiologists have difficulty in diagnosing a thyroid nodule. However, some studies reported that the diagnostic accuracy of the CAD system is as high as 100% [40,41]. An important reason why these studies differ from our study may be that these exploratory studies do not involve the clinical environment and radiologists.

In our study, after referring to the results of S-Detect, four radiologists re-evaluated the 204 thyroid nodules. As the working experience of radiologists increased, the fewer changes were made. It suggests that junior radiologists are more willing to trust the diagnosis results of S-Detect and make changes, while senior radiologists are more willing to believe in their own judgement. The diagnostic accuracy, specificity, PPV and AUC values of less experienced radiologists were significantly improved when integrating the results of S-Detect, whereas the diagnostic performance of experienced radiologists had not been significantly improved. This result shows that S-Detect is of great assistance to the less experienced radiologists. However, the role of S-Detect in improving the diagnostic performance of experienced radiologists was not significant. Yoo et al [26] reported that the sensitivity of the radiologist was significantly improved with the assistance of CAD systems (92.0% vs. 84.0%, p=0.037). Therefore, the CAD system can be reliable as a second opinion to improve the diagnostic performance of radiologists, especially the less experienced radiologists.

Barczyński et al evaluated 50 thyroid nodules (including 10 papillary thyroid carcinomas and 40 benign nodules) and found that the intra-observer variability of the surgeon with ordinary US skills was significantly higher than surgeons with professional US skills in assessing the lexicon of margin, shape, calcification and spongiform [29]. In our study, the least experienced radiologist also had significantly higher intra-observer variability than the most experienced radiologist in assessing the margin and echogenic foci of the thyroid lexicon items. The S-Detect model cannot evaluate three other items: calcification, elasticity and central vascularity. Wang et al [42] suggested that calcification found in ultrasound examination should increase the suspicion of thyroid cancer. The malignancy risk is very high when calcification is found in a solitary thyroid nodule [43]. ACR TI-RADS explained the calcification lexicon. Large comet-tail artifacts strongly indicate benign, the correlation between macrocalcifications and peripheral calcifications and malignant tumors is variable, and punctate echogenic foci are considered highly suspicious [9]. Some studies have also shown that microcalcification is valuable for predicting thyroid malignant diseases [44,45]. The European Medical and Biological Association guidelines point out that elastography is another tool for the identification of thyroid lesions and can be used for follow-up guidance of FNA negative lesions [46]. The limitation of elastography is that internal calcification may affect the stiffness of the lesion [46]. For vascularity, studies have shown that they have little value for the differentiation of thyroid nodules [47,48].

There are several limitations in our study. Firstly, the images used in the research are all static two-dimensional images and the radiologists may have missed some important information while reading the images of thyroid nodules. Secondly, the CAD diagnosis using S-Detect has a limitation in the evaluation of nodule calcification [24]. Thirdly, the histopathological results were not all obtained by surgical resection; some benign thyroid nodules in this study were derived from US-guided FNA. The 2015 edition of the American Thyroid Association (ATA) guidelines stated that if the nodule was benign on cytology, further immediate diagnostic studies or treatments were not required [49]. Therefore, some patients with benign cytology were unable to obtain pathological results. Last but not least, intra-observer variability was not assessed in this study.

In **conclusion**, the diagnostic performance of S-Detect was significantly higher than that of the less experienced radiologists and equivalent to an experienced radiologist with 9 years of work experience. In comparison with the most experienced radiologist, S-Detect had similar sensitivity and specificity but lower accuracy including the AUC value. As an auxiliary tool, S-Detect can significantly improve the diagnostic performance of less experienced radiologists.

Acknowledgments: This research was supported by the Corps Science and Technology Key Project (No.2019DB012), International Talent Cooperation Project of Henan (No. 2015GH7) and the Young and Middle-aged Medical Talents Project of Wuhan.

Conflict of interest: none

References

1. Frates MC, Benson CB, Charboneau JW, et al. Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology 2005;237:794-800.
2. Milano AF. Thyroid Cancer: 20-Year Comparative Mortality and Survival Analysis of Six Thyroid Cancer Histologic Subtypes by Age, Sex, Race, Stage, Cohort Entry Time-Period and Disease Duration (SEER*Stat 8.3.2) A Systematic Review of 145,457 Cases for Diagnosis Years 1993-2013. J Insur Med 2018;47:143-158.
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.
4. Harach HR, Franssila KO, Wasenius VM. Occult papillary carcinoma of the thyroid. A "normal" finding in Finland. A systematic autopsy study. Cancer 1985;56:531-538.
5. Hoang JK, Sosa JA, Nguyen XV, Galvin PL, Oldan JD. Imaging thyroid disease: updates, imaging approach, and management pearls. Radiol Clin North Am 2015;53:145-161.
6. Sholosh B, Borhani AA. Thyroid ultrasound part 1: technique and diffuse disease. Radiol Clin North Am 2011;49:391-416.
7. Dighe M, Barr R, Bojunga J, et al. Thyroid Ultrasound: State of the Art Part 1 – Thyroid Ultrasound reporting and Diffuse Thyroid Diseases. Med Ultrason 2017;19:79-93.
8. Dighe M, Barr R, Bojunga J, et al. Thyroid Ultrasound: State of the Art. Part 2 - Focal Thyroid Lesions. Med Ultrasony 2017;19:195-210.
9. Tessler FN, Middleton WD, Grant EG, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol 2017;14:587-595.
10. Cheng SP, Lee JJ, Lin JL, Chuang SM, Chien MN, Liu CL. Characterization of thyroid nodules using the proposed thyroid imaging reporting and data system (TI-RADS). Head Neck 2013;35:541-547.
11. Cantisani V, David E, Grazhdani H, et al. Prospective Evaluation of Semiquantitative Strain Ratio and Quantitative 2D Ultrasound Shear Wave Elastography (SWE) in Association with TIRADS Classification for Thyroid Nodule Characterization. Ultraschall Med 2019;40:495-503.
12. Cosgrove D, Barr R, Bojunga J, et al. WFUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography: Part 4. Thyroid. Ultrasound Med Biol 2017;43:4-26.
13. Dong Y, Wang WP, Xu Y, Cao J, Mao F, Dietrich CF. Point shear wave speed measurement in differentiating benign and malignant focal liver lesions. Med Ultrason 2017; 19: 259-264.
14. Ahmadi S, Oyekunle T, Jiang X, et al. A direct comparison of the ATA and TI-RADS ultrasound scoring systems. Endocr Pract 2019;25:413-422.
15. Russ G, Royer B, Bigorgne C, Rouxel A, Bienvenu-Perrard M, Leenhardt L. Prospective evaluation of thyroid imaging reporting and data system on 4550 nodules with and without elastography. Eur J Endocrinol 2013;168:649-655.
16. Friedrich-Rust M, Meyer G, Dauth N, et al. Interobserver agreement of Thyroid Imaging Reporting and Data System (TIRADS) and strain elastography for the assessment of thyroid nodules. PLoS One 2013;8:e77927.
17. Hoang JK, Middleton WD, Farjat AE, et al. Interobserver Variability of Sonographic Features Used in the American College of Radiology Thyroid Imaging Reporting and Data System. AJR Am J Roentgenol 2018;211:162-167.

18. Itani M, Assaker R, Moshiri M, Dubinsky TJ, Dighe MK. Inter-observer Variability in the American College of Radiology Thyroid Imaging Reporting and Data System: In-Depth Analysis and Areas for Improvement. Ultrasound Med Biol 2019;45:461-470.
19. Kim K, Song MK, Kim EK, Yoon JH. Clinical application of S-Detect to breast masses on ultrasonography: a study evaluating the diagnostic performance and agreement with a dedicated breast radiologist. Ultrasonography 2017;36:3-9.
20. Zhao C, Xiao M, Jiang Y, et al. Feasibility of computer-assisted diagnosis for breast ultrasound: the results of the diagnostic performance of S-detect from a single center in China. Cancer Manag Res 2019;11:921-930.
21. Cho E, Kim EK, Song MK, Yoon JH. Application of Computer-Aided Diagnosis on Breast Ultrasonography: Evaluation of Diagnostic Performances and Agreement of Radiologists According to Different Levels of Experience. J Ultrasound Med 2018;37:209-216.
22. Choi JH, Kang BJ, Baek JE, Lee HS, Kim SH. Application of computer-aided diagnosis in breast ultrasound interpretation: improvements in diagnostic performance according to reader experience. Ultrasonography 2018;37:217-225.
23. Di Segni M, de Soccio V, Cantisani V, et al. Automated classification of focal breast lesions according to S-detect: validation and role as a clinical and teaching tool. J Ultrasound 2018;21:105-118.
24. Kim HL, Ha EJ, Han M. Real-World Performance of Computer-Aided Diagnosis System for Thyroid Nodules Using Ultrasonography.Ultrasound Med Biol 2019;45:2672-2678.
25. Jeong EY, Kim HL, Ha EJ, Park SY, Cho YJ, Han M. Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. Eur Radiol 2019;29:1978-1985.
26. Yoo YJ, Ha EJ, Cho YJ, Kim HL, Han M, Kang SY. Computer-Aided Diagnosis of Thyroid Nodules via Ultrasonography: Initial Clinical Experience. Korean J Radiol 2018;19:665-672.
27. Choi YJ, Baek JH, Park HS, et al. A Computer-Aided Diagnosis System Using Artificial Intelligence for the Diagnosis and Characterization of Thyroid Nodules on Ultrasound: Initial Clinical Assessment. Thyroid 2017;27:546-552.
28. Chung SR, Baek JH, Lee MK, et al. Computer-Aided Diagnosis System for the Evaluation of Thyroid Nodules on Ultrasonography: Prospective Non-Inferiority Study according to the Experience Level of Radiologists. Korean J Radiol 2020;21:369-376.
29. Barczynski M, Stopa-Barczynska M, Wojtczak B, Czarniecka A, Konturek A. Clinical validation of S-Detect(TM) mode in semi-automated ultrasound classification of thyroid lesions in surgical office. Gland Surg 2020;9:S77-S85.
30. Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. Thyroid 2017;27:1341-1346.
31. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. Eur Thyroid Jl 2017;6:225-237.
32. Modi L, Sun W, Shafizadeh N, et al. Does a higher American College of Radiology Thyroid Imaging Reporting and Data System (ACR TI-RADS) score forecast an increased risk of malignancy? A correlation study of ACR TI-RADS with FNA cytology in the evaluation of thyroid nodules. Cancer Cytopathol 2020 Feb 20. doi:10.1002/cncy.22254.
33. Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. Biometrics 2000;56:345-351.
34. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-845.
35. Sakorafas GH. Thyroid nodules; interpretation and importance of fine-needle aspiration (FNA) for the clinician - practical considerations. Surg Oncol 2010;19:e130-e139.
36. Dietrich CF, Muller T, Bojunga J, et al. Statement and Recommendations on Interventional Ultrasound as a Thyroid Diagnostic and Treatment Procedure. Ultrasound Med Biol 2018;44:14-36.
37. Zhu LC, Ye YL, Luo WH, et al. A model to discriminate malignant from benign thyroid nodules using artificial neural network. PLoS One 2013;8:e82211.
38. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic"--screening and overdiagnosis. N Engl J Med 2014;371:1765-1767.
39. Gao L, Liu R, Jiang Y, et al. Computer-aided system for diagnosing thyroid nodules on ultrasound: A comparison with radiologist-based clinical assessments. Head Neck 2018;40:778-783.
40. Acharya UR, Faust O, Sree SV, Molinari F, Suri JS. ThyroScreen system: high resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform. Comput Methods Programs Biomed 2012;107:233-241.
41. Acharya UR, Vinitha Sree S, Krishnan MM, Molinari F, Garberoglio R, Suri JS. Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan systems. Ultrasonics 2012;52:508-520.
42. Wang N, Xu Y, Ge C, Guo R, Guo K. Association of sonographically detected calcification with thyroid carcinoma. Head Neck 2006;28:1077-1083.
43. Khoo ML, Asa SL, Witterick IJ, Freeman JL. Thyroid calcification and its association with thyroid carcinoma. Head Neck 2002;24:651-655.
44. Gungor B, Polat AK, Polat C, Seren D, Erzurumlu K. Do the calcifications in the thyroid gland predict malignancy? Bratisl Lek Listy 2012;113:552-555.
45. Chammas MC, de Araujo Filho VJ, Moyses RA, et al. Predictive value for malignancy in the finding of microcalcifi-

cations on ultrasonography of thyroid nodules. Head Neck 2008;30:1206-1210.
46. Cosgrove D, Piscaglia F, Bamber J, et al. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 2: Clinical applications. Ultraschall Med 2013;34:238-253.
47. Tamsel S, Demirpolat G, Erdogan M, et al. Power Doppler US patterns of vascularity and spectral Doppler US parameters in predicting malignancy in thyroid nodules. Clin Radiol 2007;62:245-251.
48. Moon HJ, Kwak JY, Kim MJ, Son EJ, Kim EK. Can vascularity at power Doppler US help predict thyroid malignancy? Radiology 2010;255:260-269.
49. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 2016;26:1-133.