



Clinical validation of S-Detect™ mode in semi-automated ultrasound classification of thyroid lesions in surgical office

Marcin Barczyński^{1#a}, Małgorzata Stopa-Barczyńska^{2#}, Beata Wojtczak³, Agnieszka Czarniecka⁴, Aleksander Konturek^{1b}

¹Department of Endocrine Surgery, Third Chair of General Surgery, Jagiellonian University Medical College, Kraków, Poland; ²Clinical Ward of General Surgery and Oncology, Gabriel Narutowicz Memorial Municipal Hospital, Kraków, Poland; ³Department of General, Minimally Invasive and Endocrine Surgery, Wrocław Medical University, Wrocław, Poland; ⁴Department of Oncological and Reconstructive Surgery, M. Skłodowska-Curie Institute - Oncology Centre, Gliwice, Poland

Contributions: (I) Conception and design: All authors; (II) Administrative support: B Wojtczak, A Czarniecka, A Konturek; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: M Barczyński, M Stopa-Barczyńska, A Konturek; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Marcin Barczyński, MD, PhD, FEBS-ES. Department of Endocrine Surgery, Third Department of General Surgery, Jagiellonian University Medical College, 37 Prądnicka Street, 31-202 Kraków, Poland. Email: marbar@mp.pl.

Background: In recent years well-recognized scientific societies introduced guidelines for ultrasound (US) malignancy risk stratification of thyroid nodules. These guidelines categorize the risk of malignancy in relation to a combination of several US features. Based on these US image lexicons an US-based computer-aided diagnosis (CAD) systems were developed. Nevertheless, their clinical utility has not been evaluated in any study of surgeon-performed office US of the thyroid. Hence, the aim of this pilot study was to validate s-Detect™ mode in semi-automated US classification of thyroid lesions during surgeon-performed office US.

Methods: This is a prospective study of 50 patients who underwent surgeon-performed thyroid US (basic US skills without CAD *vs.* with CAD *vs.* expert US skills without CAD) in the out-patient office as part of the preoperative workup. The real-time CAD system software using artificial intelligence (S-Detect™ for Thyroid; Samsung Medison Co.) was integrated into the RS85 US system. Primary outcome was CAD system added-value to the surgeon-performed office US evaluation. Secondary outcomes were: diagnostic accuracy of CAD system, intra and interobserver variability in the US assessment of thyroid nodules. Surgical pathology report was used to validate the pre-surgical diagnosis.

Results: CAD system added-value to thyroid assessment by a surgeon with basic US skills was equal to 6% (overall accuracy of 82% for evaluation with CAD *vs.* 76% for evaluation without CAD system; $P < 0.001$), and final diagnosis was different than predicted by US assessment in 3 patients (1 more true-positive and 2 more true-negative results). However, CAD system was inferior to thyroid assessment by a surgeon with expert US skills in 6 patients who had false-positive results ($P < 0.001$).

Conclusions: The sensitivity and negative predictive value of CAD system for US classification of thyroid lesions were similar as surgeon with expert US skills whereas specificity and positive predictive value were significantly inferior but markedly better than judgement of a surgeon with basic US skills alone.

Keywords: Thyroid lesions; thyroid cancer; thyroid ultrasound; computer-aided diagnosis (CAD); artificial intelligence

^a ORCID: 0000-0001-6062-1851.

^b ORCID: 0000-0001-8103-4082.

Submitted Oct 21, 2019. Accepted for publication Dec 10, 2019.

doi: 10.21037/gs.2019.12.23

View this article at: <http://dx.doi.org/10.21037/gs.2019.12.23>

Introduction

Thyroid lesions are common and owing to increasing prevalence of thyroid cancer worldwide the effective management of thyroid nodules with accurate selection for surgery has been an ongoing challenge (1). Ultrasound (US) is widely accepted as first-line imaging modality in thyroid diseases as it is noninvasive, cheap, easily available, and cost-effective diagnostic modality. The major obstacle of this technique is that it is observer dependent and its accuracy may differ between different specialties who undertake US examination including radiologists, endocrinologists, and surgeons. In recent years well-recognized scientific societies like American Association of Clinical Endocrinologists (AACE), American Thyroid Association (ATA), American College of Radiologists (ACR), Korean Society of Thyroid Radiology, European Thyroid Association (ETA), and others introduced guidelines for US malignancy risk stratification of thyroid nodules (1-8). In general, these guidelines categorize the risk of malignancy in relation to a combination of several US features as no single feature can reliably predict malignancy (6). Utility of these guidelines lies in providing a practical image guide for clinical usage allowing for a more accurate selection criterion for fine needle aspiration (FNA) cytology assessment based on standardized US risk of malignancy evaluation. In addition, these guidelines proposed a structured format of US thyroid lesions reporting (1). A growing awareness of these guidelines of all physicians involved in thyroid nodules management resulted in optimized patients pathways.

In recent few years, owing to technological progress in building artificial intelligence an US-based computer-aided diagnosis (CAD) system based on semi-automated US image analysis techniques has been developed and introduced to commercially available US machines software (9-11). Its utility has been initially validated in breast and thyroid tumors examined by radiologists with promising results (9-17). Nevertheless, clinical utility of this approach has not been evaluated in any study of surgeon-performed office US of the thyroid before surgery. Hence, we designed this prospective pilot study to test hypothesis that US-based CAD system can be helpful in surgeon office in preoperative assessment of thyroid nodules among patients

referred for thyroid surgery for various indications. The aim of this study was to validate s-DetectTM mode in semi-automated US classification of thyroid lesions during surgeon-performed office US.

Methods

Patients

This is a prospective study of 50 consenting patients who were qualified for thyroid surgery and were treated at the Department of Endocrine Surgery, Third Chair of General Surgery, Jagiellonian University Medical College, Krakow, Poland, from April 2019 till May 2019. All patients in this study underwent surgeon-performed thyroid US in the outpatient office as part of the preoperative workup. Surgical pathology report was used to validate the pre-surgical diagnosis.

Inclusion criterion was referral for surgery by endocrinologist in an adult patient with a solitary thyroid nodule and determined on preoperative cytology as Bethesda II (benign), V (suspicion of malignancy), or VI (malignant). Exclusion criteria were: multiple thyroid nodules, non-diagnostic (Bethesda I) or indeterminate thyroid cytology (Bethesda III, IV), autoimmune thyroid disease, thyroid surgery in history, incomplete clinical data or follow-up information.

Primary outcome was CAD system added-value to the surgeon-performed office US evaluation (defined as improved accuracy of the method).

Secondary outcomes were: diagnostic accuracy of CAD system, intra and interobserver variability in the US assessment of thyroid nodules.

The characteristics of patients in the study are presented in *Table 1*.

All the patients provided written informed consent for the storage and use of their data. The study protocol was approved by the Institutional Review Board. This study was reported in agreement with the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) statement updated in 2015 (18).

US examination

US examinations were done using an RS85 US machine

Table 1 Characteristics of patients in the study

| Parameter | Value |
|-------------------------------|-----------------|
| Mean age \pm SD, years | 47.5 \pm 15.0 |
| Gender, n [%] | |
| Male:female | 9 [18]:41 [82] |
| Mean nodule size \pm SD, mm | |
| Benign* | 34.8 \pm 8.4 |
| Malignant* | 13.6 \pm 5.6 |
| Pathological diagnosis, n (%) | |
| Benign nodule | 40 [80] |
| Papillary thyroid cancer | 10 [20] |

*, $P < 0.001$. SD, standard deviation.

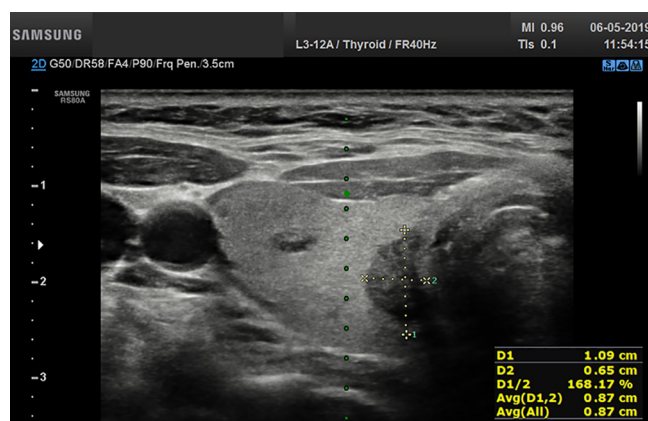


Figure 1 US greyscale transverse image of the right thyroid lobe with a suspicious US feature of a paratracheal lesion in a 42-year-old lady. US, ultrasound.

(Samsung Medison Co., Seoul, South Korea) with a linear array probe (Samsung L3-12a). The real-time CAD system software using artificial intelligence (S-Detect™ for Thyroid; Samsung Medison Co.) was integrated into the US system. S-Detect™ for Thyroid is a novel technology providing the features of the selected mass with information on 6 lexicons (composition, echogenicity, orientation, margin, spongiform, and shape) and recommending malignancy and benignancy of the mass. Among the three major imaging and data reporting systems: K-TIRADS, Russ, and ATA guidelines, the latter one was set in this study as preferred semi-automated thyroid assessment (1,6,7).

All US assessments were done twice in each patient and twice by each observer (on different days): by a surgeon

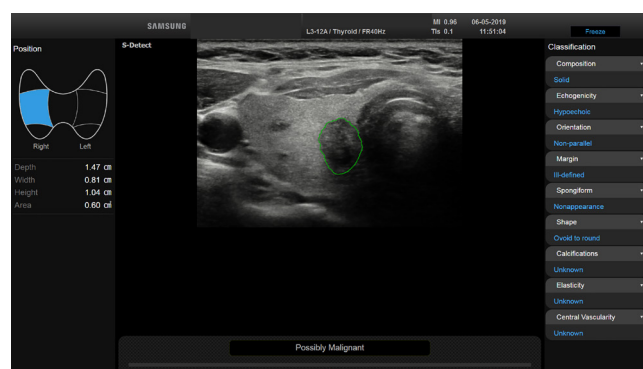


Figure 2 The same image as in Figure 1. Following manual indexing of a suspicious lesion the CAD software automatically calculates mass contours and presents US features on right side of the of screen, and diagnosis as possibly malignant nodule at bottom. CAD, computer-aided diagnosis; US, ultrasonography.

with basic skills (I) in thyroid US imaging (observer with 6 months of training in thyroid US) who used the CAD system and followed by a surgeon with expertise skills (II) in thyroid US imaging (a board certified US observer with >20 years of experience in thyroid US) without use of CAD system (Figure 1). Both surgeons were blinded to the pre-referral radiologist-performed US reports and result of FNA cytology. Participants were scanned consecutively by the observers during the same visit. Only 1 observer was present in the room at any time, and the observers were blinded to each other's measurements; the measurements of the observer were always removed from the screen of US machine before the next observer entered the examination room. All measurements were made by both observers twice. The second measurements were obtained after 1 week in a new random order. Two observers performed new US scans and measurements of thyroid nodules in all patients without knowledge of the previous results.

Before starting the study, the basic methods were discussed to establish a consensus on the use of the CAD system. The CAD data were obtained from transverse planes by manually setting a region of interest around the lesion. The software calculated the mass contours automatically (thereby distinguishing the mass from normal thyroid tissue) and evaluated the US features of the mass using reference to six lexicons described above. The CAD system ultimately diagnosed the nodules as possibly benign or possibly malignant using the previously outlined descriptors (Figures 2,3). Both US operators additionally included two other lexicons in the assessment: calcifications



Figure 3 US greyscale transverse image of the right thyroid lobe with a benign US feature of a lesion in a 53-year-old lady. Following manual selection of the lesion the CAD software automatically calculates mass contours and presents US features on right side of the of screen, and diagnosis as possibly benign nodule at bottom. CAD, computer-aided diagnosis; US, ultrasound.

and vascularity, but not elasticity.

Statistical analysis

The data are presented as means and standard deviations for continuous variables, and as the numbers for categorical variables. Characteristics of US features and CAD diagnosis were tested using Chi-square or Fisher's exact test whereas Student's *t*-test served to evaluate quantitative variables. The diagnostic performance of experienced operator *vs.* surgeon with basic US skills aided by CAD system were estimated with calculations of the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of thyroid malignancy diagnosis. McNemar's test was used to compare the diagnostic sensitivity and specificity of the CAD system and the surgeon with basic skills *vs.* expert skills in US imaging. Additionally, the areas under the receiver operating characteristic (ROC) curve for the CAD system and experienced operator were compared using the method of DeLong *et al.* (19). The intraobserver (A1 versus A2 and B1 versus B2) and interobserver (A1 versus B1) variability was assessed as percentage of agreement between two examinations. For CAD system analysis only first examination performed in each patient was used. All statistical analyses were performed using MedCalc for Windows v15.0 (MedCalc Software, Ostend, Belgium), and a *P* value of <0.05 was considered statistically significant.

Results

Of 194 patients referred for thyroid surgery during the study period 59 patients met the inclusion criterion and were found eligible for the study. The remaining 135 individuals were excluded for the following reasons: multinodular goiter (*n*=54), non-diagnostic cytology (*n*=5), indeterminate cytology (*n*=45), thyroid lesion <5 mm in largest diameter (*n*=16), autoimmune thyroid disease (*n*=15). In addition 9 patients refused to participate in the study leaving 50 patients who were enrolled.

Mean age of patients in the study was 47.5 ± 15.0 years, mean size of a solitary nodule was 30.6 ± 11.6 mm, and it was larger for benign nodules than for malignant nodules (34.8 ± 8.4 *vs.* 13.6 ± 5.6 mm, respectively; *P*<0.001). The diagnosis of benignity (*n*=40) or malignancy (*n*=10) of a solitary nodule was confirmed from a surgical specimen pathology report. All malignant nodules were classic variant of papillary thyroid carcinoma.

Primary outcome

CAD system added-value to thyroid assessment by a surgeon with basic US skills was equal to 6% (overall accuracy of 82% for evaluation with CAD *vs.* 76% for evaluation without CAD system; *P*<0.001), and final diagnosis was different than predicted by US assessment in 3 patients (1 more true-positive and 2 more true-negative results). However, CAD system was inferior to thyroid assessment by a surgeon with expert US skills in 6 patients who had false-positive results (*P*<0.001).

Secondary outcome analysis

Sensitivity, specificity, positive predictive value, negative predictive value and overall accuracy for US examination performed by a surgeon with expert skills in US imaging *vs.* a surgeon with basic US skills without CAD *vs.* a surgeon with basic US skills with CAD system are shown in Table 2. Comparison of intraobserver and interobserver differences between surgeons with expert US skills versus basic US skills in US assessment of thyroid nodules is shown in Table 3.

Discussion

In this prospective pilot study, a clinical evaluation of utility

Table 2 Diagnostic performance of surgeon performed US (expert vs. basic US skills without CAD vs. with CAD)

| Diagnostic test result | Expert US skills without CAD | Basic US skills without CAD | Basic US skills with CAD |
|------------------------|------------------------------|-----------------------------|--------------------------|
| TP, n | 9 | 8 | 9 |
| FN, n | 1 | 2 | 1 |
| FP, n | 2 | 10 | 8 |
| TN, n | 38 | 30 | 32 |
| Sensitivity, % | 90 | 80 | 90 |
| Specificity, % | 95 | 75 | 80 |
| PPV, % | 81.82 | 44.44 | 52.94 |
| NPV, % | 97.44 | 93.75 | 96.97 |
| Accuracy, % | 94 | 76 | 82 |

P<0.001 (for all: expert without CAD vs. basic without CAD; expert without CAD vs. basic with CAD; basic without CAD vs. basic with CAD; McNemar's test). US, ultrasound; CAD, computer-aided diagnosis; TP, true-positive; FN, false-negative; FP, false-positive; TN, true-negative; PPV, positive predictive value; NPV, negative predictive value.

Table 3 Comparison of intraobserver and interobserver differences between surgeons with expert US skills versus basic US skills in US assessment of thyroid nodules

| Lexicon | Indication | Intraobserver difference, % of agreement between 2 examinations | | Difference (%) (95% CI) | P |
|----------------|---------------------------|--|---------------------|----------------------------|-------|
| | | Expert US skills (%) | Basic US skills (%) | | |
| Composition | Solid | 100 | 96 | 4.0 (−3.700 to 13.460) | 0.155 |
| | Partially cystic | | | | |
| | Cystic | | | | |
| Echogenicity | Hyper/isoechoic | 98 | 90 | 8.0 (−2.204 to 19.478) | 0.093 |
| | Hypoechoic | | | | |
| Orientation | Parallel | 100 | 94 | 6.0 (−2.149 to 16.217) | 0.080 |
| | Non-parallel | | | | |
| Margin | Well-defined smooth | 94 | 80 | 14.0 (0.544 to 27.619) | 0.038 |
| | Microlobulated/spiculated | | | | |
| | Ill-defined | | | | |
| Spongiform | Appearance | 100 | 92 | 8.0 (−0.624 to 18.838) | 0.042 |
| | Nonappearance | | | | |
| Shape | Ovoid to round | 100 | 90 | 10.0 (0.897 to 21.360) | 0.022 |
| | Irregular | | | | |
| Calcifications | Micro/macroclicifications | 96 | 72 | 24.0 (9.847 to 37.968) | 0.001 |
| | No calcifications | | | | |
| Elasticity | Stiff | Not assessed | Not assessed | NA | NA |
| | Soft | | | | |
| Vascularity | Appearance | 96 | 90 | 6.0 (−5.020 to 17.723) | 0.242 |
| | Nonappearance | | | | |

All measurements without CAD; comparison of proportions for all. US, ultrasound; CAD, computer-aided diagnosis; 95% CI, 95% confidence interval; NA, not available.

of US CAD system (S-DetectTM) for initial assessment of thyroid nodules and their stratification between benign or malignant lesions was undertaken for the first time in history in surgical out-patient office by two surgeons with differing levels of US skills (basic *vs.* expert) as previous publications in the field reported outcomes of the radiologists performed examinations. In addition, the primary endpoint of this study was the CAD system added-value to the surgeon-performed office US evaluation (defined as improved accuracy of the method) which has never been reported before. Power calculation was not done for this study as it was a pilot one. This study demonstrated that the CAD system added-value to thyroid assessment by a surgeon with basic US skills was equal to 6% and the sensitivity and negative predictive value of CAD system for US classification of thyroid lesions were similar as surgeon with expert US skills whereas specificity and positive predictive value were significantly inferior but markedly better than judgement of a surgeon with basic US skills alone. In addition, surgeon with basic US skill had significantly higher intraobserver variability in the assessment of margins, spongiform *vs.* non spongiform appearance, shape and calcifications of index thyroid lesion than surgeon with expert US skills (Table 3).

The utility of CAD systems was initially reported in breast masses among radiologists with various degrees of experience in breast imaging; 1 *vs.* 7 years of experience (10,11). Cho *et al.* evaluated 119 breast masses (including 54 malignant and 65 benign lesions) with encouraging results. S-DetectTM had higher specificity than radiologists (90.8% *vs.* 49.2% and 55.4%) and positive predictive value (86.7% *vs.* 60.7% and 63.8%), respectively; $P < 0.01$ for all (10). Wu *et al.* evaluated prospectively 338 solid breast lesions (including 129 malignant and 209 benign lesions) in 269 patients with US and S-DetectTM before biopsy or surgical excision and performed multivariate logistic regression analyses to identify the factors associated with false S-DetectTM results. Larger benign lesions, the presence of lesion calcifications, and high degrees of vascularity were likely to show false-positive S-DetectTM results whereas, smaller malignant lesions and the absence of calcifications were likely to show false-negative S-DetectTM results (11).

The CAD systems like S-DetectTM (S-DetectTM for Thyroid, Samsung Medison Co., Seoul, South Korea) for thyroid lesions evaluation are brand new technological adjuncts developed to improve accuracy of radiologists performed US examinations. Chang *et al.* reported that the use of thyroid CAD to differentiate malignant from

benign lesions showed accuracy similar to that obtained via visual inspection by radiologists (12). Choi *et al.* evaluate 102 thyroid nodules from 89 patients (including 43 malignant and 59 benign lesions) and found that the CAD system showed a similar sensitivity as the experienced radiologist (90.7% *vs.* 88.4%, $P > 0.99$), but a lower specificity and a lower area under the receiver operating characteristic (AUROC) curve (specificity: 74.6% *vs.* 94.9%, $P = 0.002$; AUROC: 0.83 *vs.* 0.92, $P = 0.021$). In addition, classifications of the US characteristics (composition, orientation, echogenicity, and spongiform) between radiologist and CAD system were in substantial agreement ($\kappa = 0.659, 0.740, 0.733$, and 0.658, respectively), while the margin showed a fair agreement ($\kappa = 0.239$) (9). Gitto *et al.* evaluated 62 thyroid nodules (including 14 indeterminate to malignant and 48 benign lesions) and found that interobserver agreement between the CAD system and the radiologist was substantial for orientation ($\kappa = 0.69$), fair for composition ($\kappa = 0.36$), echogenicity ($\kappa = 0.36$), K-TIRADS ($\kappa = 0.29$), and slight for margins ($\kappa = 0.03$). Hence, the CAD system was considered to be less sensitive than an experienced radiologist and it showed slight-to-substantial agreement with the radiologist for the characterization of thyroid nodules (13). Similar data were reported by Gao *et al.* who retrospectively reviewed 342 surgically resected thyroid nodules previously assessed on typical US images using the CAD system and reviewed by an experienced radiologist using the TIRADS and ATA guidelines. The sensitivity of a thyroid US CAD system in differentiating nodules was similar to that of an experienced radiologist. However, the CAD system had lower specificity (14). Jeong *et al.* evaluated the diagnostic performance and reproducibility of a CAD system for thyroid cancer diagnosis using US based on the operator's experience in a study cohort of 100 thyroid nodules. The sensitivity and accuracy of the CAD system did not differ significantly in this study from those of the experienced radiologist while the specificity was significantly higher for the experienced radiologist. In addition, the diagnostic performance varied according to the operator's experience and they were lower for the less-experienced operators than for the experienced radiologist. The interobserver agreement was substantial for the final diagnosis and each US descriptor and moderate for the margin and composition (15). Kim *et al.* evaluated the diagnostic performance of a CAD system (S-DetectTM) for detecting thyroid cancers among 218 thyroid nodules in 106 patients. The inter-observer agreement between the CAD system and radiologist for the description of calcifications was

fair ($\kappa=0.336$), while the final diagnosis and each US descriptor showed moderate to substantial agreement for the S-DetectTM. One of the main limitations of the S-DetectTM was its inaccuracy in recognizing calcifications, which meant that differentiation had to be undertaken by the radiologist (16). Zhao *et al.* evaluated recently the diagnostic performances of CAD systems *vs.* radiologist by meta-analysis of currently published studies. Five studies with 536 patients and 723 thyroid nodules were included in this meta-analysis. The sensitivity of the CAD system in the diagnosis of thyroid nodules was similar to that of experienced radiologists. However, the CAD system had lower specificity and diagnostic odds ratio than experienced radiologists (17).

Several US findings are crucial for optimal surgical planning in patients with thyroid cancer but they are rarely mentioned on preconsultation radiologist-performed US (20). Hence, surgeon-performed US may be a useful tool in the diagnosis and accurate staging of patients with thyroid cancer. Solorzano reported that surgeon-performed US changed and enhanced the pre- and intraoperative management in more than half the patients with thyroid cancer (including identification and guidance for the FNA of nonpalpable cancers, identification of nonpalpable nodules in the contralateral lobe, preoperative diagnosis of nonpalpable metastatic lymph nodes, and intraoperative guidance for their excision) (21). Carneiro-Pla *et al.* compared the preconsultation US findings with surgeon-performed US among 136 patients with thyroid cancer and identified changes in the management of as many as 61 (45%) patients by identifying preoperatively central and/or lateral node metastasis, indicating preoperative biopsy of suspicious thyroid lesions/nodes, and pointing out thyroid intrathoracic extension (20). When compared to surgeon-performed US, the preconsultation US failed to mention node status in 101 (74%) patients, suspicious nodule features in 60/111 (54%) patients with suspicious lesions, bilateral thyroid lesions in 19/88 (22%) patients with bilateral nodules, local invasion in 5/5 (100%), and intrathoracic extension in 5/5 (100%) cases (20). Hence, US turned out to be more accurate and adequate in the evaluation of almost half of patients qualified for surgery for thyroid cancer when performed by the surgeon with expert skills in US imaging (20). These findings are also supported by many more publications in the field (22,23). Moreover, as shown by outcomes of this study the CAD system may play the potential additional role as a decision-making assistant to surgeon with basic US skills in the surgical out-

patient office in the stratification of the risk of malignancy of thyroid nodules referred for consideration for surgery. As shown by outcomes of this study among exams performed by a surgeon with expert US skills there were two false positive results in patients with index lesions appearing on US as deeply hypoechoic ones with ill-defined margins (in pathological report diagnosed as subacute thyroiditis in one case and benign hyperplastic thyroid nodule in the latter one). Among exams performed by a surgeon with basic US skills there were 10 false positive results, including 2 cases described above and additional 8 other cases classified on US as hypoechoic with microlobulated or ill-defined margin ($n=2$), non-parallel in orientation ($n=2$), and a lesion with microcalcifications ($n=4$), all of which were finally diagnosed in pathological report as benign hyperplastic thyroid nodules. Use of the CAD system improved assessment of surgeon with basic US skills in 2 cases upgrading index lesions from ill-defined to well-defined margin category. Such an aid in decision-making about whether FNA or surgery is indicated or unjustified may optimize patients' pathways in preoperative work up.

This pilot study has several limitations. First of all it was based on a relatively small population of patients preselected for surgery with a solitary thyroid lesion or nodule and treated at a single institution with prevalence of classical form of PTC equal to 20%. However, US presentation of other thyroid malignancies (follicular variant of PTC, follicular carcinoma, or medullary thyroid cancer, lymphoma) may be different from those of classical PTC. In addition, patients with non-diagnostic (Bethesda I) or indeterminate (Bethesda III, IV) thyroid cytology as well as patients with multinodular goiter were excluded which suggests need for further assessments in a less homogenous study population and more representative for generalized environment of various thyroid nodules. It is also important to note that FNA was done by the referring endocrinologist at least 1 month prior to referral for surgery. Such a time separation was considered long enough to minimize possible influence of perinodular hemorrhage on prevalence of false results of US evaluation. Moreover, this study was undertaken by only two US operators with differing level of US skills (basic *vs.* expert). To draw more generalized conclusions inter-observer agreement should be assessed in further multicenter studies. Another limitation was unavailability of the current US CAD format for evaluation of calcifications, vascularity and elasticity of thyroid nodules which should be improved in future.

Current artificial intelligence development has a diagnostic performance that is comparable with medical experts,

especially in image recognition-related fields like US thyroid imaging (24). However, future technical improvements in automatic image recognition and diagnosis systems based on deep learning using the neural network may shift the current format of CAD into a brand new dimension of real artificial intelligence leading to increased accuracy as well as diagnostic efficiency of the CAD imaging modality (25-27).

Conclusions

The sensitivity and negative predictive value of CAD system for US classification of thyroid lesions were similar as surgeon with expert US skills whereas specificity and positive predictive value were significantly inferior but markedly better than judgement of a surgeon with basic US skills alone. Hence, the system may be useful for ruling out malignancy and more appropriate selection of thyroid nodules for FNA cytology assessment at surgical office and more accurate selection for unilateral thyroid surgery. However, further multicenter studies validating clinical utility of US CAD systems in stratification of thyroid nodules should include a larger number of US operators with varying degree of experience and different populations of patients with various prevalence of malignancy and other thyroid diseases.

Acknowledgments

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

1. 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.
2. Grant EG, Tessler FN, Hoang JK, et al. Thyroid Ultrasound Reporting Lexicon: White Paper of the ACR Thyroid Imaging, Reporting and Data System (TIRADS) Committee. *J Am Coll Radiol* 2015;12:1272-9.
3. Na DG, Baek JH, Sung JY, et al. Thyroid Imaging Reporting and Data System Risk Stratification of Thyroid Nodules: Categorization Based on Solidity and Echogenicity. *Thyroid* 2016;26:562-72.
4. Gharib H, Papini E, Garber JR, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinologi. Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules—2016 Update. *Endocr Pract* 2016;22:622-39.
5. 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;14:587-95.
6. Russ G, Bonnema SJ, Erdogan MF, et al. European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. *Eur Thyroid J* 2017;6:225-37.
7. Shin JH, Baek JH, Chung J, et al. Ultrasonography Diagnosis and Imaging-Based Management of Thyroid Nodules: Revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean J Radiol* 2016;17:370-95.
8. Gao L, Xi X, Jiang Y, et al. Comparison among TIRADS (ACR TI-RADS and KWAK- TI-RADS) and 2015 ATA Guidelines in the diagnostic efficiency of thyroid nodules. *Endocrine* 2019;64:90-6.
9. Choi YJ, Baek JH, Park HS, et al. Computer-Aided Diagnosis System Using Artificial Intelligence for the Diagnosis and Characterization of Thyroid Nodules on Ultrasound: Initial Clinical Assessment. *Thyroid* 2017;27:546-52.
10. Cho E, Kim EK, Song MK, et al. 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;37:209-16.
11. Wu JY, Zhao ZZ, Zhang WY, et al. Computer-Aided Diagnosis of Solid Breast Lesions With Ultrasound:

- Factors Associated With False-negative and False-positive Results. *J Ultrasound Med* 2019. [Epub ahead of print].
12. Chang Y, Paul AK, Kim N, 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;43:554-67.
 13. Gitto S, Grassi G, De Angelis C, et al. A computer-aided diagnosis system for the assessment and characterization of low-to-high suspicion thyroid nodules on ultrasound. *Radiol Med* 2019;124:118-25.
 14. 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;40:778-83.
 15. Jeong EY, Kim HL, Ha EJ, et al. Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. *Eur Radiol* 2019;29:1978-85.
 16. Kim HL, Ha EJ, Han M. Real-World Performance of Computer-Aided Diagnosis System for Thyroid Nodules Using Ultrasonography. *Ultrasound Med Biol* 2019;45:2672-8.
 17. Zhao WJ, Fu LR, Huang ZM, et al. 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.
 18. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
 19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
 20. Carneiro-Pla D, Amin S. Comparison between preconsultation ultrasonography and office surgeon-performed ultrasound in patients with thyroid cancer. *World J Surg* 2014;38:622-7.
 21. Solorzano CC, Carneiro DM, Ramirez M, et al. Surgeon-performed ultrasound in the management of thyroid malignancy. *Am Surg* 2004;70:576-80.
 22. Sipos JA. Advances in ultrasound for the diagnosis and management of thyroid cancer. *Thyroid* 2009;19:1363-72.
 23. Mazzaglia PJ. Surgeon-performed ultrasound in patients referred for thyroid disease improves patient care by minimizing performance of unnecessary procedures and optimizing surgical treatment. *World J Surg* 2010;34:1164-70.
 24. Prochazka A, Gulati S, Holinka S, et al. Classification of Thyroid Nodules in Ultrasound Images Using Direction-Independent Features Extracted by Two-Threshold Binary Decomposition. *Technol Cancer Res Treat* 2019;18:1533033819830748.
 25. Wang L, Yang S, Yang S, et al. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World J Surg Oncol* 2019;17:12.
 26. Song J, Chai YJ, Masuoka H, et al. Ultrasound image analysis using deep learning algorithm for the diagnosis of thyroid nodules. *Medicine (Baltimore)* 2019;98:e15133.
 27. Zhou S, Chen W, Chai YJ. Artificial intelligence for thyroid ultrasound image analysis. *Ann Thyroid* 2019;4:11.

Cite this article as: Barczyński M, Stopa-Barczyńska M, Wojtczak B, Czarniecka A, Konturek A. Clinical validation of S-Detect™ mode in semi-automated ultrasound classification of thyroid lesions in surgical office. *Gland Surg* 2020;9(Suppl 2):S77-S85. doi: 10.21037/gs.2019.12.23