Original Article

J Korean Soc Radiol
https://doi.org/10.3348/jksr.2025.0054
eISSN 2951-0805

JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

Comparative Analysis of ACR TI-RADS and K-TIRADS: Inter-System Agreement and Diagnostic Performance Using a Single Study Cohort

단일 연구 집단을 이용한 ACR TI-RADS와 K-TIRADS의 비교분석: 시스템 간 일치도 및 진단 성능 평가

Seohyun Ryu[1], Hye Sun Lee, PhD[2], Jin Chung, MD[3]*, Ji Ye Lee, MD[4], Soo Yeon Hahn, MD[5], Jeoung Hyun Kim, MD[3], Won Hwa Kim, MD[6,7], Jaeil Kim, PhD[7,8], Jung Hyun Yoon, MD[1,9]

[1]Yonsei University College of Medicine, Seoul, Korea
[2]Biostatistics Collaboration Unit, Yonsei University College of Medicine, Seoul, Korea
[3]Department of Radiology, Ewha Womans University, College of Medicine, Mokdong Hospital, Seoul, Korea
[4]Department of Radiology, Seoul National University Hospital, Seoul National University, College of Medicine, Seoul, Korea
[5]Department of Radiology, Samsung Medical Center, Sungkyunkwan University, College of Medicine, Seoul, Korea
[6]Department of Radiology, School of Medicine, Kyungpook National University, Kyungpook National University Chilgok Hospital, Daegu, Korea
[7]BeamWorks Inc., Daegu, Korea
[8]School of Computer Science and Engineering, Kyungpook National University, Daegu, Korea
[9]Department of Radiology, Severance Hospital, Research Institute of Radiological Science, Yonsei University College of Medicine, Seoul, Korea

Received June 19, 2025
Revised September 15, 2025
Accepted October 29, 2025
Published Online January 8, 2026

***Corresponding author**
Jin Chung, MD
Department of Radiology,
Ewha Womans University,
College of Medicine,
Mokdong Hospital,
1071 Anyangcheon-ro,
Yangcheon-gu, Seoul 07985, Korea.

Tel 82-2-2650-5977
Fax 82-2-2655-0984
E-mail aqua0724@ewha.ac.kr

Purpose To evaluate the inter-system agreement and diagnostic performance of experienced readers using two Thyroid Imaging Reporting and Data Systems (TI-RADS).

Materials and Methods A total of 481 thyroid US images were collected retrospectively. Four experienced radiologists independently reviewed the images and documented the US features and final assessments based on the Korean TI-RADS (K-TIRADS). The final American College of Radiology (ACR) TI-RADS score was calculated using individual US feature scores. The final assessments from both TI-RADS systems were used to evaluate inter-reader and inter-system agreement. The diagnostic performance metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and unnecessary biopsy rate, were calculated and compared.

Results Among the 481 US images, 157 were benign, 184 were malignant, and 140 were

JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

negative images. In the final assessment, substantial agreement was observed between the two TI-RADS systems (κ = 0.61) for all four readers. The mean AUC did not differ significantly between ACR TI-RADS and K-TIRADS (p = 0.52). Compared with K-TIRADS, ACR TI-RADS demonstrated significantly higher sensitivity and NPV but significantly lower specificity, PPV, and unnecessary biopsy rates (p < 0.001).
Conclusion Substantial agreement was observed between the final assessments of ACR TI-RADS and K-TIRADS. The mean AUC did not differ significantly between the two TI-RADS categories, reflecting consistent image interpretation and patient management regardless of the system.

Index terms Thyroid Imaging Reporting and Data System; Thyroid Nodule; Neoplasm; Ultrasonography; Fine-Needle Aspiration Biopsy

INTRODUCTION

Despite the high prevalence of thyroid nodules in up to 68% of the general population (1-3), only 5%–15% of thyroid nodules are malignant, emphasizing the need for accurate risk stratification (4). US is the standard imaging modality for thyroid nodule evaluation because of its cost-effectiveness, widespread accessibility, noninvasiveness, and absence of radiation exposure (5, 6). It also allows real-time guidance for fine-needle aspiration biopsy, facilitating the histopathological diagnosis of suspicious nodules. Nevertheless, interobserver variability, in which differences in image interpretation can cause inconsistent classification and patient management, remains a major limitation of thyroid US (7, 8). To address this issue, several guidelines that incorporate US descriptors, collectively known as the Thyroid Imaging Reporting and Data Systems (TI-RADS), have been developed (5, 6, 9, 10). TI-RADS provides a standardized lexicon for thyroid nodule descriptions based on US features and establishes risk categories to guide clinical decision-making. However, variations in the classification criteria and clinical applications exist among the different TIRADS versions, as no universal system has yet been established.

Although the published TI-RADS systems employ similar US descriptor categories, they differ in how these features are applied to determine the final assessment category. Many pattern-based TI-RADS approaches derive the final assessment based on the presence of suspicious US features (5, 9, 10). In contrast, the American College of Radiology (ACR) TI-RADS assigns numerical points to individual US features and sums them to determine the final assessment category—representing a score-based TI-RADS (6). Although these systems share US descriptors, the final assessment categories of pattern-based versus score-based systems may differ depending on how conclusions are derived. Several previous studies comparing the performance of thyroid US risk stratification systems have relied on individual US features retrospectively extracted from examination reports, with final assessments retrospectively categorized based on different TI-RADS (11-13). Direct comparisons of the agreements and differences in performance outcomes based on the final assessment categories provided by readers using different TI-RADS scores have not been investigated.

In this study, inter-reader agreement, differences in performance measures, and inter-system agreement were evaluated among experienced readers using the score-based ACR TI-

RADS and pattern-based Korean Society of Thyroid Radiology TI-RADS (K-TIRADS) on the same image datasets.

MATERIALS AND METHODS

This retrospective study was approved by the Institutional Review Board of Yonsei University Severance Hospital, Seoul, Korea (IRB No. 2024-3991-001); Seoul National University Hospital, Seoul, Korea (IRB No. H-2311-149-1487), and Ewha Womans University Mokdong Hospital, Seoul, Korea (IRB No. 2023-09-024-001).

The requirement for informed consent was waived for access to medical records and US image review.

STUDY POPULATION

A total of 511 thyroid US examinations from 511 patients were retrospectively reviewed at Referral Centers A (n = 311) and B (n = 200) between January 2020 and December 2020. Based on the final cytopathological diagnosis, 511 representative US images were selected by experienced radiologists at Seoul National University Hospital (J.Y.L.) and Ewha Womans University Mokdong Hospital (J.C.), who had 10 and 15 years of experience in thyroid imaging, respectively. Thirteen images were excluded because the entire lesion was not completely captured, and 17 were excluded because of inadequate image quality for interpretation. Finally, 481 thyroid US images from 481 patients were included in the analysis (Fig. 1).

READING SESSIONS WITH EXPERIENCED READERS

Four board-certified radiologists with 23, 17, 16, and 3 years of experience in thyroid imaging participated in this study. US images were uploaded to a dedicated web-based system built for this reader study (Fig. 2). Each reader independently reviewed the US images for image analysis, blinded to the patients' clinical and cytopathological information.

Fig. 1. Diagram illustrating the inclusion of 481 thyroid US images.

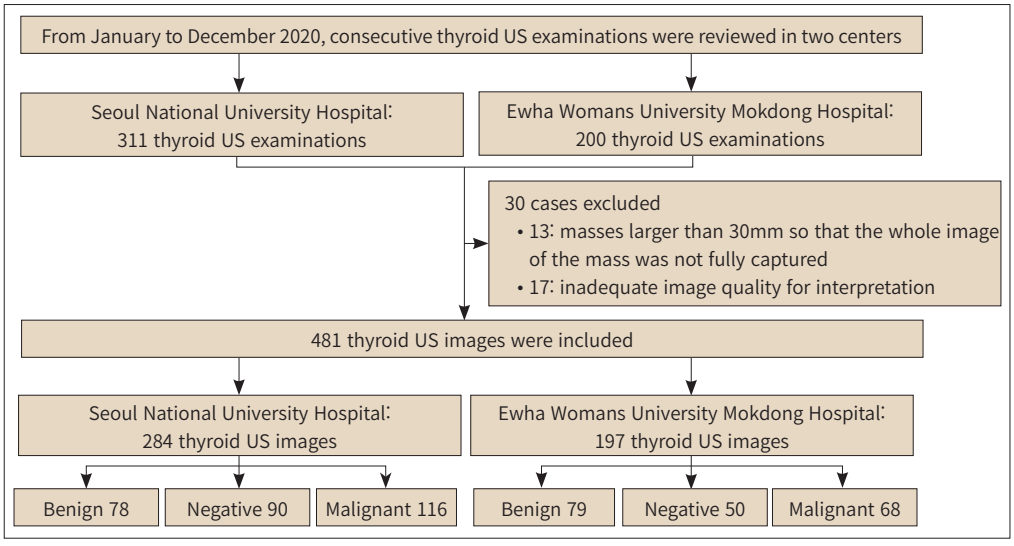
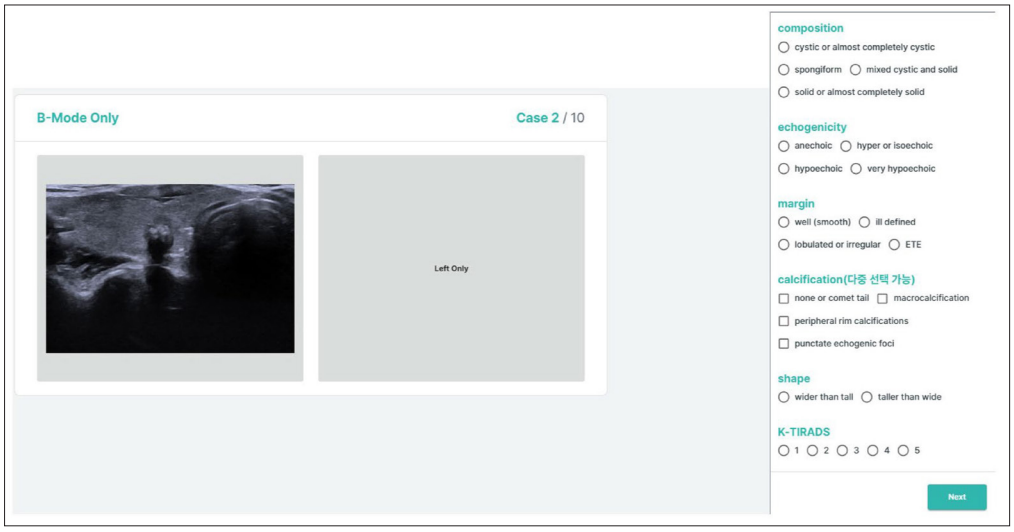


Fig. 2. Web-based interface shows a grayscale US image on the left and an input panel on the right, where readers record individual US features and final K-TIRADS assessments, while ACR TI-RADS assessments were later calculated from these recorded features.
ACR = American College of Radiology, K-TIRADS = Korean TI-RADS, TI-RADS = Thyroid Imaging Reporting and Data Systems



The readers were instructed to analyze the US images according to the following thyroid US descriptors: composition, echogenicity, shape, margin, and echogenic foci. The definitions of each US descriptor and final assessment are summarized in Table 1 (10, 14). Individual descriptors provided by the readers were converted into ACR TI-RADS scores, from which the corresponding ACR TI-RADS final assessment was determined (6). The readers also provided a final K-TIRADS assessment based on the Korean Society of Thyroid Radiology 2021 K-TIRADS (Supplementary Table 1) (10).

DATA AND STATISTICAL ANALYSIS

Histopathological results obtained from surgical resections, US-guided fine-needle aspiration, or core biopsies were used as reference standards. Those with cytopathologically confirmed benign lesions or typically benign US findings and lesions that remained stable on follow-up imaging for more than one year were also included.

Fleiss' κ statistics were used to evaluate the inter-reader agreement of the four readers on individual US descriptors and for each final ACR TI-RADS and K-TIRADS assessment. Cohen's κ values (unweighted, linearly weighted, and quadratically weighted) were used for inter-system agreement between ACR TI-RADS and K-TIRADS final assessments. κ ≤0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80, and ≥0.80 indicate slight, fair, moderate, substantial, and almost perfect agreement, respectively (15). For the final assessment category agreement, individual final assessments and binary classifications, that is, classifying TI-RADS 1–3 and 4–5 as 'negative' and 'positive' assessments, respectively, were analyzed based on the threshold at which aspiration was considered for nodules with a higher suspicion of malignancy. Stratified analyses were conducted separately based on the pathological diagnosis of benign or malignant tumors. The diagnostic performance, including sensitivity, specificity, positive predictive val-

https://doi.org/10.3348/jksr.2024.0054

Table 1. Definition of Thyroid US Descriptors and Final Assessments According to ACR TI-RADS and K-TIRADS

US Lexicon	Descriptor	Definition in ACR TI-RADS	Definition in K-TIRADS	ACR-TIRADS Score	Feature Indicated 'Suspicious' in K-TIRADS
Composition	Solid	Composed entirely or nearly entirely of soft tissue, with only a few tiny cystic spaces	No obvious cystic component	2	-
	Predominantly solid	Composed of soft tissue components occupying >50% of the volume of the nodule	Cystic portion ≤50%	1	-
	Predominantly cystic	Composed of soft tissue components occupying ≤50% of the volume of the nodule	Cystic portion >50%	1	-
	Cystic	Entirely fluid filled	No obvious solid component	0	-
	Spongiform	Composed predominately of tiny cystic spaces	Microcystic changes >50% of solid component	0	-
Echogenicity	Very hypoechoic*/marked hypoechogenicity†	Decreased echogenicity relative to adjacent neck musculature	Hypoechoic or similar echogenicity relative to the anterior neck muscles	3	-
	Hypoechoic*/mild hypoechogenicity†	Decreased echogenicity relative to thyroid tissue	Hypoechoic relative to the normal thyroid parenchyma and hyperechoic relative to the anterior neck muscles	2	-
	Isoechoic*/isoechogenicity†	Similar echogenicity relative to thyroid tissue	Same echogenicity as that of the normal thyroid parenchyma	1	-
	Hyperechoic*/hyperechogenicity†	Increased echogenicity relative to thyroid tissue	Hyperechoic relative to the normal thyroid parenchyma	1	-
Shape	Taller-than-wide*/nonparallel†	Anteroposterior diameter > horizontal diameter in the transverse plane	Anteroposterior diameter > transverse diameter in the transverse plane	3	Suspicious feature
	Wider-than-tall*/parallel†	-	Anteroposterior diameter ≤ transverse diameter in the transverse plane	0	-
Margin	Smooth	Uninterrupted, well-defined, curvilinear edge typically forming a spherical or elliptical shape	Obviously discernible smooth edges	0	-

JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

Table 1. Definition of Thyroid US Descriptors and Final Assessments According to ACR TI-RADS and K-TIRADS (Continued)

US Lexicon	Descriptor	Definition in ACR TI-RADS	Definition in K-TIRADS	ACR-TIRADS Score	Feature Indicated 'Suspicious' in K-TIRADS
	Irregular	The outer border of the nodule is spiculated, jagged, or with sharp angles ± clear soft tissue protrusions into the parenchyma.	Obviously discernible, but non-smooth edges with spiculations or microlobulations	2	Suspicious feature
	Ill-defined	Border of the nodule is difficult to distinguish from thyroid parenchyma	Poorly demarcated margins, which cannot be obviously differentiated from the adjacent thyroid tissue	0	-
	Lobulated*	Border has focal rounded soft tissue protrusions that extend into the adjacent parenchyma	-	2	-
	Halo*	Border consists of a dark rim around the periphery of the nodule	-	N/A	-
	Extrathyroidal extension*	Nodule extends through the thyroid capsule	-	3	-
Echogenic foci	Punctate echogenic foci	"Dot-like" foci having no posterior acoustic posterior artifacts	Punctate (≤1 mm) hyperechoic foci within the solid component of a nodule	3	Suspicious feature
	Macrocalcifications	When calcifications become large enough to result in posterior acoustic shadowing	Large (>1 mm) hyperechoic foci with posterior acoustic shadowing	1	-
	Peripheral calcifications*/ rim calcification†	Calcification at the periphery of nodule, which may not be completely continuous but involves the majority of the margin	Peripheral curvilinear hyperechoic line surrounding the nodule margin ± posterior shadowing	2	-
	Comet-tail artifacts*/intracystic echogenic foci with comet-tail artifact†	A type of reverberation artifact; deeper echoes become attenuated and are displayed as decreased width	Intracystic echogenic foci showing comet-like echogenic tail	0	-

*Indicate distinctive lexicons that are only described in ACR-TIRADS.
†Indicate distinctive lexicons that are only described in K-TIRADS.
ACR = American College of Radiology, K-TIRADS = Korean TI-RADS, TI-RADS = Thyroid Imaging Reporting and Data Systems

JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

ue (PPV), negative predictive value (NPV), accuracy, and area under the receiver operating characteristic curve (AUC), were compared between ACR TI-RADS and K-TIRADS. The unnecessary biopsy rate was defined as the proportion of negative or benign cases among nodules recommended for biopsy based on the size-based criteria of each system (K-TIRADS or ACR TI-RADS). The rates of the two systems were compared. The standard errors of the diagnostic performance metrics were estimated and compared using a nonparametric bootstrap method with 1,000 re-samples.

Statistical analyses were performed using R software version 4.3.2 (R Foundation for Statistical Computing, Vienna, Austria). A *p*-value of <0.05 was considered statistically significant.

RESULTS

A total of 481 patients were included in the study (392 women [81.5%] and 89 men [18.5%]; mean age, 54 years; range, 18–86 years). Among the 481 US images, 184 (38.3%) were malignant, 157 (32.6%) were benign, and 140 (29.1%) were negative without focal abnormalities. The mean size of the 341 benign and malignant nodules was 16.7 ± 8.8 mm (range: 5–51 mm).

AGREEMENT ON INDIVIDUAL US DESCRIPTORS AND FINAL ASSESSMENTS

The agreements among the four experienced readers on individual US descriptors, ACR TI-RADS, and K-TIRADS final assessments are summarized in Table 2. Moderate agreement was observed for individual US descriptors (κ range, 0.41–0.48), with the highest agreement for echogenic foci (κ = 0.48; 95% confidence interval [CI], 0.46–0.51). When stratified by pathology, the agreement for echogenic foci remained highest in both the benign and malignant subgroups (κ = 0.43; 95% CI, 0.38–0.47 and κ = 0.54; 95% CI, 0.50–0.58, respectively).

For the final assessments, the readers showed moderate ACR TI-RADS and K-TIRADS

Table 2. Agreement of Individual US Descriptors and Final Assessments for ACR TI-RADS and K-TIRADS

	Overall (*n* = 481)	Benign (*n* = 184)	Malignant (*n* = 157)
US Descriptors			
Composition	0.42 (0.39–0.44)	0.31 (0.26–0.36)	0.23 (0.17–0.28)
Echogenicity	0.41 (0.39–0.43)	0.27 (0.22–0.32)	0.27 (0.23–0.32)
Shape	0.42 (0.40–0.45)	0.10 (0.05–0.16)	0.45 (0.40–0.51)
Margin	0.41 (0.39–0.44)	0.23 (0.18–0.28)	0.35 (0.30–0.39)
Echogenic foci	0.48 (0.46–0.51)	0.43 (0.38–0.47)	0.54 (0.50–0.58)
ACR TI-RADS			
Final assessment	0.53 (0.51–0.55)	0.31 (0.27–0.35)	0.34 (0.29–0.38)
Final assessment (binary)	0.69 (0.66–0.73)	0.34 (0.27–0.40)	0.44 (0.38–0.50)
K-TIRADS			
Final assessment	0.54 (0.52–0.56)	0.29 (0.25–0.33)	0.27 (0.23–0.31)
Final assessment (binary)	0.68 (0.64–0.71)	0.38 (0.32–0.45)	0.40 (0.34–0.46)

Data are expressed as Cohen's κ (95% confidence interval).
Binary classification of final assessment: TIRADS 1–3 considered negative, 4–5 considered positive.
ACR = American College of Radiology, K-TIRADS = Korean TI-RADS, TI-RADS = Thyroid Imaging Reporting and Data Systems

JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

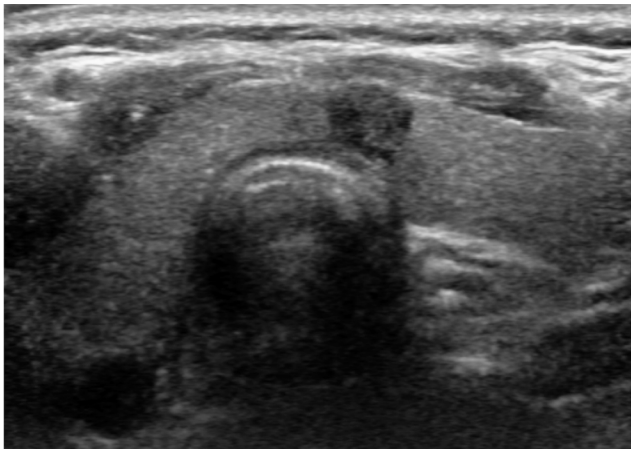
agreement (κ = 0.53; 95% CI, 0.51–0.55 and κ = 0.54; 95% CI, 0.52–0.56, respectively). ACR TI-RADS and K-TIRADS demonstrated substantial agreement when categorized into binary assessment groups (κ = 0.69, 95% CI, 0.66–0.73 and κ = 0.68, 95% CI, 0.64–0.71, respectively).

When grouped by pathology (benign and malignant), fair agreement was observed for ACR TI-RADS (κ = 0.31; 95% CI, 0.27–0.35 and κ = 0.34; 95% CI, 0.29–0.38, respectively) and K-TI-RADS (κ = 0.29; 95% CI, 0.25–0.33 and κ = 0.27; 95% CI, 0.23–0.31, respectively) (Figs. 3, 4). When categorized into binary assessment groups, fair agreement was observed for the benign ACR TIRADS subgroup (κ = 0.34; 95% CI, 0.27–0.40) and the benign and malignant K-TIRADS subgroups (κ = 0.38; 95% CI, 0.32–0.45, κ = 0.40; 95% CI, 0.34–0.46, respectively). Moderate agreement was observed in the malignant ACR TI-RADS subgroup (κ = 0.44; 95% CI, 0.38–0.50).

AGREEMENT BETWEEN ACR TI-RADS VS. K-TIRADS FINAL ASSESSMENTS

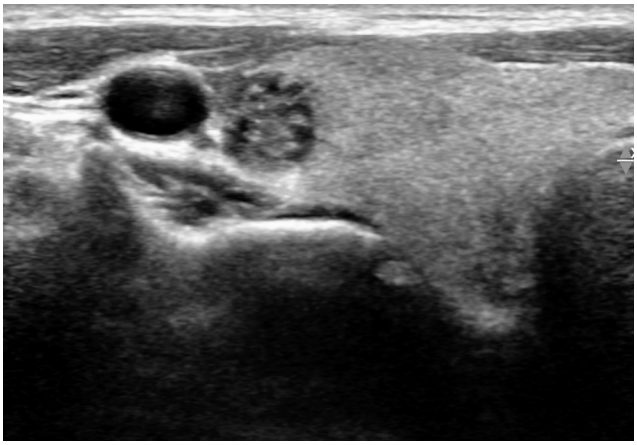
The agreement between the individual readers' ACR TI-RADS and K-TIRADS final assessments and their averages are summarized in Table 3. Moderate to substantial agreement was observed between the ACR TI-RADS and K-TIRADS assessments across the four readers (unweighted κ range, 0.53–0.71). Using unweighted measures, the average of the four readers demonstrated substantial agreement between the ACR TI-RADS and K-TIRADS assessments

Fig. 3. US image shows a 7-mm nodule in the left thyroid lobe, lower isthmic portion, in a 49-year-old woman, which was confirmed as papillary thyroid carcinoma by fine-needle aspiration and subsequent surgery, with three readers assessing it as ACR TI-RADS and K-TIRADS 4 or 5, and one reader assessing it as ACR TI-RADS 2 and K-TIRADS 3, as summarized in the table.
ACR = American College of Radiology, K-TIRADS = Korean TI-RADS, TI-RADS = Thyroid Imaging Reporting and Data Systems



	Reader 1	Reader 2	Reader 3	Reader 4
Composition	Solid or almost completely solid	Solid or almost completely solid	Mixed cystic and solid	Solid or almost completely solid
Echogenicity	Hypoechoic	Hypoechoic	Hyper or isoechoic	Hypoechoic
Margin	Lobulated or irregular	Lobulated or irregular	Well (smooth)	Lobulated or irregular
Echogenic foci	None or comet tail	None or comet tail	None or comet tail	None or comet tail
Shape	Wider than tall	Wider than tall	Wider than tall	Wider than tall
ACR TI-RADS	4	4	2	4
K-TIRADS	4	5	3	4

Fig. 4. US image shows a 10-mm nodule in the right mid-lobe of the thyroid in a 43-year-old woman, which was confirmed as benign (Bethesda II) by fine-needle aspiration, with two readers assessing it as TI-RADS 2 or 3 and the other two readers assessing it as TI-RADS 4 or 5, as summarized in the table.
ACR = American College of Radiology, K-TIRADS = Korean TI-RADS, TI-RADS = Thyroid Imaging Reporting and Data Systems



	Reader 1	Reader 2	Reader 3	Reader 4
Composition	Spongiform	Mixed cystic and solid	Mixed cystic and solid	Solid or almost completely solid
Echogenicity	Hypoechoic	Hyper or isoechoic	Hyper or isoechoic	Hypoechoic
Margin	Well (smooth)	Well (smooth)	Well (smooth)	Well (smooth)
Echogenic foci	None or comet tail	Punctate echogenic foci	None or comet tail	None or comet tail
Shape	Wider than tall	Taller than wide	Wider than tall	Wider than tall
ACR TI-RADS	2	5	2	4
K-TIRADS	2	4	3	4

Table 3. Agreement Among Readers Using ACR TI-RADS vs. K-TIRADS Final Assessment Categories

	Unweighted	Linearly Weighted	Quadratically Weighted	Binary
Reader 1	0.60 (0.55–0.65)	0.80 (0.78–0.83)	0.92 (0.90–0.94)	0.63 (0.57–0.70)
Reader 2	0.71 (0.66–0.75)	0.87 (0.84–0.89)	0.95 (0.94–0.96)	0.92 (0.89–0.96)
Reader 3	0.53 (0.48–0.58)	0.76 (0.73–0.79)	0.90 (0.89–0.92)	0.91 (0.88–0.95)
Reader 4	0.60 (0.55–0.65)	0.78 (0.75–0.82)	0.90 (0.87–0.92)	0.74 (0.68–0.80)
Average	0.61 (0.58–0.64)	0.81 (0.78–0.83)	0.92 (0.91–0.93)	0.80 (0.77–0.83)

Data are expressed as Cohen's κ (95% confidence interval).
Binary classification of final assessment: TIRADS 1–3 considered negative, 4–5 considered positive.
ACR = American College of Radiology, K-TIRADS = Korean TI-RADS, TI-RADS = Thyroid Imaging Reporting and Data Systems

(unweighted κ = 0.61; 95% CI, 0.58–0.64). When weighted measures were applied, almost perfect agreement was observed (linearly weighted κ = 0.81, 95% CI: 0.78–0.83; quadratically weighted κ = 0.92, 95% CI: 0.91–0.93). The average agreement among the four readers demonstrated almost perfect agreement when the final assessments were classified into binary assessment groups (κ = 0.80; 95% CI, 0.77–0.83). Supplementary Table 2 presents the stratified reader agreement analysis based on the pathology.

Table 4. Comparison of Diagnostic Performance Between Readers Using ACR TI-RADS and K-TIRADS for Thyroid Nodule Categorization

		Sensitivity	Specificity	PPV	NPV	Accuracy	Unnecessary Biopsy Rate	AUC
ACR TI-RADS	Reader 1	90.2* (85.8-94.6)	74.1* (69.1–79.1)	68.3* (62.6–74.1)	92.4* (89.1–95.8)	80.2* (76.7–83.8)	40.2 (33.0–47.5)	0.889 (0.882–0.924)
	Reader 2	89.1* (84.5–93.7)	88.2* (84.6–91.9)	82.4* (77.2–87.6)	92.9* (89.8–96.0)	88.6 (85.6–91.5)	25.0* (18.2–31.8)	0.925 (0.901–0.949)
	Reader 3	84.2 (78.9–89.6)	86.5* (82.5–90.6)	79.5 (73.8–85.1)	89.9 (86.3–93.4)	85.7 (82.4–89.0)	28.6* (20.9–36.3)	0.904 (0.876–0.933)
	Reader 4	89.1* (84.6–93.7)	79.1* (74.5–83.8)	72.6* (66.8–95.5)	92.2* (88.8–95.5)	83.0 (79.6–86.3)	38.0 (30.7–45.3)	0.894 (0.867–0.921)
	Average	88.2* (84.5–91.8)	82.0* (78.7–85.3)	75.7* (71.0–80.4)	91.8* (89.1–94.6)	84.4 (81.8–86.9)	33.0* (26.9–39.0)	0.903 (0.882–0.924)
K-TIRADS	Reader 1	71.2* (64.7–77.6)	92.3* (89.2–95.3)	85.1* (79.6–90.6)	83.8* (79.8–87.8)	84.2* (81.0–87.4)	37.1 (30.5–43.7)	0.896 (0.871–0.921)
	Reader 2	84.2* (79.0–89.5)	91.2* (88.1–94.4)	85.6* (80.6–90.7)	90.3* (87.0–93.7)	88.6 (85.7–91.5)	35.8* (29.5–42.1)	0.927 (0.904–0.949)
	Reader 3	86.4 (81.4–91.4)	84.5* (80.3–88.7)	77.6 (72.0–83.2)	91.0 (87.5–94.4)	85.2 (82.0–88.5)	40.4* (34.2–46.5)	0.895 (0.867–0.923)
	Reader 4	75.0* (68.5–81.5)	88.2* (84.5–92.0)	79.8* (73.6–85.9)	85.1* (81.1–89.1)	83.2 (79.7–86.6)	37.3 (31.1–43.5)	0.881 (0.853–0.910)
	Average	79.2* (74.8–83.6)	89.1* (86.4–91.7)	82.0* (77.5–86.5)	87.5* (84.4–90.6)	85.3 (82.9–87.7)	37.6* (31.8–43.5)	0.900 (0.879–0.921)

95% confidence intervals are in parentheses.

*Indicate statistically significant differences between ACR-TIRADS vs. K-TIRADS.

ACR = American College of Radiology, AUC = area under the receiver operating characteristics curve, K-TIRADS = Korean TI-RADS, NPV = negative predictive value, PPV = positive predictive value, TI-RADS = Thyroid Imaging Reporting and Data Systems

DIAGNOSTIC PERFORMANCES OF ACR TI-RADS AND K-TIRADS

The diagnostic performances of the individual readers using ACR TI-RADS, K-TIRADS, and their averages are shown in Table 4. Compared with K-TIRADS, all three readers demonstrated higher sensitivity and NPV but lower specificity and PPV when using ACR TI-RADS (all $p < 0.05$). One reader showed significantly higher specificity when using the ACR TI-RADS, whereas no significant differences were observed for the other diagnostic indices. The AUC did not differ significantly between ACR TI-RADS and K-TIRADS across the four readers (all $p < 0.05$).

Compared with K-TIRADS, average sensitivity (88.2%; 95% CI, 84.5–91.8 vs. 79.2%; 95% CI, 74.8–83.6) and NPV (91.8%; 95% CI, 89.1–94.6 vs. 87.5%; 95% CI, 84.4–90.6) were significantly higher in ACR TI-RADS (all $p < 0.001$). However, average specificity, PPV, and unnecessary biopsy rates were significantly lower in ACR TI-RADS than in K-TIRADS (82.0%, 95% CI, 78.7–85.3 vs. 89.1%, 95% CI, 86.4–91.7; 75.7%, 95% CI, 71.0–80.4 vs. 82.0%, 95% CI, 77.5–86.5; 33.0%, 95% CI, 26.9–39.0 vs. 37.6%, 95% CI, 31.8–43.5; all $p < 0.001$). When using ACR TI-RADS, the average AUC of the four readers was not significantly different from that of K-TIRADS (AUC: 0.903; 95% CI, 0.882–0.927 vs. 0.900; 95% CI, 0.879–0.921; $p = 0.52$).

DISCUSSION

This study evaluated the intersystem agreement and diagnostic performance differences among experienced readers using score-based ACR TI-RADS and pattern-based K-TIRADS on the same test dataset. The ACR TI-RADS and K-TIRADS assessments by the four experienced readers showed substantial agreement. Compared with K-TIRADS, significantly higher average sensitivity and NPV were observed for ACR TI-RADS, with a trade-off between lower specificity and PPV. The overall diagnostic performance, as measured by the AUC, did not differ significantly between ACR TI-RADS and K-TIRADS. These findings indicate that experienced readers can achieve high TI-RADS assessment agreement without significant differences in performance outcomes between the pattern- and score-based TI-RADS systems.

The agreement on individual US descriptors was moderate for all 481 images ($\kappa = 0.41$–0.48). Because these individual US descriptors form the basis of the final ACR TI-RADS and K-TIRADS assessments, the moderate agreement between the readers for ACR TI-RADS ($\kappa = 0.53$) and K-TIRADS ($\kappa = 0.54$) was expected and consistent with previous studies (7, 8, 16). However, the agreement decreased for US images containing benign or malignant nodules, showing only slight to fair agreement ($\kappa = 0.10$–0.54). The inclusion of 'negative' US images likely contributed to the higher overall agreement in the complete dataset, as experienced readers tend to demonstrate high concordance in these cases. Both systems demonstrated substantial agreement in the binary classification of the final assessments (ACR TI-RADS: $\kappa = 0.69$, K-TIRADS: $\kappa = 0.69$), reflecting strong consistency across readers when making patient management decisions.

Despite moderate agreement in the individual US descriptors of the two TI-RADS, agreement between ACR TI-RADS and K-TIRADS among the four experienced readers was moderate to substantial ($\kappa = 0.53$–0.71). The two TI-RADS differ in how individual US features are weighted and integrated to determine the final assessment category—that is, the score-based vs. pattern-based approach. To the best of our knowledge, this is the first study to evaluate

the intersystem agreement between different TI-RADS classifications. Our findings indicate that despite structural differences, agreement in the final assessment using different TI-RADS scores remained relatively consistent when assessed by experienced readers.

In this study, ACR TI-RADS showed higher sensitivity and NPV, whereas K-TIRADS exhibited higher specificity and PPV, consistent with previous studies (11, 17, 18). Among experienced readers, the overall accuracy and AUC did not differ significantly between ACR TI-RADS and K-TIRADS, indicating similar diagnostic performance despite structural differences. The average unnecessary biopsy rate was significantly lower for ACR TI-RADS than for K-TIRADS, reflecting the stricter biopsy criteria of ACR TI-RADS, which require larger nodule sizes for biopsy in category 3 (≥25 mm vs. ≥20 mm) and category 4 (≥15 mm vs. ≥10 mm) (19). Therefore, clinicians should recognize the strengths and limitations of each system and apply them appropriately in the clinical context. The average AUCs of the experienced readers for ACR TI-RADS and K-TIRADs were 0.903 and 0.900, respectively. Although the final assessment was based on US image analyses of individual US features, where moderate agreement was observed even among experienced readers, it is evident that the currently used TI-RADS offers high diagnostic performance in differentiating thyroid nodules, thereby providing reliable guidance for patient management.

This study had certain limitations. First, it used US images that were retrospectively captured by the operators and were not included as the readers. Second, for the ACR TI-RADS, the final assessments were based on the sum of the scores for the US descriptors provided by the readers instead of being directly provided by the readers. Because the readers sequentially assessed the images for individual US features and K-TIRADS, K-TIRADS assessments may have had less influence on the use of ACR TI-RADS assessments based on summed ACR scores. Third, only experienced readers were included in this study. Less experienced readers were excluded because the type of training and level of expertise critically affected individual US feature assessment. Finally, of the various pattern-based systems, K-TIRADS was used (5, 9, 10) because it is used in daily practice in Korea. Therefore, other pattern-based TI-RADS may yield different results when applied to our study samples.

In conclusion, this study demonstrated that experienced readers achieved moderate to substantial agreement in final assessments when using ACR TI-RADS and K-TIRADS. The mean AUC did not differ significantly between the two TI-RADS types, reflecting the consistency in image interpretation and patient management, regardless of the TI-RADS used.

Supplementary Materials

The Supplement is available with this article at http://doi.org/10.3348/jksr.2024.0054.

Author Contributions

Conceptualization, C.J., L.J.Y., K.W.H., K.J., Y.J.H.; data curation, R.S.; formal analysis, L.H.S.; investigation, C.J., L.J.Y., H.S.Y., K.J.H., Y.J.H.; methodology, L.H.S., C.J., Y.J.H.; project administration, C.J., L.J.Y., H.S.Y., K.J.H., K.W.H., K.J., Y.J.H.; resources, C.J., L.J.Y., H.S.Y., K.J.H.; software, K.W.H., K.J.; supervision, C.J., L.J.Y., Y.J.H.; validation, L.J.Y., H.S.Y., Y.J.H.; writing—original draft, R.S., L.H.S., C.J., Y.J.H.; and writing—review & editing, R.S., C.J., L.J.Y., H.S.Y., K.J.H., K.W.H., K.J., Y.J.H.

Conflicts of Interest

Won Hwa Kim and Jaeil Kim are the CEOs of BeamWorks. The remaining authors have declared no conflicts of interest.

ORCID iDs

Seohyun Ryu https://orcid.org/0000-0002-8718-1224
Hye Sun Lee https://orcid.org/0000-0001-6328-6948
Jin Chung https://orcid.org/0000-0001-9990-3768
Ji Ye Lee https://orcid.org/0000-0002-3929-6254
Soo Yeon Hahn https://orcid.org/0000-0002-4099-1617
Jeoung Hyun Kim https://orcid.org/0000-0003-3504-9595
Won Hwa Kim https://orcid.org/0000-0001-7137-9968
Jaeil Kim https://orcid.org/0000-0002-9799-1773
Jung Hyun Yoon https://orcid.org/0000-0002-2100-3513

Funding

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project No. 1711197554, RS-2023-00227526). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Guth S, Theune U, Aberle J, Galach A, Bamberger CM. Very high prevalence of thyroid nodules detected by high frequency (13 MHz) ultrasound examination. *Eur J Clin Invest* 2009;39:699-706
2. Hegedüs L. Clinical practice. The thyroid nodule. *N Engl J Med* 2004;351:1764-1771
3. Remonti LR, Kramer CK, Leitão CB, Pinto LC, Gross JL. Thyroid ultrasound features and risk of carcinoma: a systematic review and meta-analysis of observational studies. *Thyroid* 2015;25:538-550
4. Papini E, Guglielmi R, Bianchini A, Crescenzi A, Taccogna S, Nardi F, et al. Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. *J Clin Endocrinol Metab* 2002;87:1941-1946
5. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology* 2011;260:892-899
6. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J Am Coll Radiol* 2017;14:587-595
7. de Carlos J, Garcia J, Basterra FJ, Pineda JJ, Dolores Ollero M, Toni M, et al. Interobserver variability in thyroid ultrasound. *Endocrine* 2024;85:730-736
8. Li W, Sun Y, Xu H, Shang W, Dong A. Systematic review and meta-analysis of American College of Radiology TI-RADS inter-reader reliability for risk stratification of thyroid nodules. *Front Oncol* 2022;12:840516
9. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: the EU-TIRADS. *Eur Thyroid J* 2017;6:225-237
10. Ha EJ, Chung SR, Na DG, Ahn HS, Chung J, Lee JY, et al. 2021 Korean thyroid imaging reporting and data system and imaging-based management of thyroid nodules: Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean J Radiol* 2021;22:2094-2123
11. Yoon JH, Lee HS, Kim EK, Moon HJ, Park VY, Kwak JY. Pattern-based vs. score-based guidelines using ultrasound features have different strengths in risk stratification of thyroid nodules. *Eur Radiol* 2020;30:3793-3802
12. Hoang JK, Middleton WD, Langer JE, Schmidt K, Gillis LB, Nair SS, et al. Comparison of thyroid risk categorization systems and fine-needle aspiration recommendations in a multi-institutional thyroid ultrasound registry. *J Am Coll Radiol* 2021;18:1605-1613
13. Castellana M, Castellana C, Treglia G, Giorgino F, Giovanella L, Russ G, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. *J Clin Endocrinol Metab* 2020;105:dgz170
14. Grant EG, Tessler FN, Hoang JK, Langer JE, Beland MD, Berland LL, 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(12 Pt A):1272-1279
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:

159-174
16. Grani G, Lamartina L, Cantisani V, Maranghi M, Lucia P, Durante C. Interobserver agreement of various thyroid imaging reporting and data systems. *Endocr Connect* 2018;7:1-7
17. Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH. Diagnostic performance of four ultrasound risk stratification systems: a systematic review and meta-analysis. *Thyroid* 2020;30:1159-1168
18. Kim DH, Chung SR, Choi SH, Kim KW. Accuracy of thyroid imaging reporting and data system category 4 or 5 for diagnosing malignancy: a systematic review and meta-analysis. *Eur Radiol* 2020;30:5611-5624
19. Kim PH, Suh CH, Baek JH, Chung SR, Choi YJ, Lee JH. Unnecessary thyroid nodule biopsy rates under four ultrasound risk stratification systems: a systematic review and meta-analysis. *Eur Radiol* 2021;31:2877-2885

단일 연구 집단을 이용한 ACR TI-RADS와 K-TIRADS의 비교분석: 시스템 간 일치도 및 진단 성능 평가

류서현[1] · 이혜선[2] · 정 진[3]* · 이지예[4] · 한수연[5] · 김정현[3] · 김원화[6,7] · 김재일[7,8] · 윤정현[1,9]

목적 같은 환자군 초음파 영상을 기반으로, 두 상이한 Thyroid Imaging Reporting and Data Systems (이하 TI-RADS) 체계를 사용하였을 때 숙련된 판독자 간의 평가 일치도와 진단 성능을 비교하고자 하였다.

대상과 방법 본 후향 연구는 갑상선 초음파 영상 481건을 분석 대상으로 하였다. 갑상선 영상 판독에 숙련된 4명의 영상의학 전문의가 영상을 독립적으로 판독하였으며, 개별 기술어와 Korean TI-RADS (이하 K-TIRADS) 최종 범주를 기록하였다. 기록된 기술어를 통해 American College of Radiology (이하 ACR)-TIRADS 최종 범주를 산출하여 판독자 간 및 체계 간 일치도를 분석하였으며, 진단 성능을 비교하였다.

결과 총 481건의 영상 중 157건은 양성, 184건은 악성, 140건은 음성이었다. 두 TI-RADS 체계 간 최종 범주 일치도는 substantial ($\kappa = 0.61$) 하였다. 두 체계 간 평균 AUC는 유의한 차이를 보이지 않았다($p = 0.52$). ACR TI-RADS는 K-TIRADS에 비해 유의하게 높은 민감도와 음성예측도를 보였으나, 특이도, 양성예측도, 불필요한 생검률은 유의하게 낮았다(모두 $p < 0.001$).

결론 ACR-TIRADS와 K-TIRADS 간 최종 범주는 substantial 한 일치도를 보였다. 두 체계 간 AUC는 유의한 차이가 없었으며, 이는 분류체계에 관계없이 일관된 영상 판독 및 환자 관리가 가능함을 시사한다.

[1]연세대학교 의과대학,
[2]연세대학교 의과대학 의학통계실,
[3]이화여자대학교 의과대학 이대목동병원 영상의학과,
[4]서울대학교 의과대학 서울대학교병원 영상의학과,
[5]성균관대학교 의과대학 삼성서울병원 영상의학과,
[6]경북대학교 의과대학 칠곡경북대학교병원 영상의학과,
[7]빔웍스,
[8]경북대학교 컴퓨터학부,
[9]연세대학교 의과대학 세브란스병원 영상의학과