Journal of Ultrasound (2020) 23:169–174
https://doi.org/10.1007/s40477-020-00453-y

ORIGINAL PAPER

Computer-aided diagnostic system for thyroid nodule sonographic evaluation outperforms the specificity of less experienced examiners

Daniele Fresilli[1] · Giorgio Grani[2] · Maria Luna De Pascali[1] · Gregorio Alagna[1] · Eleonora Tassone[1] · Valeria Ramundo[2] · Valeria Ascoli[1] · Daniela Bosco[1] · Marco Biffoni[3] · Marco Bononi[4] · Vito D'Andrea[3] · Fabrizio Frattaroli[5] · Laura Giacomelli[3] · Yana Solskaya[6] · Giorgia Polti[1] · Patrizia Pacini[1] · Olga Guiban[1] · Raffaele Gallo Curcio[7] · Marcello Caratozzolo[1] · Vito Cantisani[1]

Received: 29 December 2019 / Accepted: 13 March 2020 / Published online: 3 April 2020

Abstract

Purpose Computer-aided diagnosis (CAD) may improve interobserver agreement in the risk stratification of thyroid nodules. This study aims to evaluate the performance of the Korean Thyroid Imaging Reporting and Data System (K-TIRADS) classification as estimated by an expert radiologist, a senior resident, a medical student, and a CAD system, as well as the interobserver agreement among them.

Methods Between July 2016 and 2018, 107 nodules (size 5–40 mm, 27 malignant) were classified according to the K-TIRADS by an expert radiologist and CAD software. A third-year resident and a medical student with basic imaging training, both blinded to previous findings, retrospectively estimated the K-TIRADS classification. The diagnostic performance was calculated, including sensitivity, specificity, positive and negative predictive values, and the area under the receiver operating characteristic curve.

Results The CAD system and the expert achieved a sensitivity of 70.37% (95% CI 49.82–86.25%) and 81.48% (61.92–93.7%) and a specificity of 87.50% (78.21–93.84%) and 88.75% (79.72–94.72%), respectively. The specificity of the student was significantly lower (76.25% [65.42–85.05%], $p = 0.02$).

Conclusion In our opinion, the CAD evaluation of thyroid nodules stratification risk has a potential role in a didactic field and does not play a real and effective role in the clinical field, where not only images but also specialistic medical practice is fundamental to achieve a diagnosis based on family history, genetics, lab tests, and so on. The CAD system may be useful for less experienced operators as its specificity was significantly higher.

Keywords Thyroid nodule · Medical students · Observer variation · Computer-assisted diagnosis

Abbreviations

AUROC	Area under the Receiver Operating Characteristic curve
CAD	Computer-aided diagnosis
CI	Confidence interval
FNA	Fine-needle aspiration
NPV	Negative predictive value
PACS	Picture archiving and communication system
PPV	Positive predictive value
ROI	Region of interest
TI-RADS	Thyroid imaging, reporting, and data system
US	Ultrasonography

✉ Daniele Fresilli
daniele.fresilli@hotmail.it

[1] Department of Radiological, Oncological, and Pathological Sciences, "Sapienza" University of Rome, Rome, Italy

[2] Department of Translational and Precision Medicine, "Sapienza" University of Rome, Rome, Italy

[3] Department of Surgical Sciences, "Sapienza" University of Rome, Rome, Italy

[4] Department of Surgery "P. Valdoni", Sapienza" University of Rome, Rome, Italy

[5] Department of Surgery "P. Stefanini", Sapienza" University of Rome, Rome, Italy

[6] Pauls Stradins Clinical University Hospital, Riga, Latvia

[7] Department of Health Management, Policlinico Umberto I-"Sapienza" University of Rome, Rome, Italy

Introduction

Thyroid nodules are commonly found during imaging of the neck [1, 2], but only a small proportion of these lesions subsequently prove to be clinically significant [3]. Nowadays, neck ultrasonography (US) is used to guide decisions regarding fine-needle aspiration (FNA) cytology or serial follow-up. Nodules should be carefully selected for FNA biopsy [4, 5] because of the vast number of subjects concerned, the potentially inconclusive results (both non-diagnostic [6] and indeterminate), and the risk of overdiagnosis of low-risk cancers [7]. However, while some US features are associated with nodule malignancy, the diagnostic accuracy is limited, and substantial interobserver variability has been documented [8–19].

Several classification systems that combine various US findings have been developed to estimate the likelihood of malignancy and select nodules for FNA biopsy. The application of these systems, which are endorsed by international scientific societies [1, 20–23], has proven to reduce interobserver variability, but there is room for further improvement [24]. To this end, using artificial intelligence, CAD has been proven to differentiate malignant from benign nodules, with an accuracy rate similar to expert radiologists [25–28] and may also reduce intra- and interobserver variability.

We performed a prospective analysis of sonographic examinations to evaluate the diagnostic performance of the K-TIRADS classification as estimated by an expert attending radiologist, a senior resident, a medical student, and a CAD system (S-Detect™), as well as the interobserver agreement among them. Furthermore, we evaluated the interobserver agreement between S-Detect™ and the expert radiologist in the evaluation of single sonographic features.

Methods

Cases

Between July 2016 and July 2018, 555 nodules were consecutively examined at the Head and Neck Radiology Unit of Policlinico Umberto I, "Sapienza" University of Rome (Italy), by a single radiologist. Patients were enrolled if they had no more than three thyroid nodules. Target nodules were submitted to cytological examination or thyroidectomy. The exclusion criteria were cystic lesions, nodules smaller than 5 mm (to avoid misinterpretations of the pathology report), examinations performed on an appliance unequipped with CAD software, or cytopathology provided by external services. The original examination was performed by a single attending radiologist with 18 years of experience in thyroid imaging using a Samsung RS80 system equipped with a 7–15 MHz linear probe and the S-Detect™ software. All nodules were characterized in terms of size, shape, margins, composition (solid, mixed, or cystic content), echogenicity, calcifications, hyperechoic foci, vascularity, and extrathyroidal extension. The operator then arranged the nodules according to the K-TIRADS [23] classification. Afterward, the S-Detect™ software was used to automatically determine the shape, composition, echogenicity, and margins. Other features, such as the presence of calcifications, stiffness according to elastosonography, and color Doppler vascular pattern, had to be input manually. The S-Detect™ system is semi-automatic since it requires the operator to select a region of interest (ROI), and then it performs segmentation and recognition of the nodule boundaries. If the process was incorrectly performed, the operator has to repeat it to obtain a correct recognition of the nodule.

All images were stored in a picture archiving and communication system (PACS) for retrospective analysis. A resident with three years of experience and a medical student with basic thyroid imaging training, both blinded to all clinical, pathological, and S-Detect™ findings, provided their estimation of the K-TIRADS classification for each nodule.

Reference standard

A composite reference standard was used: the final histology for nodules undergoing thyroid surgery and FNA cytology for patients with benign cytology findings. In this latter case, a nodule stability confirmed by at least a 12-months follow-up was also required. FNA specimens were prepared as direct smears and assessed by two expert cytopathologists (V. A. and D. B.) in accordance with national guidelines [29].

Statistical analysis

The classifications (K-TIRADS 2–3 were considered as test negative and K-TIRADS 4–5 as positive) were compared with the reference standard to estimate the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic (AUROC) curve, each with 95% confidence intervals (CIs) of each classification. Interobserver agreement was assessed with Krippendorff's alpha [30] for analyses involving ordinal data and more than two observers and with Cohen's kappa for analyses involving dichotomous variables and two observers. Values less than 0.20 were considered indicative of slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement;

0.61–0.80, substantial agreement; and 0.81–1.00, near-perfect agreement [31]. Sensitivity and specificity were compared using McNemar's test [32]. Data were analyzed using the IBM SPSS Statistics package, version 25.0 (IBM Corp., Armonk, New York, United States). AUROC was computed and compared using the easyROC package [33]. Institutional Review Board approval and informed consent was obtained from all subjects.

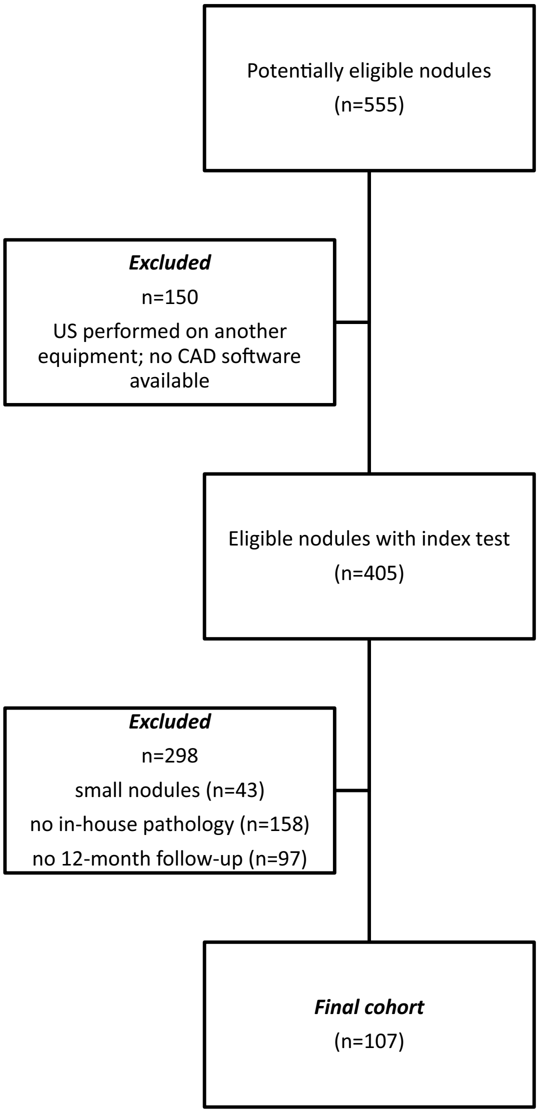
Potentially eligible nodules
(n=555)
Excluded
n=150
US performed on another equipment; no CAD software available
Eligible nodules with index test
(n=405)
Excluded
n=298
small nodules (n=43)
no in-house pathology (n=158)
no 12-month follow-up (n=97)
Final cohort
(n=107)

Fig. 1 Inclusion of the thyroid nodules in the final analysis

Results

The final cohort was composed of 76 patients (56 females and 20 males; mean age 55 years; Fig. 1) and a total of 107 nodules (size range 5–40 mm). Among them, 27 were malignant (25.2%; 17 papillary thyroid cancers, six follicular variant papillary thyroid cancers, three medullary thyroid cancers, and one poorly differentiated carcinoma), and 80 were benign (74.8%; 22 histologically confirmed and 58 diagnosed based on cytology reports and follow-ups). The sensitivity, specificity, predictive values, and AUROC are reported in Table 1.

The overall discriminant ability of S-Detect™ and the three examiners was not statistically different. However, the sensitivity of the expert radiologist seems to be better than S-Detect™ (81.5% vs. 70.4%; $p=0.25$). The expert missed five malignancies, while S-Detect™ missed eight (similar to the beginner medical student).

The S-Detect™ software had the greatest diagnostic agreement with the expert radiologist, whereas agreement decreased with less experienced examiners. This was confirmed both for the dichotomic classification (suspicious vs. not suspicious; Table 2) and for the complete K-TIRADS classification (Table 3). A substantial agreement exists between the experienced radiologist and S-Detect™ in the assessment of single US features (Table 4). However, this result may be biased because of the use of the software by the senior radiologist alone (who personally performed the ROI selection). To obtain a correct segmentation and recognition of the nodule boundaries, the examiner had to perform the ROI selection procedure for a median of three times (range 1–5 times) before achieving a correct segmentation of the nodule.

Table 1 Sensitivity, specificity, predictive values, and area under the receiver operating characteristics curve of the K-TIRADS system evaluated by S-Detect™ and three clinicians (expert attending radiologist, resident, and medical student)

	Sensitivity	Specificity	PPV	NPV	AUROC
S-Detect™	70.37% (49.82–86.25%)	87.50%‡ (78.21–93.84%)	65.52% (45.67–82.06%)	89.74% (80.79–95.47%)	0.79 (0.69–0.88)
Expert	81.48% (61.92–93.7%)	88.75%† (79.72–94.72%)	70.97% (51.96–85.78%)	93.42% (85.31–97.83%)	0.85 (0.77–0.93)
Resident	74.07% (53.72–88.89%)	85.00%* (75.26–92.0%)	62.50% (43.69–78.9%)	90.67% (81.71–96.16%)	0.80 (0.7–0.89)
Student	70.37% (49.82–86.25%)	76.25% (65.42–85.05%)	50.00% (33.38–66.62%)	88.41% (78.43–94.86%)	0.73 (0.63–0.83)

The specificity of the student is significantly lower than that of the expert and S-Detect™ ($p=0.16$ vs. the resident; $^{\dagger}p=0.02$ vs. the expert; $^{\ddagger}p=0.022$ vs. S-Detect™); no significant difference was reported between S-Detect™ and the more experienced examiners

Table 2 Agreement in dichotomic TIRADS classification (K-TIRADS 2–3 vs. 4–5) between S-Detect™ and the three examiners (Cohen Kappa coefficient ± Standard error)

	S-Detect™	Expert	Resident	Student
S-Detect™	–	0.815 ± 0.063	0.748 ± 0.071	0.634 ± 0.079
Expert	0.815 ± 0.063	–	0.888 ± 0.049	0.723 ± 0.071
Resident	0.748 ± 0.071	0.888 ± 0.049	–	0.788 ± 0.063
Student	0.634 ± 0.079	0.723 ± 0.071	0.788 ± 0.063	–

Table 3 Agreement in TIRADS classification between S-Detect™ and the three examiners (Krippendorff's alpha)

	Agreement with S-Detect™ (Krippendorff's alpha)
All three examiners	0.79 (0.73–0.85)
Expert radiologist	0.84 (0.75–0.92)
Resident	0.77 (0.66–0.87)
Medical student	0.69 (0.57–0.80)

Table 4 Agreement in single sonographic features between S-Detect™ and the expert radiologist (Krippendorff's alpha)

	Agreement between expert radiologist and S-Detect™ (Krippendorff's alpha)
Composition	0.79 (0.65–0.93)
Echogenicity	0.71 (0.59–0.81)
Margins	0.74 (0.58–0.90)
Shape	0.71 (0.38–0.94)

Discussion

In the last years, neck ultrasonography has become the cornerstone of the diagnostic algorithm of thyroid nodules, as well as ultrasound elastography [34, 35]. Its main drawbacks are the inadequate predictive values of single US features and the well-known operator dependency. Efforts have been made to improve the discriminative power of ultrasound and reduce interobserver variability through the use of image analysis [19], machine learning, and CAD systems [27, 28, 36]. Several scientific societies have proposed sonographic classification systems to help clinicians categorize and report US features of thyroid nodules. Their application, however, is time-consuming and requires specific training [24].

S-Detect™ is commercially available and has been clinically validated in previous studies [27, 28]. Choi and colleagues [26] tested S-Detect™ on 102 nodules (89 patients) and reported that a radiologist with 20 years of experience had better specificity (94.9% vs. 74.9%, $p = 0.002$) and discriminant power (AUROC 0.92 vs. 0.83, $p = 0.021$) than the software, though the sensitivity was similar. Similar figures were reported by Gao and colleagues [37]. Yoo et al. [38] also compared the diagnostic performance of a radiologist with 10 years of experience before and after using the S-Detect™ software as a "second opinion." After S-Detect™ support, the sensitivity of the examiner increased (92% vs. 84%), but a slight reduction in specificity (85.1% vs. 95.5%) and PPV (82.1% vs. 93.3%) occurred. Jeong et al. [39] studied the application of a CAD system by radiologists with different levels of experience and found that diagnostic performance varied according to operator experience; in particular, the sensitivity ranged from 70.5 to 88.6%. Overall, a meta-analysis confirmed that the sensitivity of the available CAD systems is similar to that of experienced radiologists, with lower specificity and diagnostic odds ratio [40]. Another recent study tested the diagnostic performance of a newer version of the S-Detect™ software. The authors concluded that current systems had limited specificity in the diagnosis of thyroid cancer and that, even if the new version aims to recognize calcifications, this evaluation is not accurate [41].

In our cohort, S-Detect™, a CAD system designed to simplify the scoring and reporting of thyroid nodules, achieved a sensitivity and specificity which were statistically comparable to those obtained by an expert radiologist or a senior resident. Furthermore, it outperformed the less experienced examiner. However, S-Detect™ required manual input of some features (including the presence of microcalcification, which is a crucial finding) and multiple attempts for the correct segmentation of the lesion: the selection of the ROI has a great influence on the final evaluation.

Our study has some limitations. First, this was a relatively small and select cohort of thyroid nodules, all of which had already been selected for FNA biopsy or surgery. This is reflected in the high malignancy rate (25.2%). Second, the reference standard used may have caused false-negative results, even if they are uncommon. On the contrary, all malignancies were histologically confirmed. This study was conducted in a single center, and all US examinations were performed by a radiologist with extensive experience, who was also involved in multiple research programs related to TIRADS systems [4, 18, 24, 42, 43]. This may impact the applicability of our findings to other settings. Also, the software was directly used by the senior radiologist alone.

Conclusion

S-Detect™ had a diagnostic performance similar to that of an experienced radiologist. While it does not provide a clinical advantage to the expert clinician, it may be a useful tool for the less experienced operator as its specificity was significantly higher. It may also be used for training purposes as an aid in recognizing suspicious US features and expediting learning of the TIRADS scoring process and its practical application.

Acknowledgements GG and VR contributed to this paper as part of their Ph.D. studies in Biotechnologies and Clinical Medicine at the University of Rome, Sapienza.

Author contributions MLDP, PG, PP, GO, DF, MC and VC participated in the conception and design of the study; VC collected personal classification and S-Detect data; MLDP and GA performed a subjective classification of nodule images; GA, ET, VR collected clinical data and linked nodule data with final histology reports; VA and DB provided cytology data; MB, VDA, FF, and LG provided patients' data and performed surgeries. DF and GG performed the statistical analysis and were the major contributors in writing the manuscript. All authors read and approved the final manuscript.

Funding This work received no specific funding.

Compliance with ethical standards

Conflict of interest Vito Cantisani lectured for Bracco, Samsung, Canon. The other authors declare that they have no conflict of interest.

Ethics approval and consent to participate 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 1975 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

1. Gharib H, Papini E, Garber JR et al (2016) American Association of Clinical Endocrinologists, American College of Endocrinology, and Associazione Medici Endocrinology medical guidelines for clinical practice for the diagnosis and management of thyroid nodules—2016 update. Endocr Pract 22:622–639. https://doi.org/10.4158/EP161208.GL
2. Guth S, Theune U, Aberle J et al (2009) Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. Eur J Clin Invest 39:699–706. https://doi.org/10.1111/j.1365-2362.2009.02162.x
3. Durante C, Grani G, Lamartina L et al (2018) The diagnosis and management of thyroid nodules. A review. JAMA 319:914–924. https://doi.org/10.1001/jama.2018.0898
4. Grani G, Lamartina L, Ascoli V et al (2019) Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: towards the "right" TIRADS. J Clin Endocrinol Metab 104:95–102. https://doi.org/10.1210/jc.2018-01674
5. Nabahati M, Moazezi Z, Fartookzadeh S, Mehraeen R, Ghaemian N, Sharbatdaran M (2019) The comparison of accuracy of ultrasonographic features versus ultrasound-guided fine-needle aspiration cytology in diagnosis of malignant thyroid nodules. J Ultrasound 22(3):315–321. https://doi.org/10.1007/s40477-019-00377-2**(Epub 2019 Apr 10)**
6. Grani G, Calvanese A, Carbotta G et al (2013) Intrinsic factors affecting adequacy of thyroid nodule fine-needle aspiration cytology. Clin Endocrinol 78:141–144. https://doi.org/10.1111/j.1365-2265.2012.04507.x
7. Brito JP, Davies L, Zeballos-Palacios C et al (2014) Papillary lesions of indolent course: reducing the overdiagnosis of indolent papillary thyroid cancer and unnecessary treatment. Future Oncol 10:1–4. https://doi.org/10.2217/fon.13.240
8. Choi SH, Kim EK, Kwak JY et al (2010) Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. Thyroid 20:167–172. https://doi.org/10.1089/thy.2008.0354
9. Kim HG, Kwak JY, Kim EK et al (2012) Man to man training: can it help improve the diagnostic performances and interobserver variabilities of thyroid ultrasonography in residents? Eur J Radiol 81:e352–e356. https://doi.org/10.1016/j.ejrad.2011.11.011
10. Kim SH, Park CS, Jung SL et al (2010) Observer variability and the performance between faculties and residents: US criteria for benign and malignant thyroid nodules. Korean J Radiol 11:149–155. https://doi.org/10.3348/kjr.2010.11.2.149
11. Koltin D, O'Gorman CS, Murphy A et al (2016) Pediatric thyroid nodules: ultrasonographic characteristics and inter-observer variability in prediction of malignancy. J Pediatr Endocrinol Metab 29:789–794. https://doi.org/10.1515/jpem-2015-0242
12. Lim-Dunham JE, Erdem Toslak I, Alsabban K et al (2017) Ultrasound risk stratification for malignancy using the 2015 American Thyroid Association Management Guidelines for children with thyroid nodules and differentiated thyroid cancer. Pediatr Radiol 47:429–436. https://doi.org/10.1007/s00247-017-3780-6
13. Norlen O, Popadich A, Kruijff S et al (2014) Bethesda III thyroid nodules: the role of ultrasound in clinical decision making. Ann Surg Oncol 21:3528–3533. https://doi.org/10.1245/s10434-014-3749-8
14. Park SH, Kim SJ, Kim EK et al (2009) Interobserver agreement in assessing the sonographic and elastographic features of malignant thyroid nodules. AJR Am J Roentgenol 193:W416–W423. https://doi.org/10.2214/ajr.09.2541
15. Park CS, Kim SH, Jung SL et al (2010) Observer variability in the sonographic evaluation of thyroid nodules. J Clin Ultrasound 38:287–293. https://doi.org/10.1002/jcu.20689
16. Park SJ, Park SH, Choi YJ et al (2012) Interobserver variability and diagnostic performance in US assessment of thyroid nodule according to size. Ultraschall Med 33:E186–E190. https://doi.org/10.1055/s-0032-1325404
17. Wienke JR, Chong WK, Fielding JR et al (2003) Sonographic features of benign thyroid nodules: interobserver reliability and overlap with malignancy. J Ultrasound Med 22:1027–1031
18. Grani G, Lamartina L, Ascoli V et al (2017) Ultrasonography scoring systems can rule out malignancy in cytologically indeterminate thyroid nodules. Endocrine 57:256–261. https://doi.org/10.1007/s12020-016-1148-6
19. Grani G, D'Alessandri M, Carbotta G, Nesca A, Del Sordo M, Alessandrini S, Coccaro C, Rendina R, Bianchini M, Prinzi N, Fumarola A (2015) Grey-scale analysis improves the ultrasonographic evaluation of thyroid nodules. Medicine (Baltimore). 94(27):e1129. https://doi.org/10.1097/MD.0000000000001129
20. Tessler FN, Middleton WD, Grant EG et al (2017) ACR Thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. J Am Coll Radiol 14:587–595. https://doi.org/10.1016/j.jacr.2017.01.046

21. Haugen BR, Alexander EK, Bible KC et al (2016) 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 26:1–133. https://doi.org/10.1089/thy.2015.0020
22. Russ G, Bonnema SJ, Erdogan MF et al (2017) European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. Eur Thyroid J 6:225–237. https://doi.org/10.1159/000478927
23. Shin JH, Baek JH, Chung J et al (2016) Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. Korean J Radiol 17:370–395. https://doi.org/10.3348/kjr.2016.17.3.370
24. Grani G, Lamartina L, Cantisani V et al (2018) Interobserver agreement of various thyroid imaging reporting and data systems. Endocr Connect 7:1–7. https://doi.org/10.1530/EC-17-0336
25. Chang Y, Paul AK, Kim N et al (2016) Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: a comparison with radiologist-based assessments. Med Phys 43:554. https://doi.org/10.1118/1.4939060
26. Choi YJ, Baek JH, Park HS et al (2017) A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment. Thyroid 27:546–552. https://doi.org/10.1089/thy.2016.0372
27. Di Segni M, de Soccio V, Cantisani V, Bonito G, Rubini A, Di Segni G, Lamorte S, Magri V, De Vito C, Migliara G, Bartolotta TV, Metere A, Giacomelli L, de Felice C, D'Ambrosio F (2018) Automated classification of focal breast lesions according to S-detect: validation and role as a clinical and teaching tool. J Ultrasound 21(2):105–118. https://doi.org/10.1007/s40477-018-0297-2**(Epub 2018 Apr 21)**
28. Gitto S, Grassi G, De Angelis C, Monaco CG, Sdao S, Sardanelli F, Sconfienza LM, Mauri G (2019) A computer-aided diagnosis system for the assessment and characterization of low-to-high suspicion thyroid nodules on ultrasound. Radiol Med 124(2):118–125. https://doi.org/10.1007/s11547-018-0942-z**(Epub 2018 Sep 22)**
29. Nardi F, Basolo F, Crescenzi A et al (2014) Italian consensus for the classification and reporting of thyroid cytology. J Endocrinol Invest 37:593–599. https://doi.org/10.1007/s40618-014-0062-0
30. Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. Commun Methods Meas 1:77–89. https://doi.org/10.1080/19312450709336664
31. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
32. Trajman A, Luiz RR (2008) McNemar chi2 test revisited: comparing sensitivity and specificity of diagnostic examinations. Scand J Clin Lab Invest 68:77–80. https://doi.org/10.1080/00365510701666031
33. Goksuluk D, Korkmaz S, Zararsiz G et al (2016) easyROC: an interactive web-tool for ROC curve analysis using R language environment. R J 8:213–230
34. Cantisani V, David E, Grazhdani H, Rubini A, Radzina M, Dietrich CF, Durante C, Lamartina L, Grani G, Valeria A, Bosco D, Di Gioia C, Frattaroli FM, D'Andrea V, De Vito C, Fresilli D, D'Ambrosio F, Giacomelli L, Catalano C (2019) Prospective evaluation of semiquantitative strain ratio and quantitative 2D ultrasound shear wave elastography (SWE) in association with TIRADS classification for thyroid nodule characterization. Ultraschall Med 40(4):495–503. https://doi.org/10.1055/a-0853-1821**(Epub 2019 May 28)**
35. Cantisani V, Consorti F, Guerrisi A, Guerrisi I, Ricci P, Di Segni M, Mancuso E, Scardella L, Milazzo F, D'Ambrosio F, Antonaci A (2013) Prospective comparative evaluation of quantitative-elastosonography (Q-elastography) and contrast-enhanced ultrasound for the evaluation of thyroid nodules: preliminary experience. Eur J Radiol 82(11):1892–1898. https://doi.org/10.1016/j.ejrad.2013.07.005**(Epub 2013 Aug 6)**
36. Sollini M, Cozzi L, Chiti A et al (2018) Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: where do we stand? Eur J Radiol 99:1–8. https://doi.org/10.1016/j.ejrad.2017.12.004
37. Gao L, Liu R, Jiang Y et al (2018) Computer-aided system for diagnosing thyroid nodules on ultrasound: a comparison with radiologist-based clinical assessments. Head Neck 40:778–783. https://doi.org/10.1002/hed.25049
38. Yoo YJ, Ha EJ, Cho YJ et al (2018) Computer-aided diagnosis of thyroid nodules via ultrasonography: initial clinical experience. Korean J Radiol 19:665–672. https://doi.org/10.3348/kjr.2018.19.4.665
39. Jeong EY, Kim HL, Ha EJ et al (2019) Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. Eur Radiol 29:1978–1985. https://doi.org/10.1007/s00330-018-5772-9
40. Zhao WJ, Fu LR, Huang ZM et al (2019) Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: a systematic review and meta-analysis. Medicine 98:e16379. https://doi.org/10.1097/md.0000000000016379
41. Kim HL, Ha EJ, Han M (2019) Real-world performance of computer-aided diagnosis system for thyroid nodules using ultrasonography. Ultrasound Med Biol. https://doi.org/10.1016/j.ultrasmedbio.2019.05.032
42. Grani G, Lamartina L, Biffoni M et al (2018) Sonographically estimated risks of malignancy for thyroid nodules computed with five standard classification systems: changes over time and their relation to malignancy. Thyroid 28:1190–1197. https://doi.org/10.1089/thy.2018.0178
43. Falcone R, Ramundo V, Lamartina L et al (2018) Sonographic presentation of metastases to the thyroid gland: a case series. J Endocr Soc 2:855–859. https://doi.org/10.1210/js.2018-00124