THYROID NODULE ASSESSMENT: INTEROBSERVER AGREEMENT OF ULTRASOUND FEATURES AND ACR THYROID IMAGING REPORTING AND DATA SYSTEM SCORES

Ahmed S. Abdelrahman, Mona A. Nagi, Emad H. Abdeldayem and Mai M. K. Barakat

Department of Radiology , Faculty of Medicine, Ain Shams University, Cairo, Egypt.

Corresponding author:
Mai M. K. Barakat
Mobile: +2 01006294780
E-mail:
mai_mokhtar@med.asu.edu.eg

Received: 17/03/2024
Accepted: 20/05/2024

Online ISSN: 2735-3540

ABSTRACT:

Background: *The identification of the nature of the thyroid nodules is mandatory to avoid unnecessary thyroidectomy for the benign nodule and to decrease no of biopsies. TI-RADS scoring helps radiologists in characterization of thyroid nodules and allows simple communication between radiologists, patients, and physicians.*

Purpose*: To assess ultrasound interobserver variability of (ACR) (TIRADS) categories and final scores.*

Patients and methods: *In this study, between June 2019 and July 2022, 553 thyroid nodules in 410 thyroid ultrasound examinations were retrospectively evaluated using the TIRADS features. The definitive diagnosis of benign nodules was based on the fine-needle aspiration or surgical histopathological confirmation. Whereas, the definitive diagnosis of malignant thyroid nodules was based on surgical histopathological confirmation. Additionally, receiver operating curves (ROC) were drawn to compare the actual performance of the observers, and their agreement was tested using Cohen's kappa. Also, intraclass correlation (ICC) was used to determine the interobserver agreement.*

Results: *Out of 553 thyroid nodules, 452 (81.7%) were benign and 101/553 (18.3%) were malignant. The diagnosis of benign nodules was confirmed by FNA in 324/452 (71.7%) thyroid nodules (Bethesda II) and histopathological analysis in the remaining 128/452 (28.3%) thyroid nodules. Of 101 malignant thyroid nodules that were confirmed by histopathological examination. The majority of benign nodules and TR1 were Bethesda II, whereas Bethesda VI was the most common FNA finding in malignant thyroid nodules and TR5. The interobserver agreement of composition, echogenicity, peripheral calcification, macrocalcification and shape was substantial (k = 0.752, 0.677, 0.727, 0.708, and 0.603 respectively). The agreement of margin, large comet tail, and punctate foci was moderate (k = 0.523, 0.582, 0.484 respectively). There was moderate to almost perfect interobserver agreement for each TIRADS score (K= 0.808, 0.742, 0.591, 0.631, 0.624 for TR1, TR2, TR3, TR4, and TR5 respectively). A substantial agreement was noted for the final TI-RADS score (K= 0.662; 95%CI: 0.614-0.709).*

Conclusion: *ACR TIRADS scoring system has high diagnostic performance. The interobserver agreement of TIRADS categories is moderate to substantial and substantial to almost perfect for various TIRADS scores.*

***Keywords**:Ultrasound, Thyroid Imaging Reporting and Data System. Interobserver agreement, American College of Radiology, (ACR), thyroid nodules.*

1. INTRODUCTION:

Thyroid nodules are frequent imaging findings during the neck examination [(1)], with estimated prevalence by ultrasound ranging from 20% to 76% in adults [(2)]. The ultrasound examination of the thyroid gland is considered the best 1st diagnostic method for the evaluation of thyroid nodular disease [(3&4)]. Many classification systems were constructed by the Endocrine and Radiological Societies, These scoring systems aimed to provide an accurate selection criterion for the lesion that carries a higher risk of malignancy, and hence a justified reason for fine-needle aspiration (FNA) [(5)].The (ACR) released the ultrasound lexicon white paper for the Thyroid Imaging, Reporting and Data System in 2017, which is a system with a standard reporting used to guide the FNA according to the US suspicious sonographic features and the size of the nodule [(6)]. This reporting system must be a continuous dynamic process submitted for further evaluation and development according to the feedback from clinicians, radiologists, and validation of data [(7-11)].

The identification of the nature of the thyroid nodules is an important step that determines the further management algorithm for malignant nodules and avoids unnecessary thyroidectomy for the benign nodule. The diagnostic efficiency of TIRADS is based on the ultrasound lexicon assessed by radiologists, so determining the lexicon with the highest disagreement requires a thorough examination of interobserver variability. As a result, additional efforts should be made to improve TIRADS standardization.

AIM OF THE WORK:

The main purpose of the study was to evaluate the interobserver agreement of the TIRADS categories and scores.

2. METHODS:

2.1. Patients:

This was a retrospective study acquired by retrieval of data in a single center from the medical records and the PACS from June 2019 to July 2022. The study included 410 patients with 553 thyroid nodules. The institutional ethics committee approved our study protocol and waived the need for any informed consent. Patients who had thyroid nodules over the age of 18 identified during an ultrasound examination performed and whose final thyroid nodule diagnosis had been verified by FNA or histopathology were included. We excluded patients with nodules not verified by FNA or histopathology, with nodules categorized as TR1, with small nodules less than 10 mm, and with nodules of indeterminate pathology. We also excluded prior surgical thyroid intervention.Thyroid nodules less than 5 mm were excluded from the study.

2.2. Imaging Analysis:

Ultrasound of the neck was performed using (Logiq E9 Xdclear 2.0, General Electric Healthcare, USA) and a 9L linear-array probe (9 MHz). Each nodule was imaged in longitudinal and transverse scans, with a more scan to evaluate the nodule's specific characteristics (position (left lobe, right lobe, or isthmus), the maximum diameter, composition, echogenicity, shape, margin and calcification/echogenic foci). The real-time ultrasound images were then converted into bitmap (BMP) format and saved in a separate file for each nodule. Afterward, two distinct sets were created, each containing the BMP files of all nodules but in a different order. Finally, two radiologists with 5 and 4 years of thyroid ultrasound experience (reader 1 and 2, respectively) and three years of reporting thyroid nodules using ACR TIRADS independently examined one set that included all nodule BMP files. Doppler evaluation of thyroid nodules and suspicious neck lymph nodes evaluation were not evaluated in the current study.

Each radiologist assigned points to each thyroid nodule for the separate five categories of composition, echogenicity, shape, margin, and calcification/echogenic foci, according to the TI-RADS protocols produced by ACR [(12)]. The sum of the points in each category indicated the TI-RADS category specified to each nodule, with 2 points denoting TR2 (not suspicious); 3 points, TR3 (mildly suspicious); 4–6 points, TR4 (moderately suspicious); and 7 or higher points, TR5 (highly suspicious). As per the instructions of the ACR TI-RADS committee, if margins, echogenicity, or composition could not be determined for any cause, they were assigned 0, 1, or 2 points, respectively.

Each nodule was evaluated according to a template filled by each reader for the longest diameter of the thyroid nodule, margin, composition, shape, ecopattern, and calcification or echogenic foci according to standard TIRADS lexicon feature, eventually, points were assigned to each category [(6,13)]. Composition: cystic or spongiform nodule with more than fifty percent of its volume composed of small cystic areas equal (0 points), mixed cystic areas and solid areas in nodule equal (1 point), predominantly solid nodule and nodule with a composition that can’t be assigned (2 points). Echogenicity in comparison to nearby normal thyroid parenchyma: anechoic nodule which was cystic (0 points), isoechoic or hyperechoic nodule, and if the echogenicity can`t be determined (1 point), while hypoechoic nodule equal (2 points), but very hypoechoic which was also hypoechoic compared to the strap muscle (3 points). The shape was assessed in the transverse image: wider than taller equal (0 points) and taller than wider equal (3 points). Margin: if smooth or can`t be determined equal (0 points), lobulated or irregular with spiculated or sharp angles (2 points), and beyond thyroid any extension (3 points). Calcification or any echogenic foci: large V-shaped comet-tail artifacts larger than one mm or none (0 points), macrocalcifications with posterior shadowing (1 point), peripheral calcification along nodule margin (2 points), and punctate tiny echogenic foci which show tiny (less than one mm) comet-tail artifacts (3 points). The solid part of the mixed cystic and partially solid nodule was also assessed and gained points according to the other TI-RADS features.

For the composition, echogenicity, shape, and margin categories lexicon, only one feature could be added, so the agreement was assessed for the category lexicon; however, for the calcification/echogenic foci category, more than one feature could be applied, so the agreement was assessed for each feature. Finally, the TIRADS (TR) score was given as follows; 0 points equal (TR1) Figure (1)., 2 points equal (TR2) Figure (2)., 3 points equal (TR3), 4 to 6 points equal (TR4) Figure (3)., and 7 points or more equal (TR5) Figure (4&5).

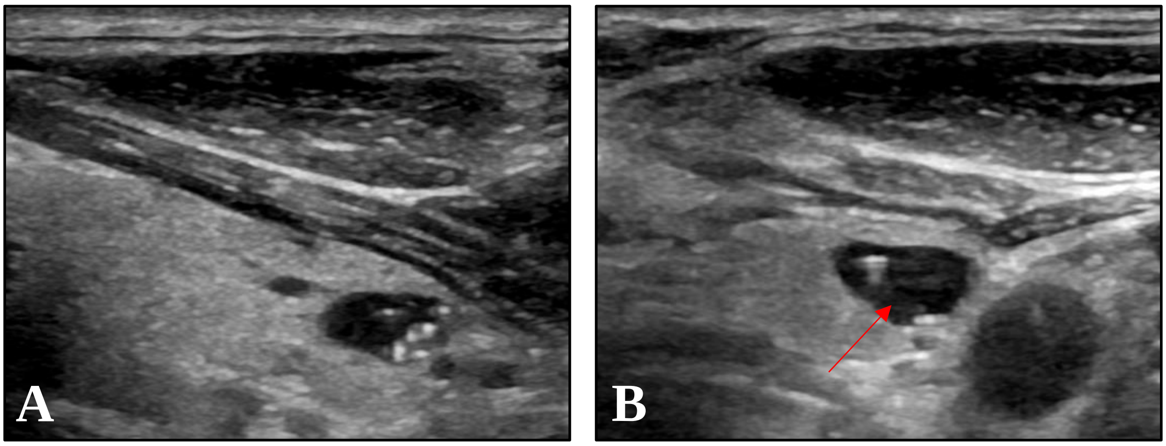

Figure 1: Thirty-five-year-old female patient with a left thyroid nodule. FNA revealed a benign nodule (Bethesda II). Longitudinal (A) and transverse (B) ultrasound images in nodule with a perfect interobserver agreement, which was interpreted by two observers as a colloid cyst with comet tail artifacts (red arrow) (total points = 0, TR = 1).

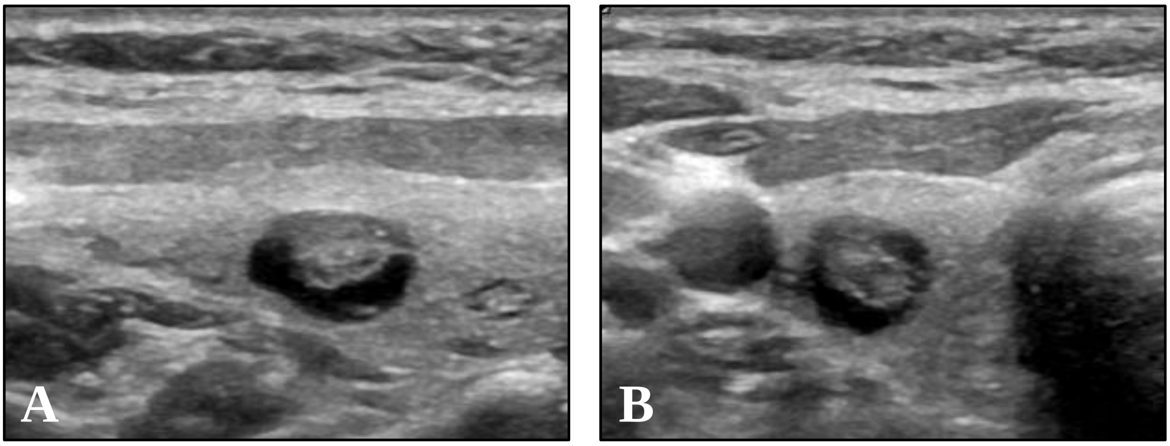

Figure 2: A Thirty-two-year-old male patient with a right thyroid nodule. FNA revealed a benign nodule (Bethesda II). Longitudinal (A) and transverse (B) ultrasound images in nodule with a perfect interobserver agreement, which was interpreted by two observers as mixed cyst/solid nodule, the solid nodule is isoechoic wider than tall, with well-defined margin and no calcification (total points = 2, TR = 2).

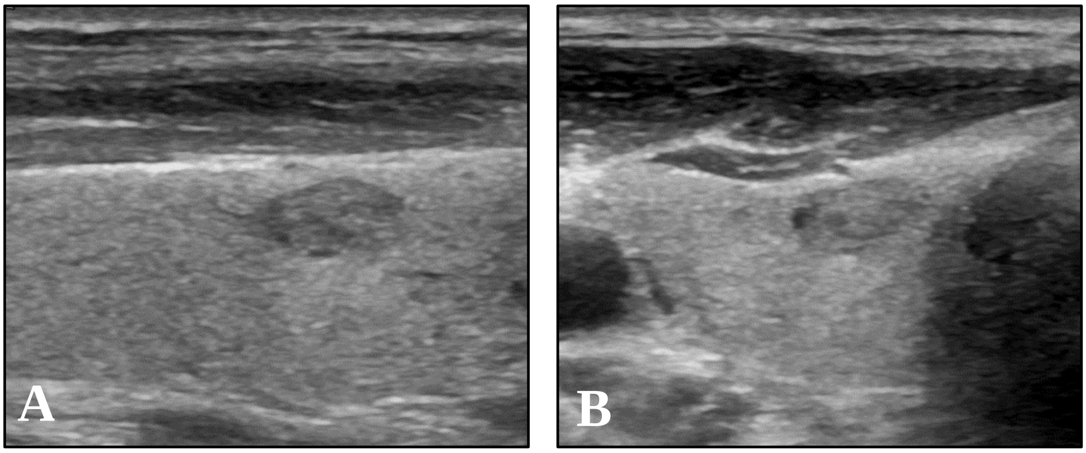

Figure 3: Twenty-eight-year-old female patient with a right thyroid nodule. Histopathology revealed follicular adenoma. Longitudinal (A) and transverse (B) ultrasound images in nodule with a poor interobserver agreement for margin. The nodule was interpreted by two observers as an isoechoic wider than tall solid nodule with no calcification, yet the first observer interpreted the margin as well-defined (total points = 3, TR = 3), and the second reader interpreted the medial and lateral margin of the nodule as irregular (total points = 5, TR = 4).

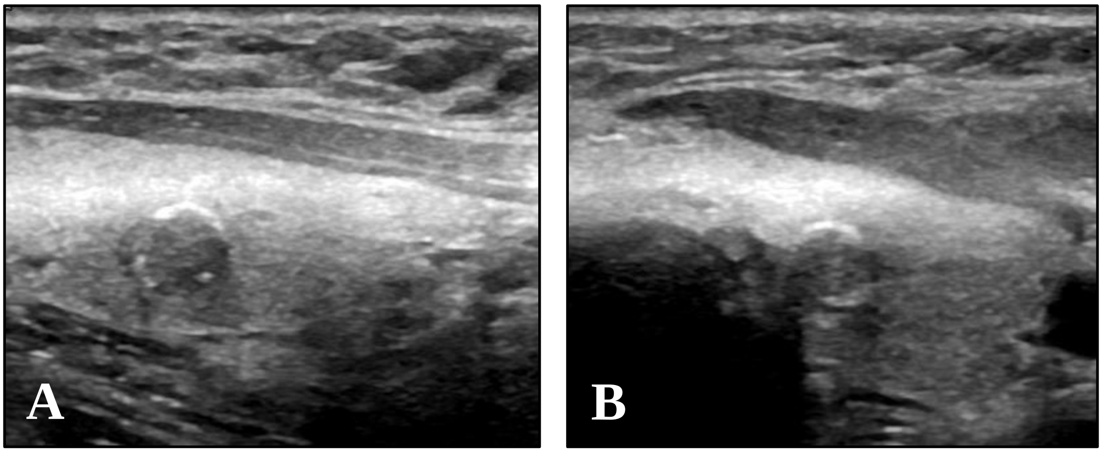

Figure 4: Twenty-seven-year-old female patient with a left thyroid nodule. Histopathology revealed follicular carcinoma. Longitudinal (A) and transverse (B) ultrasound images in nodule with a poor interobserver agreement for echogenicity and margin. The nodule was interpreted by two observers as wider than tall solid nodule with peripheral rim calcification, yet the first observer interpreted the echogenicity as hypoechoic and the margin as ill-defined (total points = 6, TR = 4), and the second reader interpreted the echogenicity as isoechoic and the margin as interrupted (irregular) (total points = 7, TR = 5).

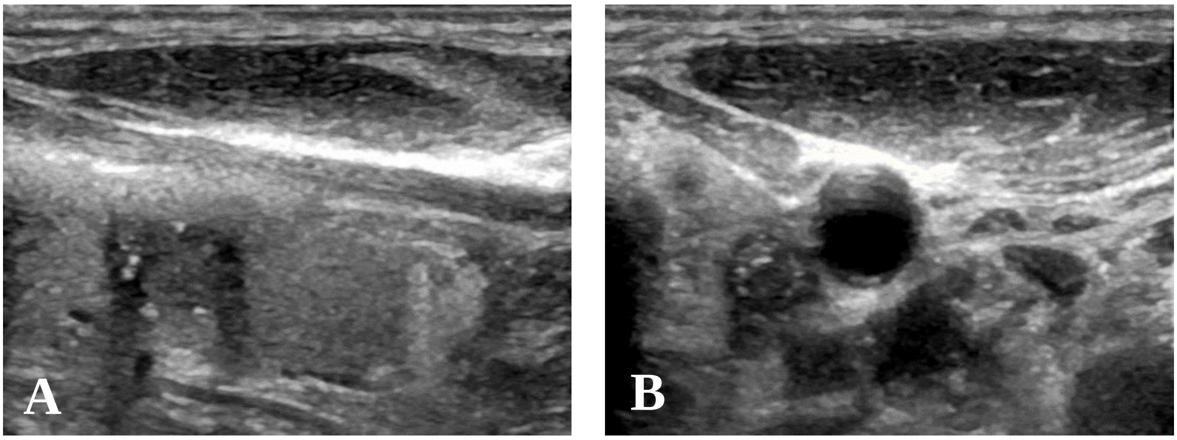

Figure 5: Thirty-eight-year-old female patient with a left thyroid nodule. Histopathology revealed papillary thyroid carcinoma. Longitudinal (A) and transverse (B) ultrasound images in nodule with a poor interobserver agreement for echogenicity and calcification. The nodule was interpreted by two observers as irregular taller than wide solid, yet the first observer interpreted the echogenicity as isoechoic and the calcification as macrocalcification (total points = 9, TR = 5), and the second reader interpreted the echogenicity as hypoechoic and the calcification as punctate echogenic foci (total points = 12, TR = 5).

2.3. Standard of reference

The definitive diagnosis of benign nodules was based on the fine-needle aspiration or surgical histopathological confirmation. Whereas, the definitive diagnosis of malignant thyroid nodules was based on surgical histopathological confirmation. FNA Bethesda category II with no new noticed suspicious features or marked increase in nodule size (50% or more increase in volume of nodule) with postoperative histopathological confirmation for benign pathology were criteria for benign nodules. Postoperative histopathological examination was obtained for our Bethesda categories III, IV, V, and VI to finally evaluate whether the thyroid nodule was malignant or benign. The thyroid nodules were classified according to the criteria recommended by the Bethesda classification [14]

2.4. Statistical analysis:

We analyzed the data by using the Statistical package of social science (IBM SPSS statistics, V. 22.0, Armonk, NY, USA). The mean and range of our quantitative data were expressed, as well as the percentage and also the number of the data category. The actual diagnostic performance of the evaluated categorical variable was done using cross-tabulation, the area under the curve (AUC) was performed to compare the diagnostic actual performance of both observers, and then the sensitivity, specificity, and accuracy were all calculated. The interobserver agreement of the evaluated categorical all variables was determined using Cohen's kappa coefficient (K), and the quadratic weighted kappa (for the final TI-RADS score). Intraclass correlation (ICC) was used to determine the interobserver agreement of nodule size. The percent agreement and 95% confidence interval were estimated. *P* value <0.05 was considered statistically significant. Interpretation of Kappa agreement was done as follows: slight (0.01–0.20), fair (0.21– 0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–0.99). interpretation of ICC was done as follows: poor (values less than 0.5), moderate (values 0.5 to 0.75), good (values 0.75 to 0.9), and excellent (values more than 0.90).

Ethics approval and consent to participate:

The study was approved by the Ethics Committee as a retrospective study at the Faculty of Medicine at Ain Shams University, Egypt (FWA 000017585) in August 2020; Reference Number: R 44 / 2020, the committee waived the written informed consent.

3. RESULTS:

The distribution of Bethesda categories according to the final diagnosis and diagnostic performance of TI-RADS score:

Our study included 410 patients with 553 thyroid nodules, 251/410 (61.2%) were females and 159/410 (38.8%) were males with patient mean age of 41.5±11.1 (SD) years (range: 21-66 years). Out of 553 thyroid nodules, 452 (81.7%) were benign and 101/553 (18.3%) were malignant. The majority of benign nodules and TR1 were Bethesda II, whereas Bethesda VI was the most common FNA finding in malignant thyroid nodules and TR5 Table (1).

Table 1: Distribution of Bethesda categories according to the final diagnosis and TI-RADS score.

	Bethesda	II+	II*	II (total)	III*	IV*	V*	VI*
		No (%)	No (%)	No (%)	No (%)	No (%)	No (%)	No (%)
Diagnosis	Benign	264 (264/264; 100%)	145 (145/149; 97.3%)	409 (409/413; 99%)	19 (19/26; 73.1%)	18 (18/26; 69.2%)	5 (5/24; 20.8%)	1 (1/64; 1.6%)
	Malignant	0 (0%)	4 (4/149; 2.7%)	4 (4/413; 1%)	7 (7/26; 26.9%)	8 (8/26; 30.8%)	19 (19/24; 79.2%)	63 (63/64; 98.4%)

Qualitative data are expressed as raw numbers followed by proportions and percentages in parentheses. TI-RADS = Thyroid imaging reporting and data system
+ No histopathological confirmation
* Histopathologically confirmed

Thyroid cancer prevalence in different TI-RADS scores and the diagnostic performance of TI-RADS among both observers:

Both observers showed a statistically significant trend of increased noticed risk of malignancy from TR1 to TR5 (P value < 0.001). Observer 1 had a risk of 0%, 3.2%, 5.8%, 14.2%, and 71.1% of a malignant thyroid tumor among TR1, TR2, TR3, TR4, and TR5 nodules respectively, while Observer 2 had a risk of 0%, 0%, 2.9%, 14.5%, 68.1% respectively. Although there was no significant noticed difference in the diagnostic performance of observers 1 and 2 for detecting malignant thyroid nodules (AUC = 0.776 and 0.810 respectively, P value

= 0.096), observer 2 had a higher sensitivity (96%; 95% CI: 90-99%) and accuracy (71.4%; 95% CI: 67-75%) than observer 1 (88.1%; 95% CI: 80-94% and 70.9%; 95% CI: 67-75% respectively), on the other hand, observer 1 had higher specificity (67%; 95% CI: 62.71%) than observer 2 (65.9%; 95% CI: 61-70%).

Thyroid imaging reporting and data system categories among both observers:

The univariate analysis of the TI-RADS categories lexicon features was demonstrated in Table (2). The solid composition was the most frequently encountered feature by both observers in the composition lexicon category, accounting for 371/553 (67.1%) for observer 1 and 375/553 (67.8%) for observer 2. The isoechoic or hyperechoic nodules were the most frequently detected nodules by observer 1 (375/553; 67.8%) and observer 2 (355/553; 64.2%). Nodules with wider shapes were detected in 476/553 (86.1%) and 492/553 (89%) by observers 1 and 2 respectively. Nodules with smooth margins were reported in 445/553 (80.5%) and 441/553 (79.7%) by observers 1 and 2 respectively, on the other hand, the nodules with lobular or irregular margins were noted by observers 1 and 2 in 108/553 (19.5%) and 112/553 (20.3%) respectively. As regards the detection of calcification/echogenic foci, the most frequently matched cases were observed in nodules with absent calcification (329/553; 59.5%), followed by macrocalcification (39/553; 7.1%) and punctate echogenic foci (33/553; 6%) and lastly the large comet-tail artifacts (29/553; 5.2%) and the peripheral calcification (23/553; 4.2%).

The interobserver agreement of TI-RADS categories:

A substantial interobserver agreement was noticed for the composition (K=0.752, 95% CI: 0.695-0.808, 88% percent agreement), echogenicity (K=0.677, 95% CI: 0.616-0.737, 83.1% percent agreement), and shape (K=0.603, 95% CI: 0.501-0.704, 91.3% percent agreement). On the other hand, a moderate agreement was noticed for margin (K=0.523, 95% CI: 0.434-0.611, 84.8% percent agreement) Table (2).

Table 2: Thyroid imaging reporting and data system categories among both observers.

TI-RADS categories	Observer 1	Observer 2	Matched cases
Composition			
Cystic/Spongiform	53 (9.6%)	44 (8%)	40 (7.2%)
Mixed solid and cystic	129 (23.3%)	134 (24.2%)	105 (19%)
Solid	371 (67.1%)	375 (67.8%)	342 (61.8%)
Echogenicity			
Anechoic	53 (9.6%)	44 (8%)	40 (7.2%)
Isoechoic or Hyperechoic	375 (67.8%)	355 (64.2%)	323 (58.4%)
Hypoechoic	84 (15.2%)	113 (20.4%)	60 (10.8%)
Very Hypoechoic	41 (7.4%)	41 (7.2%)	37 (6.7%)
Shape			
Wider	476 (86.1%)	492 (89%)	460 (83.2%)
Taller	77 (13.9%)	61 (11%)	45 (8.1%)
Margin			
Smooth	445 (80.5%)	441 (79.7%)	401 (72.5%)
Lobular or Irregular	108 (19.5%)	112 (20.3%)	68 (12.3%)
Calcification/echogenic foci			
Absent calcification/echogenic foci	361 (65.3%)	361 (65.3%)	329 (59.5%)
Large comet-tail artifacts	41 (7.4%)	53 (9.6%)	29 (5.2%)
Macrocalcification	47 (8.5%)	59 (10.7%)	39 (7.1%)
Peripheral calcification	27 (4.9%)	35 (6.3%)	23 (4.2%)
Punctate echogenic foci	65 (11.8%)	57 (10.3%)	33 (6%)

TI-RADS = Thyroid imaging reporting and data system

As regards the calcification/echogenic foci, a substantial interobserver agreement was recorded for absent calcification, peripheral rime calcification, and macrocalcification with K= 0.745, 0.727, and 0.708 respectively; 95% CI: 0.686-0.803, 0.599-0.854 and 0.606-0.809 respectively, and percent agreement of 88.4%, 97.1%, and 95% respectively. On the other hand, a moderate agreement was noticed for large comet tail artifacts and punctate foci lexicon with K= 0.582 and 0.484 respectively; 95% CI: 0.458-0.705 and 0.368-0.599 respectively, and percent agreement of 93.4% and 89.9% respectively Table (3).

Table 3: Interobserver agreement of the thyroid imaging reporting and data system composition, echogenicity, shape, and margin categories.

TI-RADS categories	Kappa	95% CI	*P* value	Percent of agreement
Composition	0.752	0.695-0.808	<0.001*	88%
Echogenicity	0.677	0.616-0.737	<0.001*	83.1%
Shape	0.603	0.501-0.704	<0.001*	91.3%
Margin	0.523	0.434-0.611	<0.001*	84.8%

TI-RADS = Thyroid imaging reporting and data system, * Significant

The interobserver agreement of TI-RADS score:

A substantial interobserver agreement was found for the final TI-RADS score with K= 0.662, (95% CI: 0.614-0.709, 73.6% percent agreement), on the other hand, the quadratic weighted Kappa showed a higher yet still substantial agreement (K=0.760, 95% CI: 0.710-0.809). The interobserver agreement for the total points was also substantial with K= 0.588 (95% CI: 0.542-0.633) and a percent agreement of 64.9% Table (4).

Table 4: Interobserver agreement of the thyroid imaging reporting and data system calcification/echogenic foci category.

Calcification/echogenic foci category	Kappa	95% CI	*P* value	Percent of agreement
No calcification/echogenic foci	0.745	0.686-0.803	<0.001*	88.4%
Peripheral rime calcification	0.727	0.599-0.854	<0.001*	97.1%
Macrocalcification	0.708	0.606-0.809	<0.001*	95%
Large comet tail artifact	0.582	0.458-0.705	<0.001*	93.4%
Punctate echogenic foci	0.484	0.368-0.599	<0.001*	89.9%

TI-RADS = Thyroid imaging reporting and data system, * Significant

As regards, the agreement of each TI-RADS score, an almost perfect agreement was observed for the TR1 score (K=0.808, 95% CI: 0.719-0.896, 96.9% percent agreement), yet the lowest agreement was noted for TR3 which revealed a moderate agreement with K= 0.591 (95% CI: 0.512-0.669, 84.7% percent agreement). TR2, TR4, and TR5 also revealed a substantial agreement with K= 0.742, 0.631, and 0.624 respectively Table (5).

Table 5: Interobserver agreement of the total point score and the final thyroid imaging reporting and data system score.

TI-RADS	Kappa	95% CI	*P* value	Percent of agreement
Total Point score	0.588	0.542-0.633	<0.001*	64.9%
TI-RADS+	0.662	0.614-0.709	<0.001*	73.6%
TI-RADS©	0.760	0.710-0.809	<0.001*	

TI-RADS = Thyroid imaging reporting and data system, * Significant, +Kappa, ©Quadratic weighted kappa

As regards, the agreement of each TI-RADS score, an almost perfect agreement was observed for the TR1 score (K=0.808, 95% CI: 0.719-0.896, 96.9% percent agreement), yet the lowest agreement was noted for TR3 which revealed a moderate agreement with K= 0.591 (95% CI: 0.512-0.669, 84.7% percent agreement). TR2, TR4, and TR5 also revealed a substantial agreement with K= 0.742, 0.631, and 0.624 respectively Table (6).

Table 6: Interobserver agreement of each thyroid imaging reporting and data system score.

TI-RADS	Observer 1	Observer 2	Matched cases	Kappa	95% CI	*P* value	Percent of agreement
TI-RADS 1	53 (9.6%)	44 (8%)	40 (7.2%)	0.808	0.719-0.896	<0.001*	96.9%
TI-RADS 2	125 (22.6%)	118 (21.3%)	97 (17.5%)	0.742	0.673-0.810	<0.001*	91.1%
TI-RADS 3	137 (24.8%)	140 (25.3%)	96 (17.4%)	0.591	0.512-0.669	<0.001*	84.7%
TI-RADS 4	141 (25.5%)	138 (25%)	101 (18.3%)	0.631	0.556-0.705	<0.001*	86.1%
TI-RADS 5	97 (17.5%)	113 (20.4%)	73 (13.2%)	0.624	0.539-0.708	<0.001*	88.4%

TI-RADS = Thyroid imaging reporting and data system, * Significant.

The histopathological results:

The diagnosis of benign nodules was confirmed by FNA in 324/452 (71.7%) thyroid nodules (Bethesda II) and histopathological analysis in the remaining 128/452 (28.3%) thyroid nodules (46 follicular adenomas, 34 colloid nodules, 31 adenomatoid nodules, and 17 lymphocytic thyroiditis). Of 101 malignant thyroid nodules that were confirmed by histopathological examination, 61 (60.4%) were papillary thyroid carcinoma, 24 (23.8%) were follicular variant-papillary carcinoma, 9 (8.9%) follicular carcinoma, 4 (3.9%) papillary microcarcinoma, and 3 (3%) follicular neoplasm with Hurthle cell features).

4. DISCUSSION:

The use of high-resolution neck ultrasound examinations increases the prevalence of discovering nodules. The FNA was deemed the most useful test to determine the nature of nodules before any surgical maneuver, reducing the number of unnecessary thyroidectomies. Several classification systems based on suspicious ultrasound features and an additional FNA recommendation were proposed [13].The TIRADS is considered a simple and practical process and in routine ultrasound practice, it has excellent diagnostic precision in the diagnosis of malignant nodules [13]. The new era of radiological reporting is the standardization of the radiological interpretation with an organized template, intending to provide a precise response to the clinician issue and sharing in the patient care by providing the proper next step in management according to the evaluated suspicious radiological features [7].

The variability in the characterization of ultrasound lexicon features influences the performance of the TIRADS and the further algorithm recommendation. The main purpose of the study was to assess the reliability and the interobserver agreement of each ACR TI-RADS lexicon and the final TIRADS score, and it did not focus on nodules with a high malignancy suspicion but instead included all types of nodules seen in everyday practice [7]. In our retrospective study, two radiologists evaluated 553 thyroid nodules using the ACR TIRADS, and they found moderate to substantial agreement for the different TI-RADS categories and substantial to almost perfect agreement for TI-RADS score with sensitivity, specificity, and accuracy of 96%, 67%, and 71.4% respectively.

Our study revealed a sensitivity of 96% and a specificity of 67% for diagnosing malignant nodules. *Lim-Dunham et al.* [15] found a comparable result, but *Schenke et al.* [16] reported a higher sensitivity (100%)

and lower specificity (40.6%) in their study, which included only small thyroid nodules (10 mm).However, *Basha et al.* reported a comparable sensitivity 98.3%,higher specificity 90.9% [17]

The current study revealed a substantial noticed agreement for the composition category, another study done by *Phuttharak et al.* [2] also revealed a substantial agreement but with a lower value (K= 0.616) between two observers who have 2 and 10 years of experience in thyroid ultrasound, while other studies with more than two observers found moderate agreement. The interobserver agreement between 6 observers was 0.53 and 0.57 in two other studies conducted by *Persichetti et al.* [5] and *Hoang et al.* [4]. A similar agreement was also noted between 3 observers in a study conducted by *Sahli et al.* [18]. Although the solid composition has traditionally been thought to be a criterion for malignancy [19], *Iannuccilli et al.* [20] found that the solid composition is reported in 60% to 83% of benign nodules. The other sonographic features of the solid part of the mixed partially solid/cystic nodules determine the level of suspicion and whether a biopsy is recommended [21&22]. Although cystic nodules are typically benign, cystic papillary carcinoma was defined as a cystic nodule with solid excrescences and echogenic foci seen inside [23].

As regards echogenicity, marked hypoechogenicity was considered a more specific criterion for malignant nodules. The hypoechoic nodules were seen in most malignant nodules and about half of the benign nodules [24&25]. Similar to our study a substantial agreement for echogenicity was noted between two radiologists who completed 4 weeks of training sessions in a study done by *Grani et al.* [1]. On the other hand, different studies that did not provide training before TIRADS thyroid nodule evaluation reported fair to moderate agreement [2&4&5&18].

In the current study, a substantial agreement was noted for the shape item, lower results were noted in studies done by *Sahli et al.* [18] and *Persichetti et al.* [5] who revealed an agreement of 0.36 and 0.47 respectively. The taller shape is specific yet insensitive to malignancy [26].

Previous studies have shown a heterogeneous agreement value for the margin lexicon category, with agreement ranging from 0.14 to 0.51 [1&2&27]. Irregular and spiculated margin is more frequently noted in the malignant nodules due to infiltrative growth patterns [28]. In the current study, a moderate agreement was noted for the margin evaluation, the interobserver agreement could be improved with more concise definitions and focused education.

The current study showed a substantial agreement for absent calcification, macrocalcification, and peripheral rim calcification. *Grani et al.* [1] and *Liu et al.* [3] studies also revealed an acceptable agreement for both calcification types, yet other studies have reported moderate agreement [4&27]. The presence of rim calcification may suggest malignant nodules specifically if an interruption in the peripheral calcification was noted [29]. Determining the nodule margin with peripheral calcification is a critical step in thyroid ultrasonography evaluation.

The presence of echogenic punctate foci in a solid nodule increases the risk of thyroid cancer by three to four-fold, they have a high predictive value yet with less sensitivity for papillary carcinoma [30]. Previous studies have found a fair agreement noticed for punctate foci [4&27], yet the current study showed a moderate agreement. A similar result was also found in *Persichetti et al.* [5] study which found a mean agreement of 0.47. The agreement of large comet-tail artifacts in our study was slightly higher than that of punctate echogenic foci, one potential cause of this variability is the similar appearance of the echogenic punctate foci and those with comet-tail artifacts in solid nodules especially if the posterior artifacts were small.

In terms of each TIRADS score, TR1 had the highest agreement, similar findings were also published by *Phuttharak et al.* [2] study, yet their study revealed slight to fair agreement for TR4 and TR5 in contrast to our study which revealed substantial agreement for TR4 and TR5. The nodules with a straightforward benign nature were easily interpreted, but those with more suspicious features were more susceptible to misinterpretation.

The final TIRADS score had a substantial agreement, which impacted the management algorithm. *Basha et al.* [17] also reported good inter-observer agreement in the final TI-RADS category. The standard reporting algorithm provides standard reports which are reliable and detailed, it also provides a practical defined language between clinicians, radiologists, and patients, and promotes cooperation between them. Continuous fine-tuning of the radiologist's knowledge, skills, and experience will maximize the benefit sought and determine the proper management of each patient.

The current study has some limitations. First, each category's features were distributed and reflected unequally. Second, although the image quality perceived by each radiologic was standardized in this retrospective study by interpreting the ultrasound images performed by one radiologist, the skills associated with conducting the US technique and determining the probe frequency and parameter used during the US examination were eliminated. The substantial agreement of the final TIRADS score, however, led to the conclusion that thyroid nodule features interpretation and TIRADS scoring on an image set are reproducible. Finally, the variability of other techniques such as thyroid elastography and the color Doppler was not evaluated; the ACR TIRADS did not include these techniques, although they are commonly used in regular thyroid imaging reporting.

5. Conclusions :

In conclusion, the overall TIRADS final score and category are in good agreement leading to uniformity of management with FNA according to TIRADS recommendations.

6. Recommendations

We recommend further studies with more patients' number with decreased range of age of the included patients to minimize the difference in results.

7. Declarations:

Consent for publication :

not applicable

Availability of data and materials:

All data are available upon reasonable request from the authors

Competing interests

Declaration of no competing interest was done by all authors

Funding:

No Funds, no financial support are to be disclosed

Authors contributions:

AS: Data curation, Supervision, methods, Revising the manuscript.

MB: Data collection and analysis, sharing in statistical analysis, sharing in manuscript writing, revising and editing, preparing figures and tables.

M.N: Reader 1, Visualization, Final draft approval., writing draft.

E.A: Reader 2, Data curation, Writing draft, Final draft approval.

Approval of the final manuscript by all authors was done.

Acknowledgements

not applicable

8. List of abbreviations:

US : Ultrasound
FNA : Fine-needle aspiration
ACR : American College of Radiology
TI-RADS: Thyroid Imaging, Reporting and Data System
BMP : Bitmap
TR : TIRADS
K : Cohen's kappa coefficient
ICC : Intraclass correlation
AUC : Area under the curve
CI : Confidence interval.
SD : Standard deviation

9. REFERENCES

1. **Grani G, Lamartina L, Cantisani V, Maranghi M et al.** Interobserver agreement of various thyroid imaging reporting and data systems. Endocr Connect. 2018;(7):1-7.
2. **Phuttharak W, Boonrod A, Klungboonkrong V, et al.** Interrater Reliability of Various Thyroid Imaging Reporting and Data System (TIRADS) Classifications for Differentiating Benign from Malignant Thyroid Nodules. Asian Pac J Cancer Prev. 2019;(20):1283-8.
3. **Liu H, Ma AL, Zhou YS, et al.** Variability in the interpretation of grey-scale ultrasound features in assessing thyroid nodules: A systematic review and meta-analysis. Eur. J. Radiol. 2021;(129):1090-50.
4. **Hoang JK, Middleton WD, Farjat AE, et al.** Interobserver Variability of Sonographic Features Used in the American College of Radiology Thyroid Imaging Reporting and Data System. AJR Am J Roentgenol. 2018;(211):162-7.
5. **Persichetti A, Di Stasio E, Coccaro C, et al.** Inter- and Intraobserver Agreement in the Assessment of Thyroid Nodule Ultrasound Features and Classification Systems: A Blinded Multicenter Study. Thyroid. 2020;(30): 237-42.
6. **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.
7. **Abdelrahman AS, Ashour MMM, Abdelaziz TT.** Predictive value of neck imaging reporting and data system (NIRADS) in CECT/CEMRI of laryngeal and oral cavity squamous cell carcinoma. Egypt J Radiol Nucl Med. 2020;(51):241-43.
8. **Abdelaziz TT, Abdel Razk AAK, Ashour MMM, et al.** Interreader reproducibility of the neck imaging reporting and data system (NI-RADS) lexicon for the detection of residual/recurrent disease in treated head and neck squamous cell carcinoma (HNSCC). Cancer Imaging. 2020;(20):61.
9. **Abdelrahman AS, Abdel-Rahman, AS, Taha NM.** The role of functional imaging; DWI, ADC, and 18F-FDG PET/CT in the evaluation of HCC treatment response after transarterial chemoembolization. Egypt J Radiol Nucl Med. 2021;(52):217.
10. **Abdelrahman AS, Ekladious MEY, Badran EM, Madkour SS.** Liver imaging reporting and data system (LI-RADS) v2018: Reliability and agreement for assessing hepatocellular carcinoma locoregional treatment response. Diagn Interv Imaging. 2022;(103):524-34.
11. **Barakat, M.M.K., Mohamed, E.K.A.E., Habib, L.A.E. et al.** The diagnostic value of digital breast tomosynthesis with complementary ultrasound in comparison with magnetic resonance imaging in assessment of postoperative changes and locoregional recurrence of breast cancer. Egypt J Radiol Nucl Med. 2022; (**53)**:17 .
12. **FN Tessler, WD. Middleton, EG Grant, et al.,** ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee, J. Am. Coll. Radiol. 2017 ; (5) : 587–595.
13. **Azab EA, Abdelrahman AS, Ibrahim MEA.** A practical trial to use Thyroid Imaging Reporting and Data System (TI-RADS) in differentiation between benign and malignant thyroid nodules. Egypt J Radiol Nucl Med. 2019;(50):17.

14. **ES Cibas, SZ Ali**, The Bethesda System for reporting thyroid cytopathology, Thyroid 2009; (11):1159–1165.
15. **Lim-Dunham JE, Toslak IE, Reiter MP, et al.** Assessment of the American College of Radiology Thyroid Imaging Reporting and Data System for Thyroid Nodule Malignancy Risk Stratification in a Pediatric Population. AJR Am J Roentgenol. 2019;(212):188-94.
16. **Schenke S, Klett R, Seifert P, et al.** Diagnostic Performance of Different Thyroid Imaging Reporting and Data Systems (Kwak-TIRADS, EU-TIRADS and ACR TI-RADS) for Risk Stratification of Small Thyroid Nodules (≤10 mm). J Clin Med . 2020;(9):236.
17. **Basha MAA, Alnaggar AA, Refaat R, El-Maghraby AM, Refaat MM, Abd Elhamed ME, Abdalla AAEM, Aly SA, Hanafy AS, Mohamed AEM, Afifi AHM, Harb O.** The validity and reproducibility of the thyroid imaging reporting and data system (TI-RADS) in categorization of thyroid nodules: Multicentre prospective study. Eur J Radiol. 2019;(117):184-192.
18. **Sahli ZT, Sharma AK, Canner JK, et al.** TIRADS Interobserver Variability Among Indeterminate Thyroid Nodules: A Single-Institution Study. J Ultrasound Med . 2019;(38):1807-13.
19. **Frates MC, Benson CB, Charboneau JW, et al.** Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology. 2005;(237):794-800.
20. **Iannuccilli JD, Cronan JJ, Monchik JM.** Risk for malignancy of thyroid nodules as assessed by sonographic criteria: the need for biopsy. J Ultrasound Med. 2004;(23):1455-64.
21. **Park JM, Choi Y, Kwag HJ.** Partially cystic thyroid nodules: ultrasound findings of malignancy. Korean J Radiol .2012;13:530-5.
22. **Kim DW, Park JS, In HS, Choo HJ, et al.** Ultrasound-based diagnostic classification for solid and partially cystic thyroid nodules. AJNR Am J Neuroradiol. 2012;(33):1144-9.
23. **Hoang JK, Lee WK, Lee M, et al.** US Features of thyroid malignancy: pearls and pitfalls. Radiographics. 2007;(27):847-60.
24. **Moon WJ, Jung SL, Lee JH, et al;** Thyroid Study Group, Korean Society of Neuro- and Head and Neck Radiology. Benign and malignant thyroid nodules: US differentiation-multicenter retrospective study. Radiology . 2008;(247):762-70.
25. **Kim EK, Park CS, Chung WY, et al.** New sonographic criteria for recommending fine needle aspiration biopsy of nonpalpable solid nodules of the thyroid. AJR Am J Roentgenol. 2002;(178):687-91.
26. **Ren J, Liu B, Zhang LL, et al.** A taller-than-wide shape is a good predictor of papillary thyroid carcinoma in small solid nodules. J Ultrasound Med. 2015;(34):19-26.
27. **Itani M, Assaker R, Moshiri M, et al.** Inter-observer Variability in the American College of Radiology Thyroid Imaging Reporting and Data System: In-Depth Analysis and Areas for Improvement. Ultrasound Med Biol. 2019;(45):461-70.
28. **Reading CC, Charboneau JW, Hay ID, et al.** Sonography of thyroid nodules: a "classic pattern" diagnostic approach. Ultrasound Q . 2005;(21):157-65.
29. **Park YJ, Kim JA, Son EJ, et al.** Thyroid nodules with macrocalcification: sonographic findings predictive of malignancy. Yonsei Med J. 2014;(55):339-44.
30. **Malhi H, Beland MD, Cen SY, et al.** Echogenic foci in thyroid nodules: significance of posterior acoustic artifacts. AJR Am J Roentgenol . 2014;(203):1310-6.

تقييم عقيدات الغدة الدرقية: اتفاق بين المراقبين على ميزات الموجات فوق الصوتية ونتائج تقارير تصوير الغدة الدرقية ونظام البيانات ACR

احمد سامى عبد الرحمن ، منى على ناجي ، عماد حامد عبد الدايم , مي مختار كمال بركات

قسم الاشعه التشخيصيه و التداخليه - كليه الطب - جامعه عين شمس

الخلفية: تحديد طبيعة عقيدات الغدة الدرقية أمر إلزامي لتجنب استئصال الغدة الدرقية غير الضروري للعقيدات الحميدة وتقليل عدد الخزعات. يساعد تسجيل TI-RADS أطباء الأشعة في توصيف عقيدات الغدة الدرقية ويسمح بالتواصل البسيط بين أطباء الأشعة والمرضى والأطباء.

غاية: لتقييم تباين الموجات فوق الصوتية بين المراقبين لفئات (ACR) (TIRADS) والنتائج النهائية.

مرضى وطرق: في هذه الدراسة، بين يونيو 2019 ويوليو 2022، تم تقييم 553 عقيدة درقية في 410 فحصًا بالموجات فوق الصوتية للغدة الدرقية بأثر رجعي باستخدام ميزات TIRADS. يعتمد التشخيص النهائي للعقيدات الحميدة على شفط الإبرة الدقيقة أو التأكيد النسيجي الجراحي. حيث أن التشخيص النهائي لعقيدات الغدة الدرقية الخبيثة يعتمد على تأكيد التشريح المرضي الجراحي. بالإضافة إلى ذلك، تم رسم منحنيات تشغيل المستقبل (ROC) لمقارنة الأداء الفعلي للمراقبين، وتم اختبار موافقتهم باستخدام كابا كوهين. كما تم استخدام الارتباط داخل الطبقة (ICC) لتحديد اتفاق بين المراقبين.

نتائج: من بين 553 عقيدات درقية، كانت 452 (81.7%) حميدة و553/101 (18.3%) خبيثة. تم تأكيد تشخيص العقيدات الحميدة بواسطة FNA في 452/324 (71.7٪) من عقيدات الغدة الدرقية (بيثيسدا II) والتحليل النسيجي المرضي في عقيدات الغدة الدرقية المتبقية 452/128 (28.3٪). من بين 101 عقيدات درقية خبيثة تم تأكيدها بالفحص النسيجي المرضي. كانت غالبية العقيدات الحميدة و TR1 هي Bethesda II، في حين كانت Bethesda VI هي أكثر اكتشافات FNA شيوعًا في عقيدات الغدة الدرقية الخبيثة و TR5. كان اتفاق المراقبين على التركيب والصدى والتكلس المحيطي والتكلس الكبير والشكل كبيرًا (k = 0.752 و0.677 و0.727 و0.708 و0.603 على التوالي). كان اتفاق الهامش وذيل المذنب الكبير والبؤر المثقوبة معتدلاً (k = 0.523، 0.582، 0.484 على التوالي). كان هناك اتفاق بين المراقبين معتدل إلى مثالي تقريبًا لكل درجة TIRADS (K = 0.808، 0.742، 0.591، 0.631، 0.624 لـ TR1، TR2، TR3، TR4، وTR5 على التوالي). ولوحظ وجود اتفاق كبير بالنسبة لدرجة TI-RADS النهائية (K= 0.662؛ 95% CI: 0.614-0.709).

خاتمة: يتمتع نظام تسجيل ACR TIRADS بأداء تشخيصي عالي. يعد اتفاق المراقبين البينيين لفئات TIRADS متوسطًا إلى جوهري وكبيرًا إلى مثالي تقريبًا لمختلف درجات TIRADS.