



Comparison of the diagnostic performance of the artificial intelligence-based TIRADS algorithm with established classification systems for thyroid nodules

Abdilkadir Bozkuş¹

Yeliz Başar¹

Koray Güven²

¹Acıbadem Maslak Hospital, Clinic of Radiology, İstanbul, Türkiye

²Acıbadem Mehmet Ali Aydınlar University Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

PURPOSE

This study aimed to evaluate and compare the diagnostic performance of various Thyroid Imaging Reporting and Data Systems (TIRADS), with a particular focus on the artificial intelligence-based TIRADS (AI-TIRADS), in characterizing thyroid nodules.

METHODS

In this retrospective study conducted between April 2016 and May 2022, 1,322 thyroid nodules from 1,139 patients with confirmed cytopathological diagnoses were included. Each nodule was assessed using TIRADS classifications defined by the American College of Radiology (ACR-TIRADS), the American Thyroid Association (ATA-TIRADS), the European Thyroid Association (EU-TIRADS), the Korean Thyroid Association (K-TIRADS), and the AI-TIRADS. Three radiologists independently evaluated the ultrasound (US) characteristics of the nodules using all classification systems. Diagnostic performance was assessed using sensitivity, specificity, positive predictive value (PPV), and negative predictive value, and comparisons were made using the McNemar test.

RESULTS

Among the nodules, 846 (64%) were benign, 299 (22.6%) were of intermediate risk, and 147 (11.1%) were malignant. The AI-TIRADS demonstrated a PPV of 21.2% and a specificity of 53.6%, outperforming the other systems in specificity without compromising sensitivity. The specificities of the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS were 44.6%, 39.3%, 40.1%, and 40.1%, respectively (all pairwise comparisons with the AI-TIRADS: $P < 0.001$). The PPVs for the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS were 18.5%, 17.9%, 17.9%, and 17.4%, respectively (all pairwise comparisons with the AI-TIRADS, excluding the ACR-TIRADS: $P < 0.05$).

CONCLUSION

The AI-TIRADS shows promise in improving diagnostic specificity and reducing unnecessary biopsies in thyroid nodule assessment while maintaining high sensitivity. The findings suggest that the AI-TIRADS may enhance risk stratification, leading to better patient management. Additionally, the study found that the presence of multiple suspicious US features markedly increases the risk of malignancy, whereas isolated features do not substantially elevate the risk.

CLINICAL SIGNIFICANCE

The AI-TIRADS can enhance thyroid nodule risk stratification by improving diagnostic specificity and reducing unnecessary biopsies, potentially leading to more efficient patient management and better utilization of healthcare resources.

KEYWORDS

Thyroid Imaging Reporting and Data Systems, thyroid cancer, artificial intelligence, thyroid nodules, ultrasonography

Handling editor: Tuğba Akıncı D'Antonoli

Corresponding author: Koray Güven

E-mail: koray.guven@acibadem.com

Received 25 April 2025; revision requested 02 June 2025; accepted 27 July 2025.



Epub: 01.09.2025

Publication date: 01.07.2026

DOI: 10.4274/dir.2025.253428

You may cite this article as: Bozkuş A, Başar Y, Güven K. Comparison of the diagnostic performance of the artificial intelligence-based TIRADS algorithm with established classification systems for thyroid nodules. *Diagn Interv Radiol.* 2026;32(4):376-381.

Thyroid nodules are a widespread clinical concern, detected in a substantial proportion of the general population through high-resolution ultrasound (US) examinations. The frequency of thyroid nodules detected in people during routine US screenings varies between 20% and 68%. This range is influenced by factors such as demographic traits and the precision of the imaging technology employed.^{1,2} Although the majority of thyroid nodules are benign, approximately 5%–15% exhibit malignant potential, necessitating accurate risk stratification to guide clinical management.³ The most important challenge in the evaluation of thyroid nodules is the distinction between benign and malignant lesions to determine the appropriate necessity for invasive investigations such as fine needle aspiration biopsy (FNAB).^{4,5}

FNAB is a cornerstone diagnostic tool for thyroid nodules with high sensitivity and specificity in differentiating benign from malignant lesions. However, it is not without drawbacks, including being invasive, having the potential for nondiagnostic or indeterminate results, and causing patient discomfort.^{6,7} Consequently, there is a pressing need for non-invasive, reliable methods to enhance the accuracy of thyroid nodule classification, thereby reducing unnecessary biopsies and associated healthcare costs.⁸

US remains the primary imaging modality for the evaluation of thyroid nodules due to its accessibility, lack of ionizing radiation, and ability to provide detailed anatomical and structural information.⁹ These classifica-

tion systems risk-stratify the probability of malignancy based on specific sonographic features such as echogenicity, composition, shape, margin, and the presence of calcifications.^{10,11}

Multiple Thyroid Imaging Reporting and Data Systems (TIRADS) have been proposed by different organizations, including the American College of Radiology (ACR-TIRADS),¹² the American Thyroid Association (ATA-TIRADS),⁴ the European Thyroid Association (EU-TIRADS),¹³ and the Korean Thyroid Association (K-TIRADS).¹⁴ Despite their widespread adoption, variability exists among these systems in criteria weighting, risk categorization, and recommended management strategies, leading to inconsistencies in clinical practice.¹⁵ This lack of consensus highlights the need for further refinement and the potential integration of advanced technologies to enhance diagnostic performance.

Artificial intelligence (AI) and machine learning have been transformative in medical imaging over the past decade, with the promise of improving traditional diagnostic practices.¹⁶ The AI-based TIRADS (AI-TIRADS) leverage computational algorithms to analyze complex patterns in US data, aiming to improve the accuracy and consistency of thyroid nodule classification.¹⁷ Preliminary studies suggest that the AI-TIRADS may exhibit higher specificity and reduced rates of unnecessary FNABs compared with the conventional TIRADS, without compromising sensitivity.^{18,19} These advancements enable better thyroid nodule assessment, where AI-assisted decision-making can enhance clinical outcomes and optimize resource utilization.

Moreover, the integration of AI into the TIRADS addresses critical issues such as interobserver variability and subjective interpretation inherent in manual US evaluations.²⁰ By providing objective, reproducible assessments of nodular characteristics, the AI-TIRADS can standardize risk stratification across different healthcare settings and practitioners.²¹ This is particularly pertinent in regions with limited access to specialized radiologists, where AI-driven tools can support primary care providers in making informed decisions.²²

Despite the promising potential of the AI-TIRADS, comprehensive evaluations comparing its diagnostic performance against established TIRADS are limited. Furthermore, the impact of AI integration on clinical workflows, patient anxiety, and overall healthcare

costs warrants thorough investigation.²³ Addressing these gaps is essential to validate the efficacy of the AI-TIRADS and facilitate its widespread adoption in routine clinical practice.

This study aims to compare and evaluate the diagnostic effectiveness of the AI-TIRADS with other currently established classification systems, such as the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS, in characterizing thyroid nodules. By analyzing a large cohort of patients with confirmed cytopathological diagnoses, this research seeks to determine the efficacy of the AI-TIRADS in improving specificity, reducing unnecessary FNABs, and enhancing overall diagnostic accuracy.

Study sample

This retrospective investigation was conducted with the approval of the local ethics committee, which also waived the requirement for informed consent due to the use of de-identified medical records (REDACTED). Approved by the Acibadem University and Acibadem Healthcare Institutions Medical Research Ethics Committee (ATADEK) on June 16, 2023, with decision number 2023-10/360. The study included adult individuals who underwent thyroid US assessments at a single tertiary care institution between April 2016 and May 2022. Initially, 1,322 thyroid nodules from 1,139 patients with confirmed cytopathological diagnoses were identified via the hospital information system and the picture archiving and communication system. All images were reviewed in Digital Imaging and Communications in Medicine format. Patients lacking complete cytopathological information were excluded, resulting in a final cohort comprising 1,110 patients with 1,292 thyroid nodules. Further details are presented as a flowchart in Figure 1. The study adhered to the Standards for Reporting of Diagnostic Accuracy guidelines to ensure integrity and transparency in the reporting process.²⁴

All thyroid nodules were evaluated and classified according to five different TIRADS: the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, the K-TIRADS, and the AI-TIRADS algorithm.^{4,12-14,17}

Wildman-Tobriner et al.¹⁷ optimized the ACR-TIRADS classification using AI. Three board-certified radiologists with more than 5 years of experience in thyroid imaging independently reviewed all nodules' sonographic features, including echogenicity, composition, shape, margin, and calcifications. Dis-

Main points

- The artificial intelligence-based Thyroid Imaging Reporting and Data System (AI-TIRADS) exhibited enhanced specificity in comparison with other recognized systems such as the American College of Radiology (ACR)-TIRADS, the American Thyroid Association TIRADS, the European TIRADS, and the Korean TIRADS, concurrently preserving elevated sensitivity.
- The AI-TIRADS notably reduced the incidence of avoidable fine needle aspiration biopsies to 40.8%, compared with 49% for the ACR-TIRADS and exceeding 51% for the other systems.
- The investigation validated that the existence of multiple concerning ultrasound characteristics is considerably correlated with an elevated risk of malignancy, whereas singular features have diminished predictive value, thereby underscoring the necessity for thorough risk evaluation.

agreements were resolved by consensus to prevent variation in classification. Independent review for ground truth verification was not performed, as inter-rater reliability testing was beyond the scope of the study.

Cytopathological diagnoses were categorized based on the Bethesda System for Reporting Thyroid Cytopathology.²⁵ Nodules were classified as benign (Bethesda category 2), indeterminate (Bethesda categories 3 and 4), or malignant (Bethesda categories 5 and 6). Nodules diagnosed as non-diagnostic or having inadequate material (Bethesda category 1) were excluded from the analysis. This classification enabled a standardized assessment of malignancy risk across the different TIRADS.

Statistical analysis

Statistical analyses were performed using the R programming language (R Core Team, 2023) within the RStudio environment (RStudio Team, 2023). All analyses were two-tailed, and a *P* value of less than 0.05 was considered statistically significant.

Descriptive statistics were used to characterize the patient cohort. Continuous variables, such as age, were presented as mean $\pm$ standard deviation for normally distributed data or as median (minimum–maximum) for non-normally distributed data. Categorical variables were expressed as frequency and percentage.

Comparative evaluations among groups were conducted using appropriate statistical methods. Independent sample *t*-tests were applied to compare continuous variables between two groups, and analysis of variance alongside Tukey's post-hoc tests was used for comparisons among multiple groups. Categorical variables were analyzed using chi-square or Fisher's exact tests, as appropriate.

The performance of all TIRADS was assessed by calculating sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) to evaluate each system's accuracy in identifying malignant and benign nodules. Histopathology results with Bethesda scores of 2, 3, and 4 were considered negative, whereas scores of 5 and 6 were considered positive. For TIRADS classification, scores of 1, 2, and 3 were considered negative, and scores of 4 and 5 were considered positive. These metrics were computed using the caret package in R.²⁶ Additionally, confidence intervals for sensitivity, specificity, PPV, and NPV were calculated to assess the precision of these estimates. Differences

in sensitivity and specificity between classification methods were evaluated using the McNemar test. Comparisons of PPV and NPV were conducted using the statistical package of Stock et al.²⁷

Results

In this study, 1,324 nodules detected in 1,139 patients were evaluated, and 32 nodules in 29 patients were excluded due to non-diagnostic or insufficient cytopathological material detected by FNAB. A total of 1,110 patients with 1,292 nodules confirmed by cytopathological diagnosis were included. Of the 1,292 nodules included, 846 (65.4%) were diagnosed as benign, 299 (23.1%) as intermediate risk, and 147 (11.3%) as malignant. The overall incidence of malignancy was approximately 11.5%. Example images with evaluations are shown in Figure 2.

Of the patients, 826 (72.5%) were women and 313 (27.5%) were men. Among the 947 nodules in women, 629 (66.4%) were benign, 217 (22.9%) were intermediate risk, and 101 (10.6%) were malignant. Of the 345 nodules in men, 216 (62.6%) were benign, 81 (23.4%) were intermediate risk, and 48 (13.9%) were malignant. A statistically significant differ-

ence was found between malignancy rates in men and women (*P* < 0.001).

The average age of the patients was 46.33 ± 11.95 years; for women, it was 46.18 ± 11.94 years, and for men, it was 47.37 ± 11.57 years. The average age of patients with benign nodules was 47.04 ± 11.62 years, whereas those with intermediate-risk nodules had an average age of 46.46 ± 11.90 years. Patients diagnosed with malignant nodules were younger, with an average age of 40.97 ± 11.93 years (*P* < 0.001). The lower average age of patients with malignant nodules was observed. Further distribution of patient characteristics is summarized in Table 1.

The diagnostic performance parameters of the AI-TIRADS classification are as follows: sensitivity, specificity, PPV, and NPV, calculated as 95.3%, 53.8%, 21.2%, and 98.8%, respectively. The statistical analysis of malignancy risk according to the AI-TIRADS guidelines is shown in Table 2.

The AI-TIRADS missed 7 cancer cases out of 149 (4.6%), the ACR-TIRADS missed 5 cancer cases out of 149 (3.3%), the ATA-TIRADS missed 3 cancer cases out of 146 (2%), the EU-TIRADS missed 3 cancer cases out of 149

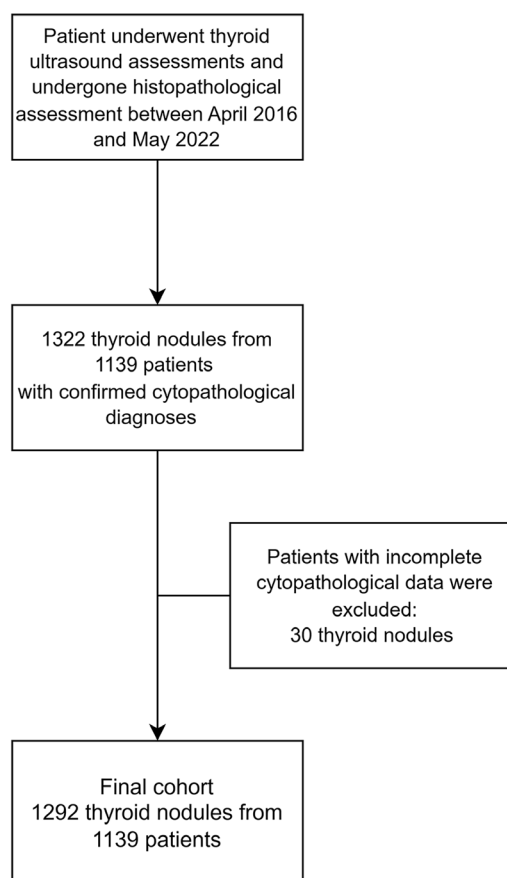


Figure 1. Flowchart of the study.

(2%), and the K-TIRADS missed 3 cancer cases out of 149 (2%).

The sensitivity and specificity for the AI-TIRADS were 94.6% and 53.6%, respectively. The AI-TIRADS showed no statistically significant difference from all other TIRADS in sensitivity but showed a statistically significant difference in specificity. In comparison, the ACR-TIRADS showed a sensitivity of 96.6% and a specificity of 44.6%. The ATA-TIRADS, the EU-TIRADS, and the K-TIRADS exhibited similar sensitivities of 97.9% but lower specificities ranging from 39.3% to 40.1%.

The PPV for the AI-TIRADS was 21.2%, which was statistically significantly higher than those of the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS, which were 17.4%, 17.9%, and 17.9%, respectively. The PPV for the AI-TIRADS was not statistically significantly different from that of the ACR-TIRADS (18.5%). Unnecessary biopsy rates were interpreted based on PPVs (higher is better). The NPV for the AI-TIRADS was 98.8% and was not statistically significantly different from those of the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS, which were 99%, 99.3%, 99.3%, and 99.3%, respectively. Examples of false-positive images are provided in Figure 3. The diagnostic performances of the classification systems used in our study are shown in Table 3.

Discussion

This study compared the diagnostic accuracy of the AI-TIRADS with other classification systems, such as the ACR-TIRADS, the ATA-TIRADS, the EU-TIRADS, and the K-TIRADS, for thyroid nodule characterization. The study revealed that the AI-TIRADS had a sensitivity of 94.6% and a specificity of 53.6%, surpassing the other systems in specificity while maintaining comparable sensitivity levels. This indicates that the AI-TIRADS can effectively reduce unnecessary FNABs without compromising the detection of malignant nodules.

The higher specificity of the AI-TIRADS aligns with recent studies suggesting that AI-enhanced algorithms improve risk stratification accuracy by minimizing false-positive results.^{17,18} The reduction in unnecessary biopsies observed with the AI-TIRADS (PPV 21.2%) compared with the ACR-TIRADS (PPV 18.5%) and other systems (PPV <18%) underscores its potential to optimize clinical workflows and alleviate patient burden.¹²

Consistent with our findings, previous research has indicated that AI-based systems

can enhance diagnostic performance by accurately analyzing complex US features beyond human capability.^{28,29} Additionally, the lack of substantial differences in malignancy rates between genders and the association of multiple suspicious US features with increased cancer risk corroborate existing literature.^{4,5}

The superior specificity of the AI-TIRADS (53.6%) compared with other systems such

as the ACR-TIRADS (44.6%), the ATA-TIRADS (39.3%), the EU-TIRADS (40.1%), and the K-TIRADS (40.1%) is crucial in this context, as it suggests a greater ability to correctly classify benign nodules within this group, thereby reducing the rate of unnecessary FNABs. Wildman-Tobriner et al.¹⁷ reported findings consistent with ours, highlighting the better specificity of the AI-TIRADS. When intermediate-risk lesions (Bethesda 3 and 4) were clas-

Table 1. Distribution of patient characteristics across different histopathological diagnoses

Patient characteristics	Benign		Malignant	P value
	Low-risk	Intermediate risk		
Sex	-	-	-	<0.001
Female	629 (66.4%)	217 (22.9%)	101 (10.7%)	-
Male	216 (62.6%)	81 (23.5%)	48 (13.9%)	-
Age	47.04 ± 11.62	46.46 ± 11.90	40.97 ± 11.93	<0.001

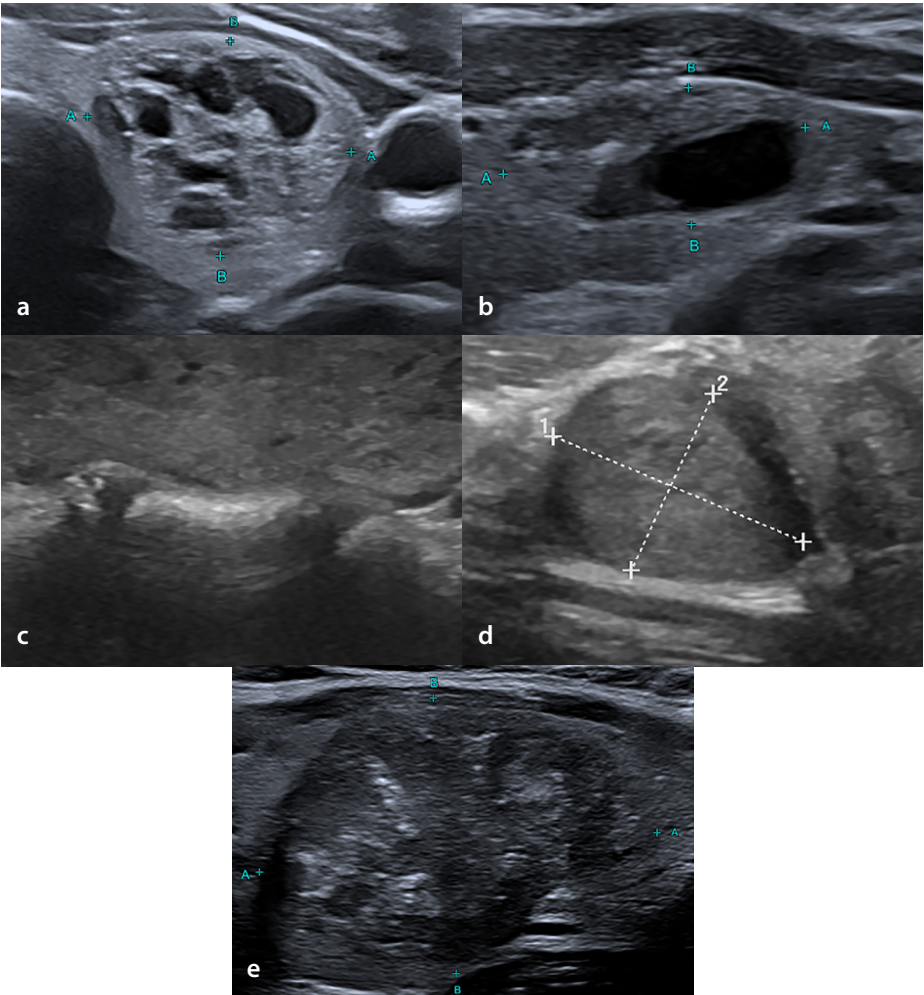


Figure 2. Examples of evaluated images. (a) AI-TIRADS 1, ACR-TIRADS 2, EU-TIRADS 2, K-TIRADS 2, and ATA-TIRADS 2 lesions. Histopathological evaluation was benign (Bethesda category 2). (b) AI-TIRADS 2, ACR-TIRADS 3, EU-TIRADS 2, K-TIRADS 2, and ATA-TIRADS 2 lesions. Histopathological evaluation was benign (Bethesda category 2). (c) AI-TIRADS 3, ACR-TIRADS 3, EU-TIRADS 3, K-TIRADS 3, and ATA-TIRADS 3 lesions. Histopathological evaluation was benign (Bethesda category 2). (d) AI-TIRADS 4, ACR-TIRADS 4, EU-TIRADS 4, K-TIRADS 4, and ATA-TIRADS 4 lesions. Histopathological evaluation was benign (Bethesda category 2). (e) AI-TIRADS 5, ACR-TIRADS 5, EU-TIRADS 5, K-TIRADS 5, and ATA-TIRADS 5 lesions. Histopathological evaluation was malignant (Bethesda category 6). AI-TIRADS, artificial intelligence-based Thyroid Imaging Reporting and Data System; ACR, American College of Radiology; EU, European Thyroid Association TIRADS; K-TIRADS, Korean Thyroid Association TIRADS; ATA, American Thyroid Association.

sified as positive, the AI-TIRADS demonstrated a PPV of 48%, which was higher than that of the other systems (43%–45%). Although other systems showed higher sensitivities in this analysis (78%–81% compared with 72% for the AI-TIRADS), they did so at the cost of lower precision and PPV. This trade-off underscores the clinical utility of the AI-TIRADS

in the intermediate-risk category, where avoiding unnecessary biopsies for benign nodules is a primary goal.

However, this study has limitations, including its retrospective design, which may lead to selection bias and potentially restrict the generalizability of the findings. In particular, it must be acknowledged that patients with histopathology results of Bethesda score 1 and only biopsied patients were included in the study. Additionally, the lack of an interobserver reliability assessment might impact the consistency of nodule classification.³⁰ To confirm the effectiveness of the AI-TIRADS and examine its incorporation into standard clinical practice, future prospective studies involving larger and more diverse populations are needed.

Moreover, variability among different TIRADS in criteria weighting and risk categorization highlights the need for standardized guidelines to ensure consistent application across healthcare settings.^{31,32} The integration of AI into the TIRADS offers a promising solution to these challenges by providing objective and reproducible assessments, thereby enhancing diagnostic accuracy and reducing interobserver variability.³³

The AI-TIRADS demonstrated improved specificity and a reduced unnecessary biopsy rate in thyroid nodule classification without sacrificing high sensitivity compared with other traditional TIRADS. The findings suggest that the AI-TIRADS can be utilized to enhance clinical decision-making, optimize resource utilization, and improve patient management in thyroid nodule assessment. Further prospective studies are required to confirm these findings and facilitate broader implementation of the AI-TIRADS in clinical practice.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

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(8):699-706. [CrossRef]
2. Tan GH, Gharib H. Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. *Ann Intern Med*. 1997;126(3):226-231. [CrossRef]

Table 2. Distribution of the AI-TIRADS classifications					
Category	Benign		Malignant (n = 149)	Malignancy risk	P value
	Low-risk (n = 845)	Intermediate risk (n = 298)			
AI-TIRADS 1	153 (18.1%)	16 (5.3%)	0 (0%)	0%	<0.001
AI-TIRADS 2	228 (27.0%)	47 (15.8%)	2 (1.3%)	0.7%	-
AI-TIRADS 3	116 (13.7%)	53 (17.8%)	5 (3.3%)	2.9%	-
AI-TIRADS 4	265 (31.3%)	154 (51.7%)	35 (22.8%)	7.7%	-
AI-TIRADS 5	83 (9.9%)	28 (9.4%)	107 (72%)	49%	-

AI, Artificial intelligence; TIRADS, Thyroid Imaging Reporting and Data System.

Table 3. Diagnostic performance of the classification systems used								
Classifications System	Sensitivity	P value*	Precision	P value*	PPV	P value*	NPV	P value
AI-TIRADS	95.3%	-	53.8%	-	21.2%	-	98.8%	-
ACR-TIRADS	96.6%	0.39	44.6%	<0.001	18.5%	0.07	99%	0.62
EU-TIRADS	97.9%	0.12	40.1%	<0.001	17.9%	0.03	99.3%	0.32
K-TIRADS	97.9%	0.12	40.1%	<0.001	17.9%	0.03	99.3%	0.32
ATA	97.9%	0.12	39.3%	<0.001	17.4%	0.04	99.3%	0.32

*Compared with the AI-TIRADS. AI, artificial intelligence; TIRADS, Thyroid Imaging Reporting and Data Systems; ACR, American College of Radiology; EU-TIRADS, European Thyroid Association TIRADS; K-TIRADS, Korean Thyroid Association TIRADS; ATA, American Thyroid Association; PPV, positive predictive value; NPV, negative predictive value.

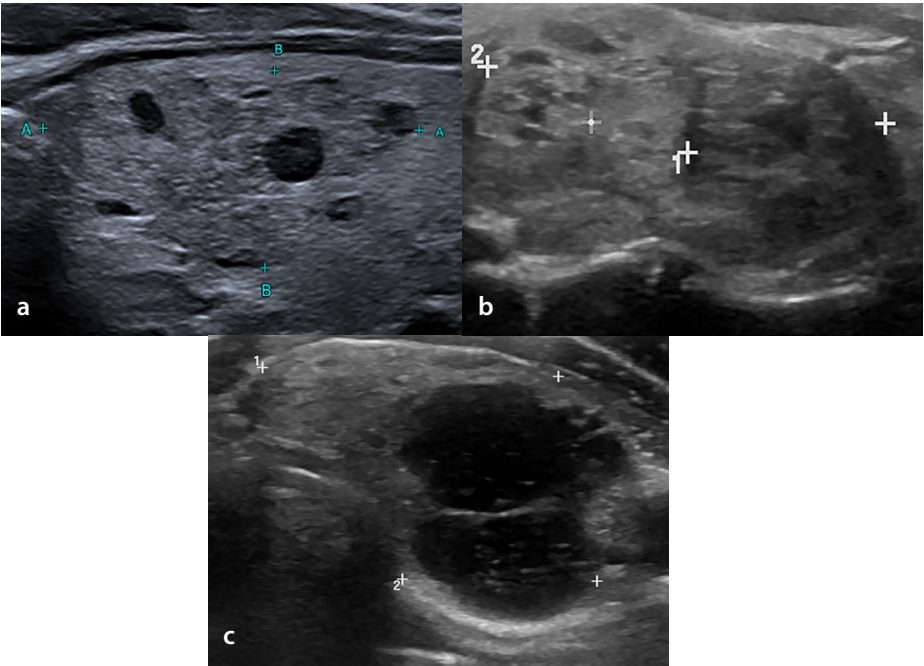


Figure 3. Examples of false-positive images. (a) AI-TIRADS 2, ACR-TIRADS 4, EU-TIRADS 5, K-TIRADS 5, and ATA-TIRADS 5 lesions. Histopathological evaluation was benign (Bethesda category 2). Biopsy could have been avoided with the AI-TIRADS. (b) AI-TIRADS 2, ACR-TIRADS 4, EU-TIRADS 5, K-TIRADS 5, and ATA-TIRADS 5 lesions. Histopathological evaluation was benign (Bethesda category 2). Biopsy could have been avoided with the AI-TIRADS. (c) AI-TIRADS 2, ACR-TIRADS 4, EU-TIRADS 5, K-TIRADS 5, and ATA-TIRADS 5 lesions. Histopathological evaluation was benign (Bethesda category 2). Biopsy could have been avoided with the AI-TIRADS. AI-TIRADS, artificial intelligence-based Thyroid Imaging Reporting and Data System; ACR, American College of Radiology; EU, European Thyroid Association TIRADS; K-TIRADS, Korean Thyroid Association TIRADS; ATA, American Thyroid Association.

3. Frates MC, Benson CB, Doubilet PM, et al. Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. *J Clin Endocrinol Metab.* 2006;91(9):3411-3417. [\[CrossRef\]](#)
4. American Thyroid Association (ATA) Guidelines taskforce on thyroid nodules and differentiated thyroid cancer; Cooper DS, Doherty GM, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid.* 2009;19(11):1167-1214. [\[CrossRef\]](#)
5. Gharib H, Papini E, Paschke R, et al. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association Medical Guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. *Endocr Pract.* 2010;16(3):468-475. [\[CrossRef\]](#)
6. Yeh MW, Demircan O, Ituarte P, Clark OH. False-negative fine-needle aspiration cytology results delay treatment and adversely affect outcome in patients with thyroid carcinoma. *Thyroid.* 2004;14(3):207-215. [\[CrossRef\]](#)
7. Weber AL, Randolph G, Aksoy FG. The thyroid and parathyroid glands. CT and MR imaging and correlation with pathology and clinical findings. *Radiol Clin North Am.* 2000;38(5):1105-1129. [\[CrossRef\]](#)
8. Burch HB. Evaluation and management of the solid thyroid nodule. *Endocrinol Metab Clin North Am.* 1995;24(4):663-710. [\[CrossRef\]](#)
9. Gharib H. Fine-needle aspiration biopsy of thyroid nodules: advantages, limitations, and effect. *Mayo Clin Proc.* 1994;69(1):44-49. [\[CrossRef\]](#)
10. Horvath E, Majlis S, Rossi R, et al. An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *J Clin Endocrinol Metab.* 2009;94(5):1748-1751. [\[CrossRef\]](#)
11. Kwak JY, Han KH, Yoon JH, 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(3):892-899. [\[CrossRef\]](#)
12. Li X, Hou XJ, Du LY, et al. Virtual touch tissue imaging and quantification (VTIQ) combined with the American College of Radiology thyroid imaging reporting and data system (ACR TI-RADS) for malignancy risk stratification of thyroid nodules. *Clin Hemorheol Microcirc.* 2019;72(3):279-291. [\[CrossRef\]](#)
13. 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(5):225-237. [\[CrossRef\]](#)
14. Ha EJ, Na DG, Baek JH. Korean thyroid imaging reporting and data system: current status, challenges, and future perspectives. *Korean J Radiol.* 2021;22(9):1569-1578. [\[CrossRef\]](#)
15. Grani G, Lamartina L, Ascoli V, et al. Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: toward the "right" TIRADS. *J Clin Endocrinol Metab.* 2019;104(1):95-102. [\[CrossRef\]](#)
16. Chen Y, Gao Z, He Y, et al. An artificial intelligence model based on ACR TI-RADS characteristics for US diagnosis of thyroid nodules. *Radiology.* 2022;303(3):613-619. [\[CrossRef\]](#)
17. Wildman-Tobriner B, Buda M, Hoang JK, et al. Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: diagnostic accuracy and utility. *Radiology.* 2019;292(1):112-119. [\[CrossRef\]](#)
18. Middleton WD, Teefey SA, Reading CC, et al. Comparison of performance characteristics of American College of Radiology TI-RADS, Korean Society of Thyroid Radiology TIRADS, and American Thyroid Association Guidelines. *AJR Am J Roentgenol.* 2018;210(5):1148-1154. [\[CrossRef\]](#)
19. Lauria Pantano A, Maddaloni E, Briganti SI, et al. Differences between ATA, AACE/ACE/AME and ACR TI-RADS ultrasound classifications performance in identifying cytological high-risk thyroid nodules. *Eur J Endocrinol.* 2018;178(6):595-603. [\[CrossRef\]](#)
20. Shah JP. Thyroid carcinoma: epidemiology, histology, and diagnosis. *Clin Adv Hematol Oncol.* 2015;13(4 Suppl 4):3-6. [\[CrossRef\]](#)
21. Borlea A, Borcan F, Sporea I, et al. TI-RADS diagnostic performance: which algorithm is superior and how elastography and 4D vascularity improve the malignancy risk assessment. *Diagnostics (Basel).* 2020;10(4):180. [\[CrossRef\]](#)
22. Nam-Goong IS, Kim HY, Gong G, et al. Ultrasonography-guided fine-needle aspiration of thyroid incidentaloma: correlation with pathological findings. *Clin Endocrinol (Oxf).* 2004;60(1):21-28. [\[CrossRef\]](#)
23. Ayvaz E, Kaplan K, Kuncan F, Ayvaz E, Türkoğlu H. Reducing operation costs of thyroid nodules using machine learning algorithms with thyroid nodules scoring systems. *Applied Sciences.* 2022;12(22):11559. [\[CrossRef\]](#)
24. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527. [\[CrossRef\]](#)
25. Acharya K, Shrivastav S, Triipathi P, et al. The Bethesda system for reporting thyroid cytopathology: validating at tribhuvan university teaching hospital. *Int Arch Otorhinolaryngol.* 2021;26(1):e097-e102. [\[CrossRef\]](#)
26. Kuhn M, Wing J, Weston S, et al. CareT: Classification and Regression Training. Published online March 21, 2023. Accessed September 15, 2024. [\[CrossRef\]](#)
27. Stock C, Hielscher T, Discacciati A. DTComPair: comparison of binary diagnostic tests in a paired study design. Published online August 20, 2023. Accessed January 2, 2024. [\[CrossRef\]](#)
28. Castellana M, Castellana C, Treglia G, et al. Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA. *J Clin Endocrinol Metab.* 2020;105(5):dgz170. [\[CrossRef\]](#)
29. Zhang J, Zhang X, Meng Y, Chen Y. Contrast-enhanced ultrasound for the differential diagnosis of thyroid nodules: an updated meta-analysis with comprehensive heterogeneity analysis. *PLoS One.* 2020;15(4):e0231775. [\[CrossRef\]](#)
30. Hoang JK, Middleton WD, Farjat AE, et al. Reduction in thyroid nodule biopsies and improved accuracy with American College of Radiology thyroid imaging reporting and data system. *Radiology.* 2018;287(1):185-193. [\[CrossRef\]](#)
31. Cappelli C, Castellano M, Pirola I, et al. The predictive value of ultrasound findings in the management of thyroid nodules. *QJM.* 2007;100(1):29-35. [\[CrossRef\]](#)
32. Bakhshaei M, Davoudi Y, Mehrabi M, et al. Vascular pattern and spectral parameters of power Doppler ultrasound as predictors of malignancy risk in thyroid nodules. *Laryngoscope.* 2008;118(12):2182-2186. [\[CrossRef\]](#)
33. Cao CL, Li QL, Tong J, et al. Artificial intelligence in thyroid ultrasound. *Front Oncol.* 2023;13:1060702. [\[CrossRef\]](#)