Computer-Aided Diagnosis Systems in Diagnosing Malignant Thyroid Nodules on Ultrasonography: A Systematic Review and Meta-Analysis

Lei Xu[a, b]   Junling Gao[a]   Quan Wang[c]   Jichao Yin[b]   Pengfei Yu[d]   Bin Bai[d]
Ruixia Pei[b]   Dingzhang Chen[d]   Guochun Yang[b]   Shiqi Wang[d]   Mingxi Wan[a]

[a]Key Laboratory of Biomedical Information Engineering of the Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China; [b]Xi'an Hospital of Traditional Chinese Medicine, Xi'an, China; [c]Laboratory of Surgical Oncology, Peking University People's Hospital, Peking University, Beijing, China; [d]Xijing Hospital, Fourth Military Medical University, Xi'an, China

Abstract
Background: Computer-aided diagnosis (CAD) systems are being applied to the ultrasonographic diagnosis of malignant thyroid nodules, but it remains controversial whether the systems add any accuracy for radiologists. ***Objective:*** To determine the accuracy of CAD systems in diagnosing malignant thyroid nodules. ***Methods:*** PubMed, EMBASE, and the Cochrane Library were searched for studies on the diagnostic performance of CAD systems. The diagnostic performance was assessed by pooled sensitivity and specificity, and their accuracy was compared with that of radiologists. The present systematic review was registered in PROSPERO (CRD42019134460). ***Results:*** Nineteen studies with 4,781 thyroid nodules were included. Both the classic machine learning- and the deep learning-based CAD system had good performance in diagnosing malignant thyroid nodules (classic machine learning: sensitivity 0.86 [95% CI 0.79–0.92], specificity 0.85 [95% CI 0.77–0.91], diagnostic odds ratio (DOR) 37.41 [95% CI 24.91–56.20]; deep learning: sensitivity 0.89 [95% CI 0.81–0.93], specificity 0.84 [95% CI 0.75–0.90], DOR 40.87 [95% CI 18.13–92.13]). The diagnostic performance of the deep learning-based CAD system was comparable to that of the radiologists (sensitivity 0.87 [95% CI 0.78–0.93] vs. 0.87 [95% CI 0.85–0.89], specificity 0.85 [95% CI 0.76–0.91] vs. 0.87 [95% CI 0.81–0.91], DOR 40.12 [95% CI 15.58–103.33] vs. DOR 44.88 [95% CI 30.71–65.57]). ***Conclusions:*** The CAD systems demonstrated good performance in diagnosing malignant thyroid nodules. However, experienced radiologists may still have an advantage over CAD systems during real-time diagnosis.

© 2019 European Thyroid Association
Published by S. Karger AG, Basel

Lei Xu, Junling Gao, and Quan Wang contributed equally to this work.

Mingxi Wan
Key Laboratory of Biomedical Information Engineering of the Ministry of Education
Department of Biomedical Engineering, School of Life Science and Technology
Xi'an Jiaotong University, Xianningxi St. 28, Xi'an 710049 (China)
E-Mail mxwan@mail.xjtu.edu.cn

Shiqi Wang
Xijing Hospital
Fourth Military Medical University
Changlexi St. 127, Xi'an 710032 (China)
E-Mail wsqfmmu@126.com

Introduction

With the development of imaging techniques and popularized medical surveillance, more thyroid nodules are detected [1, 2]. Among the general population, the incidence of thyroid nodules ranges from 19 to 68% [3], and 9–15% are determined to be malignant [4–6]. Ultrasound is the first-line method for identifying malignant thyroid nodules [3], but the diagnostic performance of ultrasound relies heavily on the clinical experience of the radiologists.

To improve the diagnostic accuracy and efficiency, machine learning-based computer-aided diagnosis (CAD) systems are being introduced in the diagnosis process. Currently, two types of machine learning method are adopted: (1) the classic machine learning method, which is based on features identified by human experts, and (2) the deep learning technique, which takes raw image pixels and corresponding class labels from medical imaging data as inputs and automatically learns feature representation in a general manner [7]. Theoretically, CAD systems may improve diagnostic accuracy by decreasing radiologists' subjectivity. However, it is unclear whether the CAD systems provide any help to radiologists in increasing diagnostic accuracy in clinical practice. Some studies were performed without external validation, and potential overfitting cannot be excluded [8–10]; some studies may have underestimated the diagnostic performance of radiologists by setting rigid diagnostic criteria and providing static ultrasound images, and the superiority of CAD systems over radiologists should be reconsidered. Additionally, it is also unclear whether deep learning-based CAD systems outperform classic machine learning-based systems in diagnosis.

Accordingly, it remains to be determined whether there is adequate evidence to support any clinical application of the current CAD systems. The present systematic review and meta-analysis was performed to assess the accuracy of CAD systems in diagnosing malignant thyroid nodules, and to compare the diagnostic performance of the CAD systems with that of radiologists.

Methods

Search Strategy and Eligibility Criteria

The present systematic review was registered in PROSPERO (CRD42019134460). The PubMed, EMBASE, and Cochrane Library databases were searched from inception until May 5, 2019, for studies that assessed the performance of CAD systems in differentiating malignant and benign thyroid nodules on ultrasound images. The search was updated on October 20, 2019. The details of the search strategy are available on https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019134460.

Study Selection and Data Extraction

The general characteristics of the included studies, and the numbers of true-positive, false-positive, false-negative, and true-negative cases were collected. Only data from the validation cohort were included in the meta-analysis to assess the diagnostic performance.

When multiple algorithms or radiologists were involved, only the one with the highest accuracy or largest AUC was selected for the analysis. When the performance of the CAD system was assessed by multiple external validation groups, only the one with the largest cohort was selected for the analysis. When more than one radiologist participated in the assessment, only the most experienced one was selected for the analysis. Both pathological examination of the surgical specimen and cytological examination of fine needle aspiration tissue were considered acceptable reference standards. A low-risk ultrasound index was also accepted as a reference standard for diagnosing benign nodules [11].

Grouping

According to whether the classification features were set in advance or automatically recognized, the CAD systems were classified into a classic machine learning group and a deep learning group. According to their availability for application in real-time clinical diagnosis, the CAD systems were further classified into a real-time subgroup and an ex post subgroup, and their diagnostic performances were assessed. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [12].

Study Quality Assessment

The methodological quality of each study was assessed by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) rating system [13].

Data Analysis

The statistical analysis was performed with STATA version 15.0 software for Windows (StataCorp, College Station, TX, USA). Hierarchical summary ROC curves were constructed. The pooled sensitivity, specificity, diagnostic odds ratio (DOR), and AUC with 95% CI were calculated using the bivariate model. Meta-regression analysis was not conducted, due to the small number of included studies. Publication bias was evaluated using Deeks' test for funnel plot asymmetry. Interstudy heterogeneity was assessed by the DerSimonian-Laird random-effects model and the index of inconsistency (I^2). The combined estimates for sensitivity and specificity were performed by a random-effects model if $I^2 < 50\%$ and by a fixed-effect model if $I^2 \geq 50\%$. A p value <0.05 was considered statistically significant.

Results

Literature Searches and Description of Studies

The flow diagram of the literature search is shown in Figure 1. Nineteen studies with 4,781 nodules used in external validation sets were included in the study, including 6 studies on classic machine learning-based CAD sys-
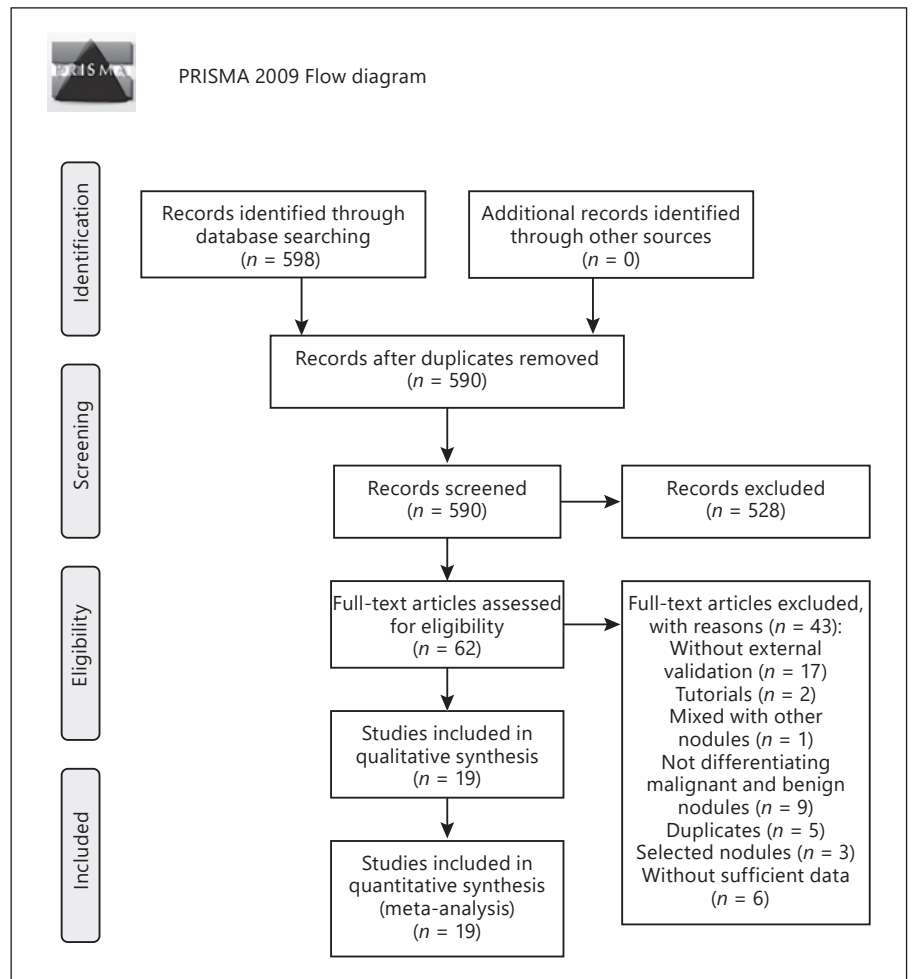
Fig. 1. Search, inclusion, and exclusion flow diagram.

tems [14–19] and 13 studies on deep learning-based CAD systems [7, 20–31]. The general characteristics of the included studies are shown in Table 1, and the detailed characteristics are demonstrated in online supplementary Table 1 (see www.karger.com/doi/10.1159/000504390 for all online suppl. material). No significant publication bias in the studies on deep learning-based CAD systems was demonstrated by Deeks' funnel plot ($p = 0.39$) (online suppl. Fig. 1).

Methodological Quality of the Included Studies

The quality of the included studies is summarized in online supplementary Table 2. The risk of bias from patient selection was judged to be high or unclear in 13 of the included studies: 4 studies limited the nodule size within a certain scope [16, 17, 21, 25]; 5 studies excluded difficult-to-diagnose nodules [15, 25–27, 31]; and 4 studies were unclear about whether there were selected co-

horts and inappropriate exclusions [14, 19, 23, 29]. The risk of bias from the reference standard was considered to be unclear in 2 of the included studies [14, 23]. The risk of bias from flow and timing was considered to be high or unclear in 7 of the included studies, and these studies adopted pathological examination, fine needle aspiration, and ultrasound as reference standards for diagnosing benign nodules [19, 22, 25, 27, 28, 30, 31].

Diagnostic Performance of Classic Machine Learning-Based CAD Systems

There were 6 studies which investigated the performance of classic machine learning-based CAD systems [14–19]. The CAD systems in 4 studies were developed according to similar parameters, such as shape, margin, composition, echogenicity, internal composition, microcalcification, and peripheral halo [15, 17–19]. The sensitivity and specificity of the CAD systems ranged from

Xu/Gao/Wang/Yin/Yu/Bai/Pei/Chen/
Yang/Wang/Wan

Table 1. Study characteristics

Study [Ref.]	Year	Country	Design	Nodule size, mm	Real-time diagnosis	Deep learning	Training set size	Validation set size	
								M	B
Zhu et al. [18]	2013	China	R	4–52	No	No	464	148	77
Song et al. [14]	2015	China	R	NR	No	No	155	21	20
Wu et al. [15]	2016	China	R	NR	No	No	485	260	225
Yu et al. [16]	2017	China	P	>2	No	No	610	17	33
Thomas et al. [19][a]	2017	USA	R	NR	No	No	410	11	61
Zhang et al. [17]	2019	China	R	≤25	No	No	1,238	118	708
Choi et al. [25]	2017	South Korea	P	≥5	Yes	Yes	–	43	59
Jeong et al. [27]	2019	South Korea	P	≥10	Yes	Yes	–	44	56
Yoo et al. [28]	2018	South Korea	P	≥5	Yes	Yes	–	50	67
Gitto et al. [26]	2019	Italy	R	18±7	Yes	Yes	–	14	48
Kim et al. [31]	2019	South Korea	R	1.2±0.9 (B) 1.2±0.9 (M)	Yes	Yes	–	86	132
Gao et al. [20]	2018	China	R	17±14 (B) 10±7 (M)	No	Yes	3,700	239	103
Song et al. [23]	2019	China	R	NR	No	Yes	6,228	180	187
Li et al. [7]	2019	China	R	NR	No	Yes	312,399	118	156
Ko et al. [21]	2019	China	R	10–20	No	Yes	594	100	50
Song et al. [22]	2019	South Korea	P	NR	No	Yes	1,358	50	50
Wang et al. [24]	2019	China	R	NR	No	Yes	5,007	242	109
Luo et al. [29][a]	2018	China	NR	NR	No	Yes	NR	292	208
Guan et al. [30]	2019	China	R	NR	No	Yes	2,437	209	190

B, benign; M, malignant; NR, not reported; P, prospective; R, retrospective. [a] Conference abstract.

0.82 to 0.92 and from 0.65 to 0.97, respectively. The pooled sensitivity, specificity, AUC, and DOR are demonstrated in Figure 2a.

Diagnostic Performance of Deep Learning-Based CAD Systems

Thirteen studies with 1,667 malignant and 1,415 benign nodules were included in the analysis [7, 20–31]. The pooled sensitivity, specificity, AUC, and DOR are demonstrated in Figure 2b. Eleven of the 13 studies compared the diagnostic performances of CAD systems and radiologists [7, 20, 21, 23–28]. The pooled sensitivity, specificity, AUC, and DOR were comparable between the CAD systems and the radiologists (Fig. 2c, d).

Diagnostic Performance of the CAD System and Real-Time Diagnosis of Radiologists

Five studies with 237 malignant and 362 benign thyroid nodules were included in the analysis [25–28, 31]. All 5 studies compared the diagnostic performances of CAD systems and radiologists. The pooled sensitivity, specificity, AUC, and DOR were comparable between the CAD systems and the radiologists (Fig. 2e, f). However, in individual studies, the radiologists outperformed the CAD

system either in sensitivity (0.79 vs. 0.21; $p = 0.008$) [26] or specificity (0.75 vs. 0.95, $p = 0.002$; 0.96 vs. 0.84, $p = 0.016$; and 0.96 vs. 0.83, $p < 0.001$) [25, 27, 31] or in the positive predictive value (0.93 vs. 0.83, $p = 0.076$) [28].

Discussion and Conclusion

The present study reviewed the current research on the performance of CAD systems in differentiating malignant and benign thyroid nodules, and the results suggest that CAD systems, both classic machine learning- and deep learning-based systems, demonstrate comparable diagnostic accuracy to that of radiologists with 5–20 years of experience in thyroid ultrasound scanning. Nonetheless, experienced radiologists may retain a diagnostic advantage over CAD systems in real-time diagnosis.

The good performance of classic machine learning-based CAD systems may benefit from the automatic and mandatory standardized diagnostic process they follow. The strategies for nodule character classification were based on several classic parameters, such as shape, margin, composition, echogenicity, internal composition, microcalcification, and peripheral halo [15, 17, 18]. These char-

(For legend see next page.)

Xu/Gao/Wang/Yin/Yu/Bai/Pei/Chen/
Yang/Wang/Wan
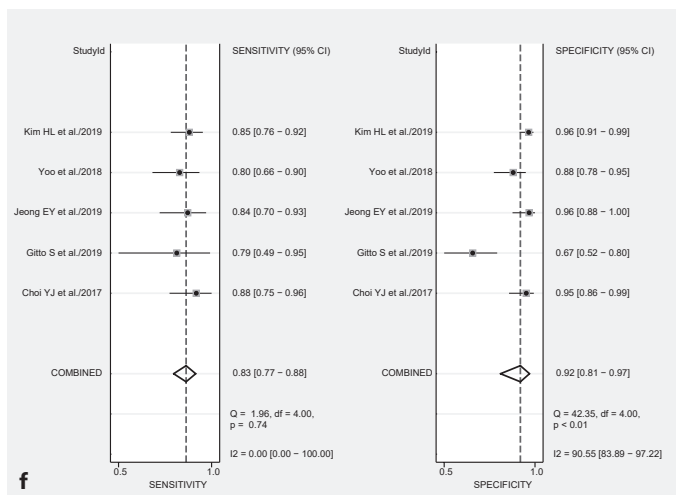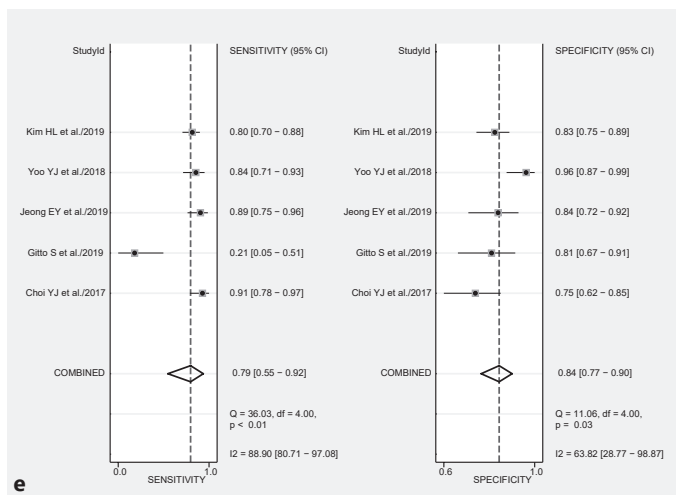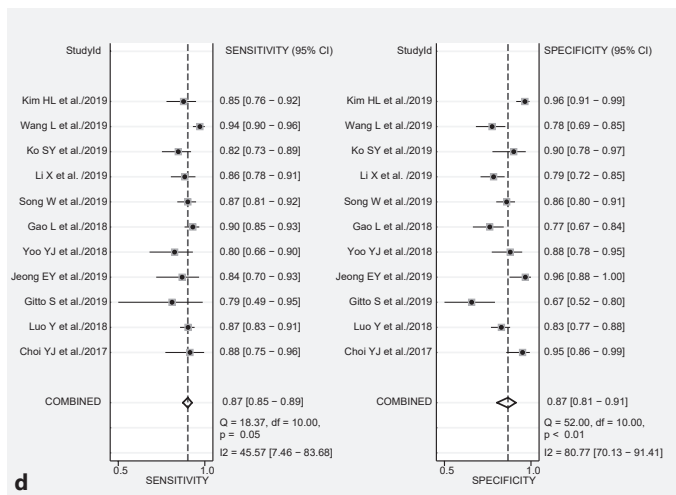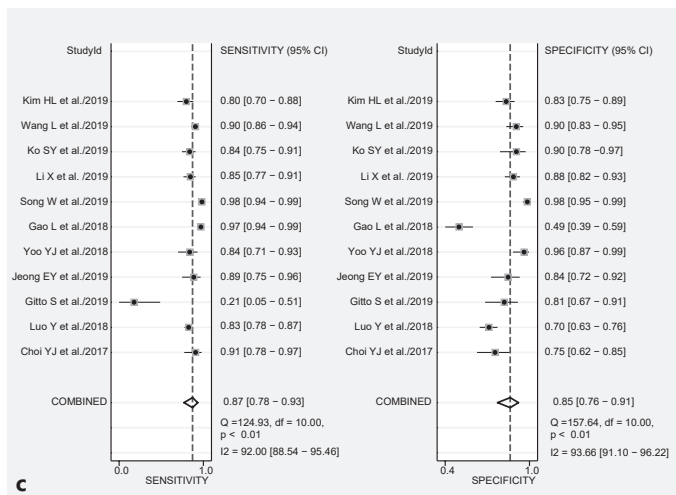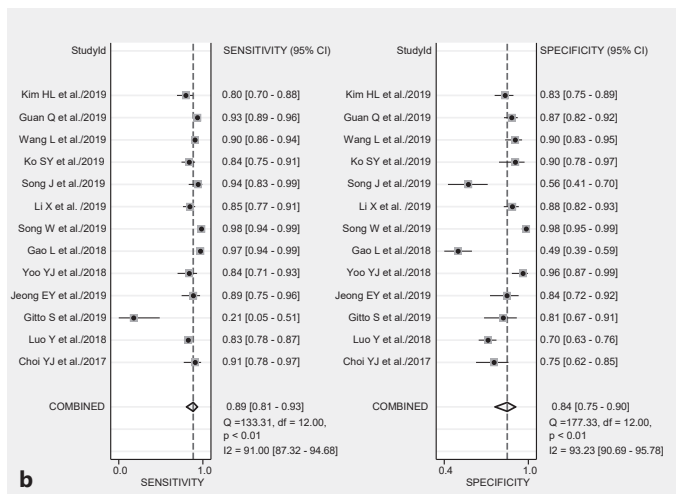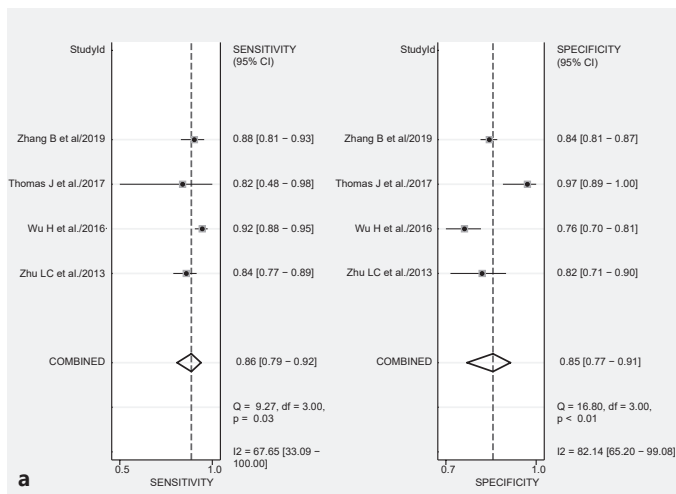
acteristics are very similar to features that are proposed by the thyroid imaging reporting guidelines [3, 32, 33], and the diagnostic accuracy may be improved by systematically perceiving and interpreting all the features. The standardized process will benefit inexperienced or nonspecialist radiologists in improving their diagnostic accuracy [15]. However, classic machine learning-based CAD systems merely simulate the presentational diagnostic strategy of radiologists, and it is difficult to transcend the limits of their human teachers, the experienced radiologists.

Compared with classic machine learning, deep learning may further improve the diagnostic performance of CAD systems by further decreasing subjectivity during the diagnostic process. The deep learning technique can automatically extract multilevel features that are not limited by the engineered features used by radiologists [7]. However, in actual fact, sensitivity, specificity, and accuracy values comparable to those of radiologists were achieved by deep learning-based CAD systems, and no significant superiority in accuracy over radiologists was demonstrated. This negative result may be related to the small sample sizes of the training sets used in most of the included studies, in which the number of images ranged from 594 to 6,228 for the training set (Table 1). Generally, hundreds of thousands of well-selected images are required to develop stable, high-performance systems. In the only study with a larger training set, including 312,399 images [7], the deep learning-based CAD system outperformed most of the experienced radiologists. However, the ex post test and the rigid cutoff levels for diagnostic interpretation may underestimate the diagnostic performance of radiologists. First, the performance of the radiologists was likely to have been limited by the static images provided during the ex post tests. The overall characteristics of one nodule may not be well reflected by merely one or several images. During real-time clinical diagnosis, ultrasound image segments are dynamically observed, and other characteristics beyond nodule images, such as cervical lymph nodes, age, and medical history, were also considered. Second, the rigid cutoff levels that were adopted to determine the diagnostic conclusion of the radiologists may also have influenced the per-formance of the radiologists. For instance, points 4a, 4b, and 5 of the TI-RADS criteria were adopted by researchers to determine the conclusion of radiologists during the diagnostic process [7, 21, 24]. It is probable that a different conclusion would have been drawn if the cutoff level had been adjusted.

Radiologists may regain their competitiveness in real-time clinical diagnosis, as suggested by the 5 studies comparing the diagnostic performance of a CAD system and real-time diagnosis by radiologists without fixed cutoff levels [25–28, 31]. During real-time diagnosis, radiologists demonstrated superior sensitivity [26] or specificity [25–27, 31], and no inferior diagnostic performance by any evaluation index was demonstrated in any individual study. The pooled result also demonstrated a potentially higher pooled sensitivity (0.83 vs. 0.79), specificity (0.84 vs. 0.92), and DOR (55.93 [95% CI 17.72–176.54] vs. 19.82 [95% CI 5.92–66.35]) of the experienced radiologists compared with the CAD system (Fig. 2e, f).

There are some limitations of the present study. First, various artificial intelligence models were combined in the meta-analysis, and this may have introduced statistical heterogeneity. To decrease this kind of heterogeneity, classic machine learning- and deep learning-based CAD systems were analyzed separately. Furthermore, a subgroup analysis of studies applying the same CAD system (the S-Detect system) was also performed. Second, in 3 of the included studies [25, 28, 30], ultrasonic diagnosis was adopted as the reference standard for benign nodules; this might possibly be related to the increased diagnostic accuracy of the radiologists. However, the benign nodule diagnosis was considered only if ultrasonic findings were of very low suspicion, and the risk of malignancy of such nodules is exceedingly low [11].

In conclusion, our results suggest that CAD systems may provide an accuracy comparable to that of radiologists with 5–20 years of experience in thyroid ultrasound scanning with regard to diagnosing malignant thyroid nodules using static ultrasound images. However, most of the CAD systems are currently unavailable for real-time clinical diagnosis. Considering the variation in classifica-

Fig. 2. Forest plots of computer-aided diagnosis (CAD) systems and the radiologist counterparts. The sensitivity and specificity of the individual studies are represented by gray squares, and the pooled results are represented by rhombi. The confidence interval (CI) is indicated by error bars. **a** Diagnostic performance of classic machine learning-based CAD systems: AUC 0.93 (95% CI 0.90–0.95) and DOR 37.41 (95% CI 24.91–56.20). **b** Diagnostic performance of deep learning-based CAD systems: AUC 0.93 (95% CI 0.90–0.95) and DOR 40.87 (95% CI 18.13–92.13). **c, d** Comparison between deep learning-based CAD systems (**c**) and radiologists (**d**): AUC 0.93 (95% CI 0.90–0.95) versus 0.92 (95% CI 0.89–0.94) and DOR 40.12 (95% CI 15.58–103.33) versus 44.88 (95% CI 30.71–65.57). **e, f** Comparison between deep learning-based real-time CAD systems (**e**) and radiologists (**f**): AUC 0.88 (95% CI 0.85–0.91) versus 0.88 (95% CI 0.85–0.91) and DOR 19.82 (95% CI 5.92–66.35) versus 55.93 (95% CI 17.72–176.54).

tion algorithms, sample sizes of training sets, clinical experience of the image-interpreting staff, diagnostic criteria used by radiologists, clinical experience of the radiologists, and reference standards, the diagnostic conclusions drawn from any of the current CAD systems on thyroid nodules should be accepted with caution.

Statement of Ethics

All analyses were based on previously published studies and ethical approval, and patient consent forms were not required.

Disclosure Statement

The authors have no conflicts of interest to declare.

References

1 Davies L, Welch HG. Current thyroid cancer trends in the United States. JAMA Otolaryngol Head Neck Surg. 2014 Apr;140(4):317–22.
2 Vaccarella S, Franceschi S, Bray F, Wild CP, Plummer M, Dal Maso L. Worldwide Thyroid-Cancer Epidemic? The Increasing Impact of Overdiagnosis. N Engl J Med. 2016 Aug;375(7):614–7.
3 Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, 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 Jan;26(1):1–133.
4 Frates MC, Benson CB, Charboneau JW, Cibas ES, Clark OH, Coleman BG, et al.; Society of Radiologists in Ultrasound. Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology. 2005 Dec;237(3):794–800.
5 Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. J Clin Endocrinol Metab. 2002 May;87(5):1941–6.
6 Nam-Goong IS, Kim HY, Gong G, Lee HK, Hong SJ, Kim WB, et al. Ultrasonography-guided fine-needle aspiration of thyroid incidentaloma: correlation with pathological findings. Clin Endocrinol (Oxf). 2004 Jan;60(1):21–8.
7 Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. Lancet Oncol. 2019 Feb;20(2):193–201.
8 Ma J, Wu F, Zhu J, Xu D, Kong D. A pretrained convolutional neural network based method for thyroid nodule diagnosis. Ultrasonics. 2017 Jan;73:221–30.
9 Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. J Digit Imaging. 2017 Aug;30(4):477–86.
10 Lim KJ, Choi CS, Yoon DY, Chang SK, Kim KK, Han H, et al. Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography. Acad Radiol. 2008 Jul;15(7):853–8.
11 Durante C, Costante G, Lucisano G, Bruno R, Meringolo D, Paciaroni A, et al. The natural history of benign thyroid nodules. JAMA. 2015 Mar;313(9):926–35.
12 Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. Ann Intern Med. 2009 Aug;151(4):264–9, W64.
13 Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al.; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct;155(8):529–36.
14 Song G, Xue F, Zhang C. A Model Using Texture Features to Differentiate the Nature of Thyroid Nodules on Sonography. J Ultrasound Med. 2015 Oct;34(10):1753–60.
15 Wu H, Deng Z, Zhang B, Liu Q, Chen J. Classifier Model Based on Machine Learning Algorithms: Application to Differential Diagnosis of Suspicious Thyroid Nodules via Sonography. AJR Am J Roentgenol. 2016 Oct;207(4):859–64.
16 Yu Q, Jiang T, Zhou A, Zhang L, Zhang C, Xu P. Computer-aided diagnosis of malignant or benign thyroid nodes based on ultrasound images. Eur Arch Otorhinolaryngol. 2017 Jul;274(7):2891–7.
17 Zhang B, Tian J, Pei S, Chen Y, He X, Dong Y, et al. Machine Learning-Assisted System for Thyroid Nodule Diagnosis. Thyroid. 2019 Jun;29(6):858–67.
18 Zhu LC, Ye YL, Luo WH, Su M, Wei HP, Zhang XB, et al. A model to discriminate malignant from benign thyroid nodules using artificial neural network. PLoS One. 2013 Dec;8(12):e82211.
19 Thomas J, Hupart KH, Radhamma RK, Ledger GA, Singh G. Thyroid ultrasound malignancy score (TUMS): a machine learning model to predict thyroid malignancy from ultrasound features. Thyroid. 2017;27:A187.
20 Gao L, Liu R, Jiang Y, Song W, Wang Y, Liu J, et al. Computer-aided system for diagnosing thyroid nodules on ultrasound: a comparison with radiologist-based clinical assessments. Head Neck. 2018 Apr;40(4):778–83.
21 Ko SY, Lee JH, Yoon JH, Na H, Hong E, Han K, et al. Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck. 2019 Apr;41(4):885–91.
22 Song J, Chai YJ, Masuoka H, Park SW, Kim SJ, Choi JY, et al. Ultrasound image analysis using deep learning algorithm for the diagnosis of thyroid nodules. Medicine (Baltimore). 2019 Apr;98(15):e15133.

23 Song W, Li S, Liu J, Qin H, Zhang B, Shuyang Z, et al. Multitask Cascade Convolution Neural Networks for Automatic Thyroid Nodule Detection and Recognition. IEEE J Biomed Health Inform. 2019 May;23(3): 1215–24.

24 Wang L, Yang S, Yang S, Zhao C, Tian G, Gao Y, et al. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. World J Surg Oncol. 2019 Jan;17(1):12.

25 Choi YJ, Baek JH, Park HS, Shim WH, Kim TY, Shong YK, 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 Apr;27(4): 546–52.

26 Gitto S, Grassi G, De Angelis C, Monaco CG, Sdao S, Sardanelli F, et al. A computer-aided diagnosis system for the assessment and characterization of low-to-high suspicion thyroid nodules on ultrasound. Radiol Med (Torino). 2019 Feb;124(2):118–25.

27 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 Apr;29(4):1978–85.

28 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 Jul-Aug; 19(4):665–72.

29 Luo Y, Xie F. Artificial intelligence-assisted ultrasound diagnosis for thyroid nodules. Thyroid. 2018;28:A1.

30 Guan Q, Wang Y, Du J, Qin Y, Lu H, Xiang J, et al. Deep learning based classification of ultrasound images for thyroid nodules: a large scale of pilot study. Ann Transl Med. 2019 Apr;7(7):137.

31 Kim HL, Ha EJ, Han M. Real-World Performance of Computer-Aided Diagnosis System for Thyroid Nodules Using Ultrasonography. Ultrasound Med Biol. 2019 Oct;45(10):2672–8.

32 Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. Radiology. 2011 Sep;260(3):892–9.

33 Kim EK, Park CS, Chung WY, Oh KK, Kim DI, Lee JT, et al. New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. AJR Am J Roentgenol. 2002 Mar;178(3):687–91.