



Computer-Aided Diagnosis System for the Evaluation of Thyroid Nodules on Ultrasonography: Prospective Non-Inferiority Study according to the Experience Level of Radiologists

Sae Rom Chung, MD¹, Jung Hwan Baek, MD, PhD¹, Min Kyoung Lee, MD¹, Yura Ahn, MD¹, Young Jun Choi, MD, PhD¹, Tae-Yon Sung, MD, PhD², Dong Eun Song, MD, PhD³, Tae Yong Kim, MD, PhD⁴, Jeong Hyun Lee, MD, PhD¹

Departments of ¹Radiology and Research Institute of Radiology, ²Surgery, ³Pathology, and ⁴Endocrinology and Metabolism, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

Objective: To determine whether a computer-aided diagnosis (CAD) system for the evaluation of thyroid nodules is non-inferior to radiologists with different levels of experience.

Materials and Methods: Patients with thyroid nodules with a decisive diagnosis of benign or malignant nodule were consecutively enrolled from November 2017 to September 2018. Three radiologists with different levels of experience (1 month, 4 years, and 7 years) in thyroid ultrasound (US) reviewed the thyroid US with and without using the CAD system. Statistical analyses included non-inferiority testing of the diagnostic accuracy for malignant thyroid nodules between the CAD system and the three radiologists with a non-inferiority margin of 10%, comparison of the diagnostic performance, and the added value of the CAD system to the radiologists.

Results: Altogether, 197 patients were included in the study cohort. The diagnostic accuracy of the CAD system (88.5%, 95% confidence interval [CI] = 82.7–92.5) was non-inferior to that of the radiologists with less experience (1 month and 4 year) of thyroid US (83.0%, 95% CI = 76.5–88.0; $p < 0.001$), whereas it was inferior to that of the experienced radiologist (7 years) (95.8%, 95% CI = 91.4–98.0; $p = 0.138$). The sensitivity and negative predictive value of the CAD system were significantly higher than those of the less-experienced radiologists were, whereas no significant difference was found with those of the experienced radiologist. A combination of US and the CAD system significantly improved sensitivity and negative predictive value, although the specificity and positive predictive value deteriorated for the less-experienced radiologists.

Conclusion: The CAD system may offer support for decision-making in the diagnosis of malignant thyroid nodules for operators who have less experience with thyroid US.

Keywords: Computer-aided diagnosis; Thyroid nodule; Thyroid cancer; Ultrasonography

INTRODUCTION

Thyroid nodules are common, with a prevalence of 19–68% in the adult population based on detection

using ultrasound (US) (1-3). Although the vast majority of these incidentally detected nodules will ultimately prove to be benign, approximately 5–15% of patients with either solitary or multiple nodules will be diagnosed

Received August 7, 2019; accepted after revision November 4, 2019.

This research was supported by a research fund and equipment from SAMSUNG MEDISON Co., Ltd.

Corresponding author: Jung Hwan Baek, MD, PhD, Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.

• Tel: (822) 3010-4348 • Fax: (822) 476-0090 • E-mail: radbaek@naver.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

with thyroid cancer (4-6). US is the main diagnostic modality for evaluating thyroid nodules and differentiating between benign and malignant nodules. However, the main limitations of US are its operator dependence and interobserver variability, which is moderate to substantial (7-10).

A computer-aided diagnosis (CAD) system was recently introduced for the characterization and interpretation of the US features of thyroid nodules (11-19). Several studies have found that this CAD system affords a diagnostic performance similar to that of an experienced radiologist and that it can offer support for decision-making in thyroid cancer diagnosis (12, 13, 16, 18, 19). However, the study populations of all of these studies included substantial proportions of malignant thyroid nodules (42.2–69.9%), rates that are much higher than the general prevalence of thyroid cancer (12, 13, 16, 18, 19). Furthermore, these studies were performed by experienced radiologists and there are no studies comparing diagnostic performance between the CAD system and radiologists with less experience. The usefulness of the CAD system for US may differ according to the level of experience of the operator. The purpose of this study was therefore to determine whether the CAD system for evaluation of thyroid nodules is non-inferior to radiologists with different experience levels.

MATERIALS AND METHODS

Patients

This prospective study was approved by our Institutional Review Board and written informed consent was obtained from all patients before they underwent US. Patients who visited the thyroid clinic of the radiology department of Asan Medical Center for the evaluation of thyroid nodules were recruited between November 2017 and September 2018. The inclusion criteria used to select patients were 1) underwent US-guided core needle biopsy (CNB) or fine needle aspiration (FNA) or 2) follow-up for a thyroid nodule with decisive diagnosis. Patients were excluded from the study population if they 1) had a thyroid nodule less than 1 cm or 2) were younger than 18 years of age.

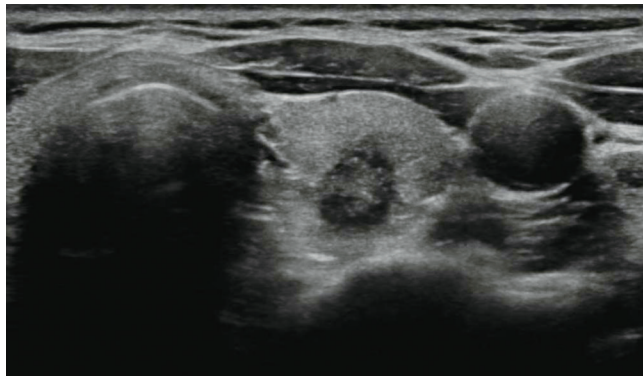
A decisive diagnosis consisted of a malignant or benign diagnosis. A malignant diagnosis was made when malignancy was confirmed based on a surgical specimen or using CNB histology or FNA cytology. A diagnosis of a benign nodule was made when any one of the following criteria was met: 1) confirmation using a surgical specimen; 2) benign CNB

histology or FNA cytology findings; or 3) US findings of benign nodule (simple cyst, predominantly cystic nodule with reverberating artifact, or nodule with a spongiform appearance) (8) with no change over at least 1 year.

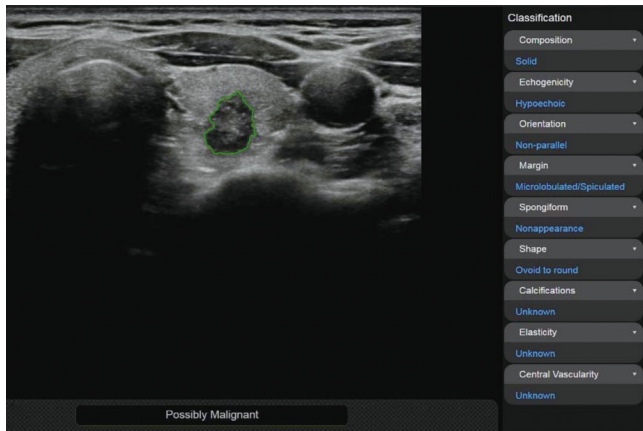
US Image Acquisition and Analysis

US examinations were performed using an RS80A US system (Samsung Medison Co., Ltd., Seoul, Korea) equipped with L3-12A (linear high-frequency probe; frequency range, 3–12 MHz). Real-time CAD system software (S-Detect™ for Thyroid, Samsung Medison Co., Ltd.) was integrated into the US system. This real-time CAD software provides two points to indicate the top left and bottom right of a region of interest (ROI) box enclosing a thyroid nodule on the US system. Based on the given box, the software automatically calculates the contour of the mass to distinguish it from normal thyroid tissue (segmentation) and evaluates the US features of the mass, including size (maximum diameter in captured image), composition (solid, partially cystic, or cystic), shape (oval-to-round or irregular), orientation (parallel or non-parallel), margins (well-defined, ill-defined, or spiculated), echogenicity (hyperechoic/isoechoic or hypoechoic or marked hypoechoic), and spongiform nature. These US features are quantified into computerized values and are presented as features to describe the thyroid nodule. Consequently, the software displays a diagnosis as to whether the nodule is possibly benign or possibly malignant (Figs. 1, 2). One radiologist with 7 years of experience in performing thyroid US drew the ROI box enclosing a target thyroid nodule on transverse and longitudinal images and then evaluated the quality of the nodule segmentation. If the segmentation did not properly define the contours of the nodule, manual correction was made.

All images were reviewed using a local picture archiving and communication system (PACS) monitor and digital imaging and communications in medicine image-viewing software (PetaVision, Asan Medical Center, Seoul, Korea). The US images were independently analyzed by three radiologists with different levels of thyroid US experience: resident (1 month, 200 cases), fellow (4 years, 1000 cases), and staff (7 years, 80000 cases). Resident and fellow radiologists were categorized as less-experienced radiologists and the staff radiologist categorized as an experienced radiologist. None of the reviewers had any information regarding the patients' clinical histories, previous imaging results, or previous biopsy results.



A

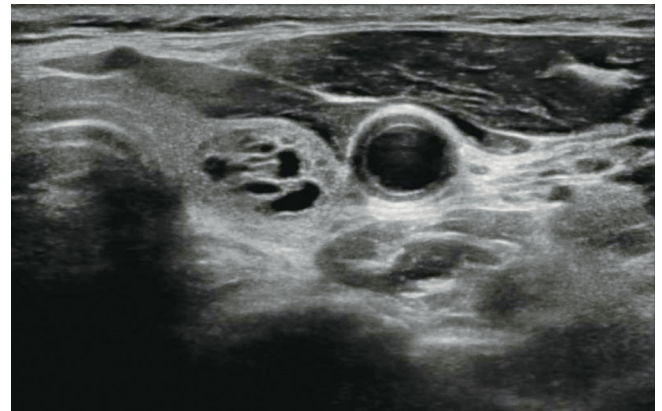


B

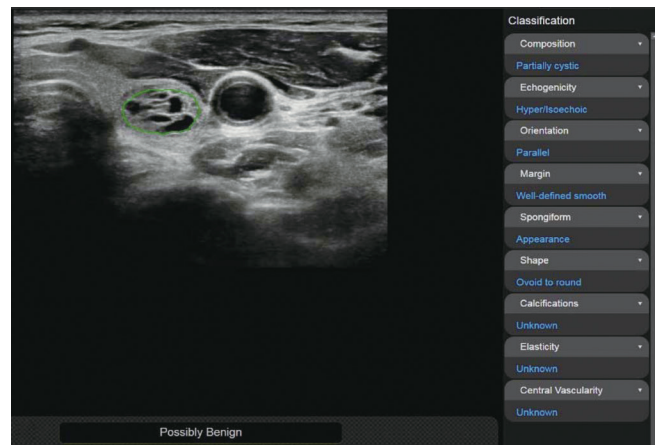
Fig. 1. Representative case of malignant thyroid nodule.

US image (A) and automatically calculated mass contour, US features, and diagnosis presented by CAD system (B). Both CAD system and radiologist diagnosed it as malignant nodule. CAD system performed excellent segmentation of thyroid nodule. CAD system and radiologist demonstrated concordance regarding US characteristics of composition (solid), shape (ovoid-to-round), orientation (non-parallel), echogenicity (hypoechoic), and margin (spiculated). CAD = computer-aided diagnosis, US = ultrasound

Two sets of US images for each patient were arranged in the folders of our local PACS; one set included gray scale images only and the other set included gray scale images and CAD system images together. The gray scale images included in the analysis were a transverse and longitudinal image of the target nodule used as a reference in the CAD system. Two separate US image analysis sessions were performed. First, each radiologist reviewed the grayscale images only and evaluated the following features: composition (solid, partially cystic, or cystic), echogenicity (hyperechoic/isoechoic or hypoechoic or marked hypoechoic), shape (ovoid-to-round or irregular), orientation (parallel or non-parallel), margin (smooth, spiculated/microlobulated, or ill-defined), and calcification (none, microcalcification, macrocalcification, or rim



A



B

Fig. 2. Representative case of benign thyroid nodule.

US image (A) and automatically calculated mass contour, US features, and diagnosis presented by CAD system (B). Both CAD system and radiologist diagnosed it as benign nodule. CAD system performed satisfactory segmentation of thyroid nodule. CAD system and radiologist demonstrated concordance regarding US characteristics of composition (partially cystic), shape (ovoid-to-round), orientation (parallel), echogenicity (hyperechoic/isoechoic), and margin (well-defined).

calcification). Each radiologist concluded their diagnosis as to whether the nodule was benign or malignant according to the previous report by Moon et al. (8). Each radiologist then re-evaluated the same grayscale images while referring to the CAD system and made a subjective diagnostic decision based on the grayscale US and CAD. Additionally, the conjunctive combination was also analyzed, with a finding of "malignant" on either the grayscale US or CAD system being defined as malignant.

Study End Point

The primary end point of this study was the diagnostic accuracy of the CAD system for diagnosis of malignant thyroid nodules in comparison with radiologists with

different levels of experience. The secondary end points included the diagnostic performance and added value of the CAD system for diagnosis of malignant thyroid nodules, and the added value of the CAD system for interobserver agreement of US features between the three radiologists.

Statistics

This study was primarily designed as a non-inferiority study and the sample size was estimated to determine the non-inferiority of the CAD to the radiologists regarding the primary end point (diagnostic accuracy). Non-inferiority was defined as a diagnostic accuracy that was no more than 10 percentage points below the estimated diagnostic accuracy of the radiologist. The diagnostic accuracy of thyroid US for the assessment of thyroid nodules is approximately 78%. To obtain a statistical power of 80% with a one-sided P-level of 0.05, a sample size of 157 patients in each group was required (20). Therefore, to allow for study dropouts, 200 patients were enrolled in this study.

The diagnostic performance of the CAD system and each radiologist was evaluated by calculating the sensitivities, specificities, positive predictive values (PPVs), and negative predictive values (NPVs), and then comparing those using generalized estimating equations for matched data.

To assess the added value of the CAD system, the sensitivity, specificity, PPV, and NPV were compared between the results from the grayscale images and those from the grayscale images with CAD. Generalized estimating equations were used for the two comparisons to account for clustering from the same patient.

Finally, the extent of interobserver agreement (the multiple kappa value) between the three radiologists in terms of the descriptions of the US characteristics was determined for the nodule evaluations using only the grayscale images and those made using both the grayscale images and CAD system. The level of agreement for Cohen's kappa was defined as follows: < 0.20, poor agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and > 0.80, good agreement. Statistical analyses were performed using SAS software version 9.4 (SAS Institute Inc., Cary, NC, USA).

RESULTS

Of the 200 patients initially enrolled in the study, 197 were included in the final cohort for analysis (Table 1). Three patients were excluded because the size of their

Table 1. Characteristics of Study Subjects

| Parameter | Value |
|--------------------------------------|--------------|
| Mean age* (year) | 51.4 (20–94) |
| Sex, n (%) | |
| Women | 163 (82.7) |
| Men | 34 (17.3) |
| Sonographic features | |
| Composition | |
| Solid | 68 |
| Partially cystic | 105 |
| Cystic | 24 |
| Echogenicity | |
| Iso/hyperechogenicity | 140 |
| Hypoechoogenicity | 51 |
| Marked hypoechoogenicity | 6 |
| Shape | |
| Ovoid-to-round | 195 |
| Irregular | 2 |
| Orientation | |
| Parallel | 179 |
| Non-parallel | 18 |
| Margin | |
| Smooth | 149 |
| Spiculated/microlobulated | 21 |
| Ill-defined | 27 |
| Calcification | |
| None | 13 |
| Microcalcification | 34 |
| Macrocalcification/rim calcification | 26 |
| FNA/CNB result | |
| Category 1 | 2 |
| Category 2 | 30 |
| Category 3 | 133 |
| Category 4 | 6 |
| Category 5 | 3 |
| Category 6 | 21 |
| Final diagnosis, n (%) | 165 |
| Benign | 140 (84.8) |
| Malignant | 25 (15.2) |
| Diagnostic method, n (%) | |
| Benign | |
| Surgery | 1 (0.7) |
| CNB or FNA | 135 (96.4) |
| US finding | 4 (2.9) |
| Malignant | |
| Surgery | 19 (76) |
| CNB or FNA | 7 (24) |

*Data are expressed as median with range in parenthesis. CNB = core needle biopsy, FNA = fine needle aspiration, US = ultrasound

target nodule was smaller than 1 cm. Final diagnosis of the thyroid nodule was confirmed in 165 patients (benign: $n = 140$, 84.8%; malignant: $n = 25$, 15.2%). The pathological subtypes of the malignant nodules were papillary thyroid carcinoma ($n = 24$) and follicular thyroid carcinoma ($n = 1$).

Diagnostic Performance and Added Value of the CAD System

Diagnostic performance was assessed using the 165 patients with thyroid nodules with a final diagnosis. The diagnostic accuracy was 88.5% (95% confidence interval [CI], 82.7–92.5) for the CAD system and 83.0% (95% CI, 76.5–88.0) for the resident and fellow radiologists, which leads us to conclude that, with a 10% non-inferiority margin, the CAD system shows non-inferiority compared with the resident and fellow radiologists ($p = 0.001$). However, the diagnostic accuracy of the CAD was inferior to that of the staff radiologist (95.8%, 95% CI = 91.4–98.0; $p = 0.138$) (Table 2).

The sensitivity and NPV of the CAD system were significantly higher than those of the resident and fellow radiologists were, but were not significantly different to those of the staff radiologist. The diagnostic accuracy, specificity, and PPV of the CAD system were significantly lower than those of the staff radiologist were, but were not significantly different to those of the resident and fellow radiologists (Table 3).

When the radiologists subjectively combined the grayscale images and CAD, the conclusion was changed in six cases reviewed by the resident, three cases reviewed by the

fellow, and one case reviewed by the staff. The diagnostic accuracy, sensitivity, PPV, and NPV slightly improved for all three radiologists, although the differences did not reach statistical significance (Table 4). The specificity was also slightly improved for the resident but showed no change for the fellow and staff radiologists. For the conjunctive combination analysis of the CAD system and grayscale

Table 4. Added Value of CAD System for Diagnosis of Malignant Thyroid Nodules according to Experience Level of Radiologists

| Diagnostic Measures (%) | Without CAD | With CAD | <i>P</i> |
|-------------------------|---------------------|---------------------|----------|
| Accuracy | | | |
| Resident | 83.03 (76.52–88.02) | 84.24 (77.86–89.04) | 0.414 |
| Fellow | 83.03 (76.52–88.02) | 84.85 (78.54–89.55) | 0.083 |
| Staff | 95.76 (91.37–97.96) | 96.36 (92.14–98.36) | 0.317 |
| Sensitivity | | | |
| Resident | 64 (44.00–80.09) | 68 (47.84–83.12) | 0.317 |
| Fellow | 72 (51.78–86.03) | 72 (51.78–86.03) | N/A |
| Staff | 84 (64.31–93.86) | 84 (64.31–93.86) | N/A |
| Specificity | | | |
| Resident | 86.43 (79.70–91.17) | 87.14 (80.51–91.75) | 0.655 |
| Fellow | 85.00 (78.09–90.01) | 87.14 (80.51–91.75) | 0.083 |
| Staff | 97.86 (93.57–99.31) | 98.57 (94.47–99.64) | 0.317 |
| PPV | | | |
| Resident | 45.71 (30.22–62.09) | 48.57 (32.74–64.70) | 0.396 |
| Fellow | 46.15 (31.35–61.67) | 50.00 (34.22–65.78) | 0.084 |
| Staff | 87.50 (67.62–95.91) | 91.30 (71.12–97.82) | 0.308 |
| NPV | | | |
| Resident | 93.08 (87.23–96.36) | 93.85 (88.18–96.89) | 0.290 |
| Fellow | 94.44 (88.80–97.33) | 94.57 (89.05–97.39) | 0.141 |
| Staff | 97.16 (92.69–98.93) | 97.18 (92.74–98.94) | 0.368 |

NA = not applicable

Table 2. Non-Inferiority Test for Diagnostic Accuracy between CAD and Radiologists

| Diagnostic Measures (%) | CAD | Resident | Fellow | Staff | <i>P</i> * | <i>p</i> [†] | <i>p</i> [‡] |
|-------------------------|------------------|--------------------|--------------------|--------------------|------------|-----------------------|-----------------------|
| Accuracy | 88.5 (82.7–92.5) | 83.0 (76.57–87.99) | 83.0 (76.57–87.99) | 95.8 (91.50–97.93) | 0.001 | 0.001 | 0.138 |

**p* value of CAD system versus resident radiologist, [†]*p* value of CAD system versus fellow radiologist, [‡]*p* value of CAD system versus staff radiologist. CAD = computer-aided diagnosis

Table 3. Diagnostic Performance of CAD System and Three Radiologists with Different Levels of Experience

| Diagnostic Measures (%) | CAD | Resident | Fellow | Staff | <i>P</i> * | <i>p</i> [†] | <i>p</i> [‡] |
|-------------------------|---------------------|---------------------|---------------------|---------------------|------------|-----------------------|-----------------------|
| Diagnostic accuracy | 88.48 (82.65–92.53) | 83.03 (76.52–88.02) | 83.03 (76.52–88.02) | 95.76 (91.37–97.96) | 0.139 | 0.083 | 0.005 |
| Sensitivity | 92 (73.06–97.99) | 64 (44.00–80.09) | 72 (51.78–86.03) | 84 (64.31–93.86) | 0.008 | 0.025 | 0.157 |
| Specificity | 87.9 (81.33–92.32) | 86.4 (79.70–91.17) | 85.0 (78.09–90.01) | 97.9 (93.57–99.31) | 0.715 | 0.394 | < 0.001 |
| PPV | 57.50 (41.96–71.69) | 45.71 (30.22–62.09) | 46.15 (31.35–61.67) | 87.50 (67.62–95.91) | 0.158 | 0.092 | < 0.001 |
| NPV | 98.40 (93.83–99.60) | 93.08 (87.23–96.36) | 94.44 (88.80–97.33) | 97.16 (92.69–98.93) | 0.007 | 0.021 | 0.212 |

**p* value of CAD system versus resident radiologist, [†]*p* value of CAD system versus fellow radiologist, [‡]*p* value of CAD system versus staff radiologist. NPV = negative predictive value, PPV = positive predictive value

Table 5. Added Value of CAD System for Diagnosis of Malignant Thyroid Nodules according to Experience Level of Radiologists with Conjunctive Combination Analysis

| Diagnostic Measures (%) | Without CAD | With CAD | P |
|-------------------------|---------------------|---------------------|---------|
| Accuracy | | | |
| Resident | 83.03 (76.52–88.02) | 80.00 (73.20–85.42) | 0.317 |
| Fellow | 83.03 (76.52–88.02) | 82.42 (75.85–87.50) | 0.808 |
| Staff | 95.76 (91.37–97.96) | 88.48 (82.65–92.53) | 0.005 |
| Sensitivity | | | |
| Resident | 64 (44.00–80.09) | 92 (73.06–97.99) | 0.008 |
| Fellow | 72 (51.78–86.03) | 92 (73.06–97.99) | 0.025 |
| Staff | 84 (64.31–93.86) | 92 (73.06–97.99) | 0.157 |
| Specificity | | | |
| Resident | 86.43 (79.70–91.17) | 77.86 (70.23–83.97) | 0.005 |
| Fellow | 85.00 (78.09–90.01) | 80.71 (73.34–86.43) | 0.083 |
| Staff | 97.86 (93.57–99.31) | 87.86 (81.33–92.32) | < 0.001 |
| PPV | | | |
| Resident | 45.71 (30.22–62.09) | 42.59 (30.20–55.99) | 0.871 |
| Fellow | 46.15 (31.35–61.67) | 46.00 (32.82–59.77) | 0.973 |
| Staff | 87.50 (67.62–95.91) | 57.50 (41.96–71.69) | < 0.001 |
| NPV | | | |
| Resident | 93.08 (87.23–96.36) | 98.20 (93.08–99.55) | 0.009 |
| Fellow | 94.44 (88.80–97.33) | 98.26 (93.32–99.56) | 0.026 |
| Staff | 97.16 (92.69–98.93) | 98.40 (93.83–99.60) | 0.212 |

images, the sensitivity, and NPV of the resident and fellow were significantly improved, while the specificity and PPV decreased (Table 5). The decreased specificity and PPV related to the increased overall number of positive cases, which increased the number of false positive cases. However, for the staff radiologist, the conjunctive combination resulted in significantly decreased specificity, PPV, and diagnostic accuracy, without a significant improvement in sensitivity or NPV.

Interobserver Variability in the Recording of US Characteristics and the Effect of the CAD System

Table 6 shows a summary of the interobserver variability in the US characteristics defined by the three radiologists before and after application of the CAD system. With the exception of shape ($\kappa = 0.034$), moderate agreement was seen for all characteristics ($\kappa = 0.473$ – 0.634) and they showed no significant difference after the CAD system was applied. The kappa value for shape was very low because of the possibility of the agreement occurring by chance. The proportions of ovoid shapes were 100%, 94.9%, and 99% for the resident, fellow, and staff radiologists, respectively.

Table 6. Comparison of Kappa Values between Three Radiologists before and after Application of CAD System

| US Characteristics | Kappa Value | |
|--------------------|----------------|-----------------------------|
| | 3 Radiologists | CAD-Assisted 3 Radiologists |
| Composition | 0.603 | 0.627 |
| Echogenicity | 0.596 | 0.601 |
| Shape | 0.034 | 0.034 |
| Orientation | 0.533 | 0.563 |
| Margin | 0.473 | 0.479 |
| Calcification | 0.634 | 0.638 |

DISCUSSION

This study demonstrates the diagnostic performance of a thyroid CAD system in comparison with radiologists of various levels of experience. In terms of the primary outcome, the diagnostic accuracy of the CAD system was non-inferior to that of the resident and fellow radiologists with 1 month and 4 years of experience in thyroid US, respectively, whereas the CAD system was not demonstrated to be non-inferior to the staff radiologist. In terms of the secondary outcomes, the sensitivity and NPV of the CAD system were significantly higher than those of the resident and fellow radiologists but were not significantly different to those of the staff radiologist. The conjunctive combination of the grayscale US and CAD system significantly improved sensitivity and NPV for the resident and fellow radiologists but resulted in a reduced specificity and PPV. Therefore, we suggest that the CAD system may offer support for decision-making in thyroid cancer diagnosis for operators with less experience in thyroid US.

Recently, several studies have reported comparable diagnostic performance for the CAD system and experienced radiologists (12, 13, 16, 18, 19, 21). In the present study, we also found that the CAD system had a high sensitivity (92%) and NPV (98.4%), with these values not statistically different to those of the staff radiologist (84% and 97.2%, respectively). However, the specificity and NPV of the staff radiologist were significantly higher (97.9%) than those of the CAD system (87.9%). Overall, the diagnostic accuracy of the CAD system was inferior to that of the highly experienced radiologist. These results are consistent with previous original articles and meta-analyses (13, 21, 22). For the less-experienced radiologists, the situation was different with the sensitivity and NPV of the CAD system being significantly higher than those of the less-experienced radiologists, and there being no significant differences in

specificity or PPV. Overall, the diagnostic accuracy of the CAD system was non-inferior to the diagnostic accuracies of the less-experienced radiologists. One previous study reported that the CAD system showed lower sensitivity and higher specificity than those of an experienced radiologist, which may be due to the different level of experience of the experienced radiologist (23).

The effect of the CAD system on the radiologists' performance in diagnosing malignant thyroid nodules on US was also assessed according to the experience level of the radiologists. When the radiologists used the grayscale US and CAD system in combination subjectively, their diagnostic performance slightly increased, although the difference did not reach statistical significance. The conjunctive combination analysis showed a significant increase in the diagnostic sensitivity and NPV of the less-experienced radiologists, although the specificity decreased. The high sensitivity and NPV would be useful for ruling out diseases in clinical practice. Therefore, the less-experienced radiologists may benefit from a conjunctive combination of grayscale images and the CAD system to rule out malignant thyroid nodules and ultimately avoid unnecessary FNA.

In this study, we recruited consecutive patients who visited the outpatient clinic for thyroid nodules and this led to a realistic proportion of malignant thyroid nodules (15.2%). In previously published studies concerning the application of the CAD system to thyroid US (12, 13, 16, 18, 19), the proportion of malignant thyroid nodules ranged from 42% to 69%. This difference may be the cause of our finding of a relatively low PPV for the CAD system (57.5%) in comparison with other studies (72.2–83.3%). Considering that only 5–15% of patients with thyroid nodules will be diagnosed with thyroid cancer, the low PPV should be considered when the CAD system is used in clinical practice. In addition, considering that one of the limitations of US is the moderate to substantial interobserver variability, we analyzed whether agreement in the characterization of nodules was improved by the addition of CAD. Although the kappa score showed a slight increase in all characteristics, there was no significant difference. The low effect of the CAD system on improving interobserver variability between radiologists may be related to the poor segmentation of the nodules (13). Future technical improvements to segmentation would be of substantial benefit for nodule characterization.

This study has several limitations. First, this study was performed in a single tertiary referral center, which means

that there could be some selection bias. Large-scale multicenter studies are needed in the future to validate and generalize the findings. Second, the value of the CAD system was not evaluated for thyroid nodules with indeterminate cytological results because thyroid nodules without a decisive diagnosis were excluded. Furthermore, thyroid nodules smaller than 1 cm were excluded to enable clear CAD diagnoses. These exclusion criteria might have influenced the diagnostic performance of the CAD system. Third, the thyroid US CAD system could not evaluate the calcification of thyroid nodules. Further technical developments are needed to improve the performance of the CAD system in this respect (21). Finally, image acquisition using the CAD system was performed by an experienced radiologist. Considering the operator dependency of the CAD system (19), the diagnostic performance of the CAD system in our study may be overestimated.

In conclusion, the diagnostic accuracy of the CAD system was non-inferior to that of radiologists with less experience in thyroid US, whereas it was inferior to that of a staff radiologist. The conjunctive combination of the grayscale US and CAD system significantly improved sensitivity and NPV for the less-experienced radiologists, even though it caused a deterioration in specificity and PPV. Therefore, less-experienced radiologists may benefit from a conjunctive combination of grayscale US and the CAD system to rule out malignant thyroid nodules and ultimately avoid unnecessary FNA.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Acknowledgments

The authors thank the Division of Statistics in Medical Research Collaborating Center at Seoul Asan Medical Center for statistical analyses.

ORCID iDs

Jung Hwan Baek

<https://orcid.org/0000-0003-0480-4754>

Sae Rom Chung

<https://orcid.org/0000-0003-4219-7166>

REFERENCES

1. Guth S, Theune U, Aberle J, Galach A, Bamberger CM.

- Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest* 2009;39:699-706
2. Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med* 1997;126:226-231
 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;26:1-133
 4. Frates MC, Benson CB, Doubilet PM, Kunreuther E, Contreras M, Cibas ES, et al. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. *J Clin Endocrinol Metab* 2006;91:3411-3417
 5. 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;60:21-28
 6. 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;87:1941-1946
 7. Choi SH, Kim EK, Kwak JY, Kim MJ, Son EJ. Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. *Thyroid* 2010;20:167-172
 8. Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH, et al. Benign and malignant thyroid nodules: US differentiation—Multicenter retrospective study. *Radiology* 2008;247:762-770
 9. Park CS, Kim SH, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer variability in the sonographic evaluation of thyroid nodules. *J Clin Ultrasound* 2010;38:287-293
 10. Kim SH, Park CS, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. *Korean J Radiol* 2010;11:149-155
 11. Acharya UR, Sree SV, Krishnan MM, Molinari F, Zieleźnik W, Bardales RH, et al. Computer-aided diagnostic system for detection of Hashimoto thyroiditis on ultrasound images from a Polish population. *J Ultrasound Med* 2014;33:245-253
 12. Chang Y, Paul AK, Kim N, Baek JH, Choi YJ, Ha EJ, et al. Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: a comparison with radiologist-based assessments. *Med Phys* 2016;43:554
 13. 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;27:546-552
 14. Li LN, Ouyang JH, Chen HL, Liu DY. A computer aided diagnosis system for thyroid disease using extreme learning machine. *J Med Syst* 2012;36:3327-3337
 15. 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;15:853-858
 16. 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
 17. 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;274:2891-2897
 18. 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;40:778-783
 19. 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
 20. Liu JP, Hsueh HM, Hsieh E, Chen JJ. Tests for equivalence or non-inferiority for paired binary data. *Stat Med* 2002;21:231-245
 21. 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
 22. Zhao WJ, Fu LR, Huang ZM, Zhu JQ, Ma BY. Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: a systematic review and meta-analysis. *Medicine (Baltimore)* 2019;98:e16379
 23. 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* 2019;124:118-125