


Comparative Analysis of AI-SONICTM Thyroid System and Six Thyroid Risk Stratification Guidelines in Papillary Thyroid Cancer: A Retrospective Cohort Study

Mingyan Wang^{1-3,*}, Siyuan Yang^{1-3,*}, Linxin Yang¹⁻³, Ning Lin¹⁻³ 

¹Ultrasound Department of Shengli Clinical Medical College of Fujian Medical University, Fuzhou City, Fujian Province, 350001, People's Republic of China; ²Ultrasound Department of Fujian Provincial Hospital, Fuzhou City, Fujian Province, 350001, People's Republic of China; ³Ultrasound Department of Fuzhou University Affiliated Provincial Hospital, Fuzhou City, Fujian Province, 350001, People's Republic of China

*These authors contributed equally to this work

Correspondence: Ning Lin, Ultrasound Department of Fujian Provincial Hospital, 134 East Street, Fuzhou City, 350001, Fujian Province, People's Republic of China, Tel +86 0591-87557768, Email NingLin2413@163.com

Aim: The study aimed to compare the diagnostic performance of AI-SONICTM Thyroid System (AI-SONICTM) with six thyroid nodule ultrasound risk stratification systems, as well as the interobserver agreement among different-year ultrasound examiners using the same diagnostic approach.

Methods: This retrospective study included patients who underwent thyroid ultrasound examination and surgery between 2010 and 2022. Three ultrasound examiners with 2, 5, and 10 years of experience, respectively, used AI-SONICTM and six guidelines to risk-stratify the nodules. The diagnostic performance and interobserver agreement were assessed.

Results: A total of 370 thyroid nodules were included, including 195 papillary thyroid carcinomas (PTC) and 175 benign nodules. For physicians of varying seniority from low to high, AI-SONICTM had a moderate sensitivities of 82.56%, 83.08%, 84.62%, respectively, while AACE/ACE/AME had the highest diagnostic sensitivities (96.41%, 95.38%, 96.41%, respectively); And relatively higher specificities were 85.14%, 85.71%, 85.71% for KSThR, while moderate specificities with values of 84.0%, 85.14%, and 85.71%, respectively were found for AI-SONICTM; The accuracy was highest for ATA (excluding non-classifiable nodules), with values of 87.26%, 87.93%, and 88.82%, respectively, while the accuracy for AI-SONICTM were 83.24%, 84.05%, and 85.14%, respectively. The Kendall's tau coefficient indicated strong or moderate interobserver agreement among all examiners using different diagnostic methods (Kendall's tau coefficient >0.6, P<0.001). AI-SONICTM showed the highest interobserver agreement (Kendall's tau coefficient=0.995, P<0.001). A binary probit regression analysis showed that nodules with cystic components had a significantly higher regression coefficient value of 0.983 (P=0.002), indicating that AI-SONICTM may have higher accuracy for nodules with cystic components.

Conclusion: AI-SONICTM and the six thyroid nodule ultrasound risk stratification systems showed high diagnostic performance for papillary thyroid carcinoma. All examiners showed strong or moderate interobserver agreement when using different diagnostic methods. AI-SONICTM may have higher accuracy for nodules with cystic components.

Keywords: AI-SONICTM thyroid system, six thyroid risk stratification guidelines, papillary thyroid cancer

Introduction

The thyroid gland, is one of the largest endocrine glands in the human body. The prevalence rate of thyroid lesions is 4%–7%, which is often found on the gland. Most of thyroid lesions are asymptomatic, and thyroid hormone secretion is normal.¹ Thyroid cancer is one of the most common malignant tumors in the endocrine system. Approximately 5% of adult thyroid nodule patients have no history of radiation exposure, and the proportion of malignant nodules is relatively high.² Thyroid cancer, a prevalent malignant tumor, has been on the rise worldwide in recent years. Papillary thyroid

cancer (PTC) accounts for approximately 90% of all thyroid malignancies,³ making it the most frequently encountered malignant tumor of the thyroid gland. Ultrasonography, as the primary diagnostic modality for thyroid nodules, plays a crucial role in the early detection and management of thyroid cancer.

To standardize thyroid ultrasound exams and reports, multiple professional associations have issued guidelines with varying approaches to thyroid nodule risk stratification. These include six commonly recognized versions, namely the 2015 American Thyroid Association (ATA) guidelines,⁴ the 2016 guidelines from the American Association of Clinical Endocrinologists (AACE) and the American College of Endocrinology (ACE), as well as those from the Italian Association of Clinical Endocrinologists (AME),⁵ the 2017 American College of Radiology (ACR) guidelines,⁶ the 2017 European Thyroid Association (ETA) guidelines,⁷ and the 2020 guidelines from both the Chinese Medical Association's Ultrasound Medical Branch⁸ and the Korean Society of Thyroid Radiology (KSThR).⁹ These guidelines provide reliable diagnostic tools for ultrasound physicians in effectively identifying thyroid cancer.

With the rapid advancement of artificial intelligence (AI) technology, its application in the medical field has been increasingly recognized as a potential game-changer.¹⁰ Among them, medical imaging, particularly in the context of thyroid nodule diagnosis, has emerged as a highly promising sub-specialty. A key example of this is the AI-SONICTM Thyroid System (AI-SONICTM) developed by China's DeShangYunXing Imaging Technology Co., Ltd.¹¹ which utilizes deep learning techniques and an independently modified convolutional neural network (CNN) to demonstrate superior performance in the diagnosis of thyroid nodules. The system was trained using a dataset of 200,000 thyroid nodules expert-annotated and has shown the capability to accurately and rapidly identify lesions and evaluate their malignancy using a large volume of data.¹² This approach has significantly improved diagnostic speed and accuracy, potentially filling the diagnostic gap left by human doctors. Thyroid ultrasound examination, a common screening tool in population-based health checks, imposes a heavy workload on medical professionals. However, due to limitations in diagnostic accuracy, workload demands, and fatigue, instances of misdiagnosis or missed diagnosis are not uncommon. The objective, consistent, and efficient nature of AI diagnosis can effectively mitigate these challenges, enhancing diagnostic accuracy and efficiency.

There have been many articles comparing the diagnostic efficacy of different guidelines, but very few comprehensive comparisons of different guideline diagnostic tools with AI diagnostic tools, especially the interobserver agreement among different-year ultrasound examiners using the same diagnostic tools. There are more than 20 thyroid ultrasound risk stratification systems in clinical use. AI-SONICTM, as a new diagnostic tool, is newly applied in clinical practice. Six representative diagnostic tools used by the vast majority of sonographers, which were by various professional organisations and widely recognised, were selected to compare with AI-SONICTM in this study. Does AI-SONICTM, as a new diagnostic tool, have the same diagnostic capabilities as the diagnostic tools currently used by human physicians? Can AI-SONICTM be expected to make up for the shortcomings of human physicians and thus improve the efficiency and accuracy of diagnosis?

Therefore, this study aimed to assess the utility of different thyroid nodule ultrasound risk stratification systems in predicting nodule malignancy and compare their diagnostic efficacies, thus to evaluate the diagnostic accuracy of the AI-SONICTM in thyroid nodule diagnosis. Moreover, we will evaluate the inter-observer consistency among doctors with different experience levels using the same guidance in their diagnosis, in order to explore the impact of baseline and ultrasound characteristics of the nodule on the AI system's diagnostic accuracy.

Materials and Methods

Research Subjects

A retrospective selection of 370 thyroid nodule patients who underwent thyroid ultrasound examination and surgical treatment at Fujian Provincial Hospital between 2010 and 2022 were included. Inclusion criteria: 1) Complete clinical data and relevant examinations of the patients; 2) Complete, clear ultrasound images of thyroid nodules with accurate ultrasound report descriptions; 3) Patients underwent thyroid surgery at our hospital with clear postoperative pathological results, and the pathological types of malignant nodules only included papillary thyroid carcinoma; 4) For multiple lesions, only one nodule was selected, and the selected thyroid nodule could be corresponding to its pathological results

one by one. The exclusion criteria were as follows: (1) unclear thyroid ultrasound image data like unclear and incomplete archival images; (2) no clear pathological diagnosis results; (3) incomplete relevant baseline data.

Data Collection

Initially, three ultrasound physicians with 2, 5, and 10 years of experience respectively jointly studied six guidelines, proficiently grasping the similarities and differences of the “ultrasound feature dictionary” and risk stratification systems of each guideline. The three physicians independently analyzed the ultrasound images of thyroid nodules, recording the baseline and ultrasound characteristics of the nodules. Baseline characteristics included patient gender, age, presence or absence of diffuse thyroid disease, nodule location, and size. Ultrasound characteristics encompassed nodule structure, echo, margin, boundary, aspect ratio, calcification, and extrathyroidal extension. The six thyroid risk stratification systems were used to stratify the risk of thyroid nodules. Subsequently, the three physicians individually selected a representative ultrasound grayscale image for each nodule, imported it into the AI-SONICTM, which automatically located and framed the thyroid nodule region in the image, promptly providing a benign or malignant probability value. None of the three physicians were aware of the patient’s clinical data, pathological results, or the diagnoses made by other physicians.

AI-SONICTM Thyroid System Operating Procedure

The AI-SONICTM Thyroid system was developed by Zhejiang DeShangYunXing Imaging Technology Co., Ltd., China,¹¹ which utilizes deep learning techniques and an independently modified convolutional neural network (CNN) to demonstrate superior performance in the diagnosis of thyroid nodules. The system was trained using a dataset of 200,000 thyroid nodules expert-annotated and has shown the capability to identify lesions and evaluate their malignancy using a large volume of data accurately and rapidly.¹² The three physicians collected and screened the 2D static images of thyroid nodules in DICOM format that meet the inclusion criteria from the ultrasound working system of our hospital, then opened the AI-SONICTM Thyroid Diagnostic System, imported the images into the system. The interface of AI-SONICTM Thyroid Diagnostic System will display the images and identify each thyroid nodule lesion in the image immediately and automatically. For example, the system framed each nodule along the edges of the nodule on the image, and display the scores of the nodules on the top of the border. Scores of 0–0.4 were considered as possibly benign, while scores of 0.41–0.99 were considered as possibly malignant (with scores of 0.41–0.60 in a moderately suspicious state and scores of 0.65–0.99 indicating a higher likelihood of malignancy). In a small number of cases, the AI detection system’s automatic identification of thyroid nodules is not accurate enough, such as: The large size of the nodules, the blurred boundary of the nodules, the extremely uneven internal echo of the nodules, the nodules breaking through the edge of the thyroid envelope, the uneven echo of the thyroid parenchyma, the adjacent growth of multiple nodules, and the obvious rear sound shadow caused by internal calcification may all lead to the inconsistency between the outline edge and the actual edge of the nodules. In this case, the nodes automatically selected by the system were incorrect, the physician could select “Manual Outline” from the system toolbar to re-outline the nodes along their edges, and the system will automatically display the new scores.

The Definition and Meaning of the Image

Boundaries of the nodule: whether the nodule is well demarcated from the surrounding thyroid parenchyma is classified as well demarcated or poorly demarcated.

Nodule margins: the shape of the margins of the thyroid nodule, either smooth or lobulated/irregular.

Thyroid nodule size: maximum width diameter measured in transverse section, maximum length and thickness diameter measured in longitudinal section, measurement includes the acoustic halo of the nodule;

Nodule location: classified as left lobe, right lobe and isthmus

Structure: classified as cystic, solid, or cystic-solid (cystic: solid component <5%; solid: cystic component <5%; cystic-solid: 5% < solid component <95%, different guidelines have different bases for classifying the structure of the nodule, and the structure of the nodule in the grading of each guideline shall be based on that guideline;

Texture: homogeneous and heterogeneous echoes within the nodule;

Internal echoes: classified as very hypoechoic, hypoechoic, isoechoic, and hyperechoic; very hypoechoic if the echoes are lower than the cervical strap muscles, hypoechoic if the echoes are lower than the echoes of the thyroid parenchyma, isoechoic if the echoes are the same as the echoes of the thyroid parenchyma, and hyperechoic if the echoes are higher than the thyroid parenchyma echoes;

Boundaries: the boundary between the nodule and the thyroid parenchyma is clear or blurred;

Margins: regular: round and oval; irregular: lobulated or acicular margins;

Coarse calcification: rough and strong echoes within the nodule ≥ 1 mm in diameter, may be accompanied by acoustic shadows.

Microcalcifications: fine strong echoes < 1 mm in diameter scattered in the nodule;

Aspect ratio: assessed in transverse or longitudinal section, < 1 , > 1 or $= 1$;

Extrathyroidal extension: contiguous to the peritoneum: the nodule is adjacent to the thyroid peritoneum; encroachment on the peritoneum: interruption of continuity where the thyroid peritoneum meets the nodule; disruption of the periphery: the nodule breaks through the peritoneum and encroaches on the surrounding tissues;

Blood flow is typed by different guidelines and is summarised in [Supplemental Table 1](#).

Equipment and Examination Methods

The equipment below were applied in this study: ultrasound diagnostic machine (Philips iU22, EPIQ5/EPIQ7, Siemens ACUSON S2000, S3000), and 6~13MHz line array probe. The patients were examined by three sonographers in accordance with the code of practice, and the isthmus and bilateral lobes of the thyroid gland are examined in transverse and longitudinal views, and two-dimensional images of the largest sections of abnormal nodules and blood flow maps are retained.

Statistics

Data analysis was performed using SPSS 26.0 and MedCalc 22.009 software. Normally distributed measurement data were expressed as $x \pm s$, with comparisons between groups using one-way ANOVA; non-normally distributed measurement data were represented by median and interquartile range, with comparisons between groups using Kruskal–Wallis *H*-tests. Count data were expressed in case numbers and percentages, with comparisons between groups using Pearson's chi-square test or Fisher's exact probability method. ROC curves were used to evaluate the diagnostic efficacy of various diagnostic methods, selecting the point with the highest Youden index as the diagnostic cutoff value and calculating sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and the area under curve (AUC). Kendall's coefficient was used to test the consistency of different observers diagnosing PTC using the same method, with a Kendall coefficient > 0.8 indicating excellent consistency. Binary Probit regression analysis was used to investigate the impact of different characteristics on the diagnostic accuracy of the AI-SONICTM. A *P*-value < 0.05 was considered statistically significant.

Results

Baseline Characteristics

A total of 370 thyroid nodules were included in this study, of which 195 were PTC and 175 were benign nodules, and the pathological types of benign nodules included nodular goitre (147), follicular tumour (14) and adenoma (14). There were 52 patients with thyroid nodules combined with diffuse thyroid lesions (39 cases of PTC and 13 cases of benign nodules) in this study, and the pathological types included Hashimoto's thyroiditis (35 cases of PTC and 11 cases in the benign nodules group), hyperthyroidism (3 cases in the PTC group and 1 case in the benign nodules group), hyperthyroidism combined with Hashimoto's thyroiditis (1 case in the PTC group), and subacute thyroiditis (1 case in the benign nodules group). There was no significant difference in age and gender between the two groups in this study. Benign thyroid nodules were larger in all diameters and volumes than PTC ($P < 0.001$), and those located in the middle and lower parts of the thyroid gland were more often benign ($P < 0.001$), while PTC was more often combined with diffuse thyroid lesions, as shown in [Table 1](#).

Table 1 Comparison of General Characteristics

Parameters	PTC Group	Benign Group	P-value
Age (years)	46.83±12.03	48.22±12.62	0.276
Gender (female)	155(79.49)	128(73.14)	0.151
Nodule Location [#]			
Upper	68(34.87)	60(34.48)	0.938
Middle	85(43.59)	132(75.86)	<0.001*
Lower	63(32.31)	124(70.86)	<0.001*
Isthmus	13(6.70)	10(5.71)	0.695
Nodule Dimensions (mm)			
Length	8.700(5.7,15.9)	28.750(17.5,38.4)	<0.001*
Width	7.875(5.4,13.2)	22.300(14.9,29.7)	<0.001*
Height	8.520(5.9,11.6)	16.600(11.5,22.1)	<0.001*
Nodule Volume [§] (cm ³)	0.300(0.1,1.3)	5.745(1.6,12.5)	<0.001*
Concomitant Diffuse Thyroid Disease	39(20.00)	13(7.43)	<0.001*

Notes: *P<0.05. [#]If a nodule is large and involves multiple locations, each involved location is counted once. [§]Nodule volume was estimated using the formula "Volume = Length x Width x Height x 0.52".

Abbreviation: PTC, papillary thyroid carcinomas.

Ultrasonic Characteristics of Thyroid Nodules

In this study, the differences in ultrasound features were statistically significant ($P < 0.001$), except for gross calcification. The PTC group was more likely to have grey-scale ultrasound features such as realistic structures, more homogeneous echogenicity, hypoechoic or very hypoechoicity, indistinct borders, lobulated or acicular margins, aspect ratio > 1 , microcalcification, and extra-thyroidal extensions, than the benign nodule group. In terms of colour Doppler ultrasound features, the PTC group was more likely to have a flow fraction with a lack of blood supply, see [Table 2](#).

Table 2 Gray-Scale Ultrasonic Features and Blood Flow Patterns of Thyroid Nodules

Ultrasonic Features of Thyroid Nodules		PTC Group (%)	Benign Nodule Group (%)	P-value
Structure	Cyst-solid	3(1.54)	60(34.29)	<0.001 [#]
	Cystic	0(0.00)	13(7.43)	
	Solid	192(98.46)	102(58.29)	
Texture	Heterogeneous	143(73.33)	156(89.14)	<0.001 [#]
	Homogeneous	52(26.67)	19(10.86)	
Echo Level	Low	120(61.54)	44(25.14)	<0.001 [#]
	None	0(0.00)	4(2.29)	
	Very low	52(26.67)	5(2.86)	
	Equal	23(11.79)	120(68.57)	
Boundary	High	0(0.00)	2(1.14)	<0.001 [#]
	Unclear	37(18.97)	8(4.57)	
	Clear	158(81.03)	167(95.43)	
Margin	Lobulated/Spiculated	162(83.08)	22(12.57)	<0.001 [#]
	Regular	33(16.92)	153(87.43)	
Aspect Ratio	<1	98(50.26)	159(90.86)	<0.001 [#]
	=1	1(0.51)	0(0.00)	
	>1	96(49.23)	16(9.14)	
*Extrathyroidal Extension	None	84(43.08)	149(85.14)	<0.001 [#]
	Adjacent to capsule	58(29.74)	22(12.57)	
	Invasion of capsule	34(17.44)	3(1.71)	
	Disruption of surrounding area	19(9.74)	1(0.57)	

(Continued)

Table 2 (Continued).

Ultrasonic Features of Thyroid Nodules		PTC Group (%)	Benign Nodule Group (%)	P-value
Coarse Calcification	Present	56(28.72)	38(21.71)	0.122
	Absent	139(71.28)	137(78.29)	
Microcalcification	Present	97(49.74)	33(18.86)	<0.001 [#]
	Absent	98(50.26)	142(81.14)	
KSThR Blood Flow	I	50(25.64)	26(14.86)	<0.001 [#]
	2	55(28.21)	35(20.00)	
	3	53(27.18)	43(24.57)	
	4	37(18.97)	71(40.57)	
ETA Blood Flow	I	50(25.64)	26(14.86)	<0.001 [#]
	II	108(55.38)	78(44.57)	
	III	12(6.15)	6(3.43)	
	Unable to classify	25(12.82)	65(37.14)	
C-TIRADS Blood Flow	I	50(26.46)	26(14.94)	<0.001 [#]
	2	55(29.10)	35(20.11)	
	3	15(7.94)	22(12.64)	
	4	12(6.35)	6(3.45)	
	5	57(30.16)	85(48.85)	
	Unable to classify	6(3.08)	1(0.57)	

Notes: *Suspected extrathyroidal extensions include: adjacent to the peritoneum (suspected malignant nodule immediately adjacent to the thyroid peritoneum), invasion of the peritoneum (suspected disruption of the continuity of the thyroid peritoneum adjacent to the nodule), and disruption of the periphery (disruption of the structure of the thyroid tissue adjacent to the nodule). [#]P<0.05.

Abbreviations: KSThR, Korean Society of Thyroid Radiology; ETA, European Thyroid Association; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; PTC, papillary thyroid carcinomas.

Comparison of Diagnostic Efficacy of Different Diagnostic Modalities Used by Physicians of Varying Years of Experience

The efficacy of the different diagnostic modalities was higher for physicians of different years of experience (AUC >0.75), with ACR, ATA (exclusion of unclassifiable nodules), KSThR, and Chinese Thyroid Imaging Reporting and Data System (C-TIRADS) being better (AUC >0.85), and AI-SONICTM having higher diagnostic efficacy (AUC 0.84–0.85). Diagnostic efficacy was similar across years of experience, with higher years of experience having slightly higher diagnostic efficacy than lower years of experience.

For physicians of different years of experience from low to high, the highest diagnostic sensitivities were American Association of Clinical Endocrinologists/ American College of Endocrinology/Italian Association of Clinical Endocrinologists guidelines (AACE/ACE/AME) with 96.41%, 95.38%, and 96.41%, respectively, and the specificities were C-TIRADS 88.57%, KSThR 85.71%, KSThR, and AI-SONICTM Thyroid 85.71%, respectively, and positive predictive values were The highest were C-TIRADS 88.17%, KSThR 87.11%, KSThR 87.11%, respectively; the highest negative predictive values were all AACE/ACE/AME, 93.75%, 92.56%, 94.62%, respectively, and the highest accuracy were all ATA (excluding unclassifiable nodules), 87.26%, 87.93%, respectively, 88.82%, and the highest AUC were all C-TIRADS, 0.896, 0.894, and 0.900, respectively, as shown in Tables 3–5, and Figure 1A–C.

In this study, 24 and 27 thyroid nodules in the ultrasound images selected from low to high seniority physicians were automatically recognised as abnormal by the AI-SONICTM Thyroid system and were outlined by the sonographers themselves. Examples are shown in Figure 2A and B. Correct diagnosis examples using AI-SONICTM are shown in Figure 3A and B.

Consistency Comparison of Different Diagnostic Methods Used by Different-Yearly Trained Doctors

Consistency in the use of different diagnostic modalities was strong or better among physicians of different years of experience (Kendall concordance coefficient >0.6, P value <0.001), with the highest consistency in the use of AI-

Table 3 Efficacy Comparison of Different Diagnostic Methods for 2-Year Experienced Physicians

Diagnostic approach	Threshold	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC
ACR	TR4	88.72	78.29	81.99	86.16	83.78	0.855
ATA*	Moderate	95.38	77.30	83.76	93.16	87.26	0.869
AACE/ACE/AME	Moderate	96.41	60.00	72.87	93.75	79.19	0.786
KSThR	KTR4	83.59	85.14	86.24	82.32	84.32	0.869
ETA	EUTR4	95.38	59.43	72.37	92.04	78.38	0.775
C-TIRADS	C-TR4A	76.41	88.57	88.17	77.11	82.16	0.896
AI-SONICTM	Moderate	82.56	84.00	85.19	81.22	83.24	0.846

Notes: *Excludes 56 thyroid nodules that were unclassifiable by the ATA guidelines, of which 22 were PTC and 34 were benign nodules. [non-classifiable nodules: It is not possible to classify all thyroid nodules according to the ATA guidelines, eg solid or cystic solid nodules are “unclassifiable” when the solid component is isoechoic or hypoechoic with suspected malignant features, and the majority of studies have shown that the risk of malignancy for this type of nodule is approximately 16–18%].

Abbreviations: ACR, American College of Radiology; ATA, American Thyroid Association; AACE/ACE/AME, American Association of Clinical Endocrinologists/ American College of Endocrinology/Italian Association of Clinical Endocrinologists; KSThR, Korean Society of Thyroid Radiology; ETA, European Thyroid Association; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; AI-SONICTM, AI-SONICTM Thyroid System; AUC, Area under curve.

Table 4 Efficacy Comparison of Different Diagnostic Methods for 5-Year Experienced Physicians

Risk Stratification System	Threshold	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC
ACR	TR4	87.18	81.71	84.16	85.12	84.59	0.867
ATA*	Moderate	94.41	79.86	85.35	92.00	87.93	0.880
AACE/ACE/AME	Moderate	95.38	64.00	74.70	92.56	80.54	0.799
KSThR	KTR4	86.67	85.71	87.11	85.23	86.22	0.885
ETA	EUTR4	94.87	64.57	74.90	91.87	80.54	0.797
C-TIRADS	C-TR4A	93.85	69.71	77.54	91.04	82.43	0.894
AI-SONICTM	Moderate	83.08	85.14	86.17	81.87	84.05	0.850

Notes: *Excludes 47 thyroid nodules that were unclassifiable by the ATA guidelines, of which 16 were PTC and 31 were benign nodules. [non-classifiable nodules: It is not possible to classify all thyroid nodules according to the ATA guidelines, eg solid or cystic solid nodules are “unclassifiable” when the solid component is isoechoic or hypoechoic with suspected malignant features, and the majority of studies have shown that the risk of malignancy for this type of nodule is approximately 16–18%].

Abbreviations: ACR, American College of Radiology; ATA, American Thyroid Association; AACE/ACE/AME, American Association of Clinical Endocrinologists/ American College of Endocrinology/Italian Association of Clinical Endocrinologists; KSThR, Korean Society of Thyroid Radiology; ETA, European Thyroid Association; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; AI-SONICTM, AI-SONICTM Thyroid System; AUC, Area under curve.

Table 5 Efficacy Comparison of Different Diagnostic Methods for 10-Year Experienced Physicians

Risk Stratification System	Threshold	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC
ACR	TR4	88.72	80.57	83.57	86.50	84.86	0.868
ATA*	Moderate	95.51	81.05	85.43	93.94	88.82	0.889
AACE/ACE/AME	Moderate	96.41	70.29	78.33	94.62	84.05	0.835
KSThR	KTR4	86.67	85.71	87.11	85.23	86.22	0.883
ETA	EUTR4	95.90	70.86	78.57	93.94	84.05	0.836
C-TIRADS	C-TR4A	94.36	73.71	80.00	92.14	84.59	0.900
AI-SONICTM	Moderate	84.62	85.71	86.84	83.33	85.14	0.865

Notes: *Excludes 39 thyroid nodules that were unclassifiable by the ATA guidelines, of which 17 were PTC and 22 were benign nodules. [non-classifiable nodules: It is not possible to classify all thyroid nodules according to the ATA guidelines, eg solid or cystic solid nodules are “unclassifiable” when the solid component is isoechoic or hypoechoic with suspected malignant features, and the majority of studies have shown that the risk of malignancy for this type of nodule is approximately 16% - 18%].

Abbreviations: ACR, American College of Radiology; ATA, American Thyroid Association; AACE/ACE/AME, American Association of Clinical Endocrinologists/ American College of Endocrinology/Italian Association of Clinical Endocrinologists; KSThR, Korean Society of Thyroid Radiology; ETA, European Thyroid Association; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; AI-SONICTM, AI-SONICTM Thyroid System; AUC, Area under curve.

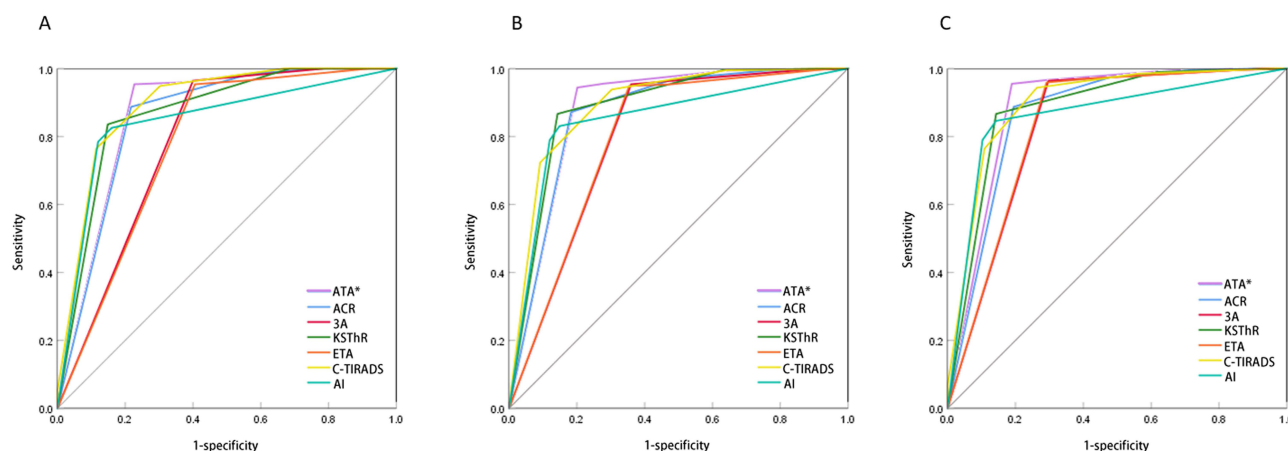


Figure 1 (A) ROC curve comparing the diagnostic performance of AI-SONICTM Thyroid system versus six guidelines in thyroid nodule diagnosis by a two-year experienced physician; (B) ROC curve comparing the diagnostic performance of AI-SONICTM Thyroid system versus six guidelines in thyroid nodule diagnosis by a five-year experienced physician; (C) ROC curve comparing the diagnostic performance of AI-SONICTM Thyroid system versus six guidelines in thyroid nodule diagnosis by a ten-year experienced physician.

Abbreviations: ACR, American College of Radiology; ATA, American Thyroid Association; 3A: AACE/ACE/AME, American Association of Clinical Endocrinologists/American College of Endocrinology/Italian Association of Clinical Endocrinologists; KSThR, Korean Society of Thyroid Radiology; ETA, European Thyroid Association; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; AI, AI-SONICTM, AI-SONICTM Thyroid System.

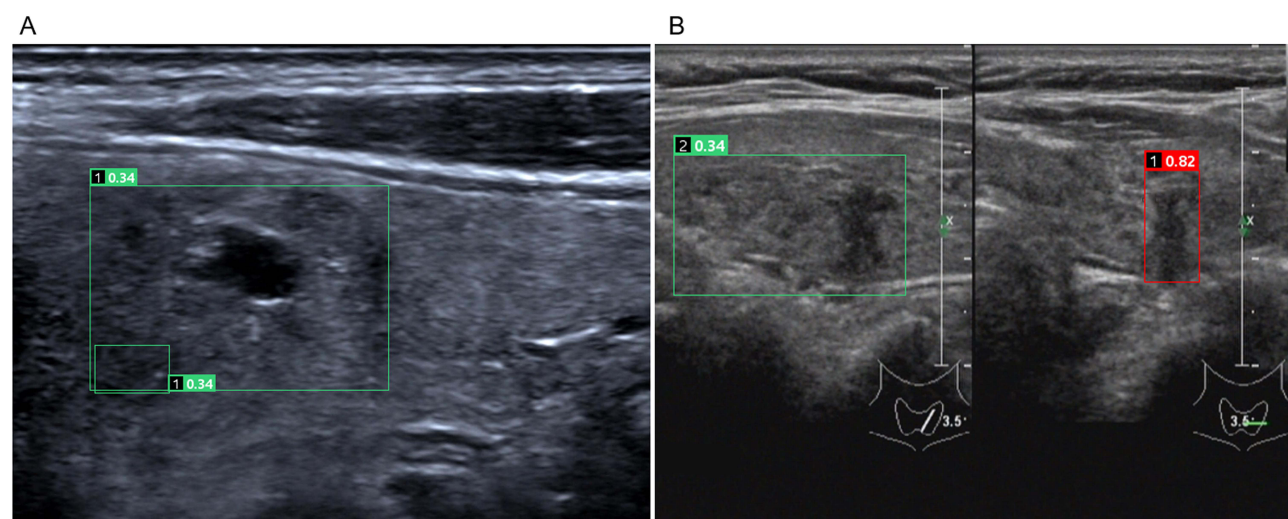


Figure 2 (A) 57-year-old male patient with a left thyroid nodule. Postoperative pathological examination revealed a nodular goiter. The AI-SONICTM Thyroid system automatically identified an abnormality by “non-existent nodule”(box selected). The physician redrew the nodule contour, but the score remained unchanged; (B) 58-year-old female patient with a left thyroid nodule. Postoperative pathological examination revealed papillary thyroid cancer with Hashimoto's thyroiditis. The AI-SONICTM Thyroid system automatically identified an abnormality by including the surrounding thickened and uneven thyroid parenchyma in the box selection. The physician redrew the nodule contour, resulting in a change in AI score from benign to malignant. The box marks on the images and the scores above them are automatically generated by the AI-SONICTM Thyroid system, which frames the nodules along their edges, with a green border for a score of 0–0.4 inclined to benign, an Orange border for a score of 0.41–0.60 moderately suspicious (no corresponding image provided), and a red border for a score of 0.65–0.99 inclined to malignant.

SONICTM Thyroid (Kendall concordance coefficient 0.995, P value <0.001), followed by C-TIRADS and ACR (Kendall concordance coefficients of 0.952 and 0.951, respectively, p -value < 0.001), and ATA was slightly inferior to the other guidelines (Kendall's coordination coefficient of 0.787, p -value < 0.001), see Table 6.

Impact of Different Characteristics on Diagnostic Accuracy of AI-SONICTM

The number of nodules correctly determined to be benign or malignant by AI-SONICTM Thyroid was 315 (165 PTC and 150 benign nodules). The regression coefficient value of 0.983 ($P=0.002$) for nodules with cystic components in their structure had a significant positive effect on the correctness of diagnosis by AI-SONICTM Thyroid, with a marginal effect

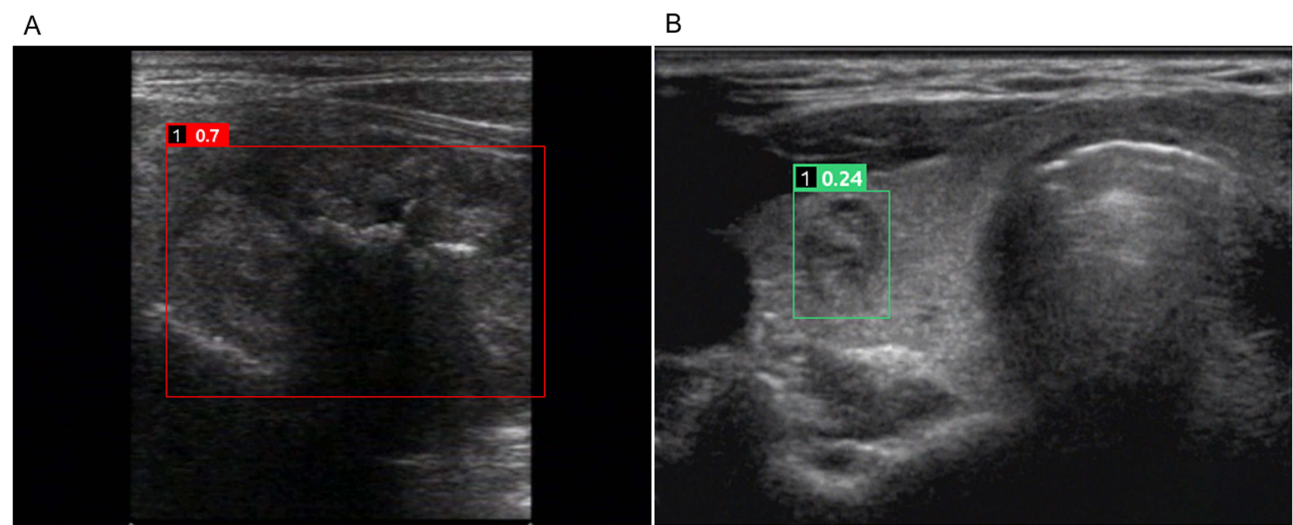


Figure 3 (A) 44-year-old female patient with a right thyroid nodule. Postoperative pathological examination revealed papillary thyroid cancer. Due to the lack of typical ultrasound features of malignancy, all six guidelines had low to intermediate suspicion, while the AI-SONICTM Thyroid system classified it as a malignant nodule; (B) 62-year-old female patient with a right thyroid nodule. Postoperative pathological examination revealed nodular goiter. Due to the presence of ultrasound features suggestive of malignancy such as an aspect ratio greater than 1, all six guidelines had a high suspicion of malignancy, while the AI-SONICTM Thyroid system classified it as a benign nodule. The box marks on the images and the scores above them are automatically generated by the AI-SONICTM Thyroid system, which frames the nodules along their edges, with a green border for a score of 0–0.4 inclined to benign, an Orange border for a score of 0.41–0.60 moderately suspicious (no corresponding image provided), and a red border for a score of 0.65–0.99 inclined to malignant.

of 0.154, ie, the proportion of nodules correctly diagnosed by AI-SONICTM Thyroid for nodules with cystic components was 15.4% greater than that for solid nodules. The various baseline and ultrasound characteristics of the remaining thyroid nodules did not have a significant effect on the correctness of the AI-SONICTM Thyroid diagnosis ($p > 0.05$, Table 7).

Needle Biopsies Avoided by Applying AI-SONICTM

By the data analyzed, unnecessary needle biopsies can be avoided by applying AI-SONICTM, as shown in Table 8. In this study, the application of AI-SONICTM avoided 40.63–55.56% of the unnecessary needle biopsies that raised by

Table 6 Consistency Comparison of Different Diagnostic Methods Used by Different-Yearly Trained Doctors

Diagnostic Method	Kendall Coordination Coefficient	P-value
ACR	0.951	<0.001*
ATA	0.787	<0.001*
AACE/ACE/AME	0.891	<0.001*
KSThR	0.949	<0.001*
ETA	0.885	<0.001*
C-TIRADS	0.952	<0.001*
AI-SONICTM	0.995	<0.001*

Note: * $P < 0.05$.

Abbreviations: ACR, American College of Radiology; ATA, American Thyroid Association; AACE/ACE/AME, American Association of Clinical Endocrinologists/ American College of Endocrinology/Italian Association of Clinical Endocrinologists; KSThR, Korean Society of Thyroid Radiology; ETA, European Thyroid Association; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; AI-SONICTM, AI-SONICTM Thyroid System.

Table 7 Impact of Different Characteristics on Diagnostic Accuracy of AI-SONICTM

Feature	Regression Coefficient	P-value	Marginal Effect
Pathology: PTC	0.066	0.773	0.014
Isthmus	0.259	0.528	0.05
Upper	0.201	0.362	0.042
Middle	0.127	0.575	0.028
Lower	0.421	0.073	0.091
Volume (cm^3)	−0.011	0.305	−0.002
Presence of thyroiditis	0.27	0.303	0.053
Cystic Component Present	0.983	0.002*	0.154
Homogeneous texture	−0.318	0.182	−0.063
High/isoechoic	0.087	0.722	0.019
Indeterminate borders	−0.173	0.484	−0.04
Irregular/microlobulated borders	0.268	0.326	0.059
Width-to-length ratio >1	0.194	0.378	0.041
Microcalcifications	−0.103	0.583	−0.023
Coarse calcifications	−0.023	0.910	−0.005
Suspected extrathyroidal extension	0.153	0.451	0.033

Note: * P<0.05.
Abbreviation: PTC, papillary thyroid carcinomas.

Table 8 The Unnecessary Fine-Needle Aspiration Biopsies Avoided by Applying AI-SONICTM(n,%)

Guidelines	Nodules Suggested of Needle Biopsies by Guidelines	Malignant Nodules by Needle Biopsies	Nodules Avoided Unnecessary Needle Biopsies by AI-SONICTM	AI-SONICTM Missed Nodules that Should Needle Biopsies
ACR	171 (46.22)	80 (46.78)	83 (48.54)	15 (8.77)
ATA*	189 (57.10)	72 (38.10)	105 (55.56)	11 (5.82)
AACE/ACE/AME	201 (54.32)	87 (43.28)	103 (51.24)	20 (9.95)
KSThR	245 (66.22)	126 (51.43)	102 (41.63)	20 (8.16)
ETA	218 (58.92)	89 (40.83)	118 (54.13)	21 (9.63)
C-TIRADS	160 (43.24)	85 (53.13)	65 (40.63)	21 (13.13)

Notes:* Excludes 39 thyroid nodules that were unclassifiable by the ATA guidelines, of which 17 were PTC and 22 were benign nodules. [non-classifiable nodules: It is not possible to classify all thyroid nodules according to the ATA guidelines, eg solid or cystic solid nodules are “unclassifiable” when the solid component is isoechoic or hypoechoic with suspected malignant features, and the majority of studies have shown that the risk of malignancy for this type of nodule is approximately 16% - 18%].
Abbreviations: ACR, American College of Radiology; ATA, American Thyroid Association; AACE/ACE/AME, American Association of Clinical Endocrinologists/ American College of Endocrinology/Italian Association of Clinical Endocrinologists; KSThR, Korean Society of Thyroid Radiology; ETA, European Thyroid Association; C-TIRADS, Chinese Thyroid Imaging Reporting and Data System; AI-SONICTM, AI-SONICTM Thyroid System; AUC, Area under curve.

other guidelines. At the same time, AI-SONICTM missed some nodules that should do needle biopsies, the rate was 5.22%-13.13%.

Discussion

The diagnostic efficacy of different models used by doctors with different experience was generally higher (all AUC > 0.75), and there was no significant difference in the diagnostic efficacy. The diagnostic efficacy of senior doctors was slightly higher than that of junior doctors. The sensitivity of AACE/ACE/AME diagnosis is most advantageous for experienced physicians. ACR, ATA (excluding unclassifiable nodules), KSTHR, and C-TIRADS performed better, and the diagnostic value of AI-SONICTM for the thyroid was comparable to that of KSTHR and C-TIRADS. Previous

studies have shown that the six guidelines and the AI-SONIC™ system are of satisfactory value in distinguishing between benign and malignant Thyroid nodule.¹³ Yang et al finds the ATA and ACR guidelines and AI-SONIC™ system can efficiently differentiate malignant from benign thyroid nodules and C-TIRADS exhibited the best performance.¹¹ This study found that physicians of different seniority demonstrated high diagnostic value and good concordance when using these diagnostic methods, and that the AI-SONICTM could serve as an effective adjunct diagnostic tool, to improve the accuracy and efficiency of diagnostic papillary thyroid cancer.

Guo et al¹⁴ conducted a study that included 1092 ultrasound images and compared the diagnostic time and value between AI-SONICTM and human physicians using ACR guidelines as the diagnostic tool. The results showed that AI-SONICTM had an average diagnostic time that was significantly shorter than that of human physicians (0.146s vs 2.8–4.5 minutes, respectively).¹⁴ Moreover, AI-SONICTM demonstrated an overall diagnostic value comparable to that of senior physicians, with values of 91.7% and 91.2%, respectively. Although AI-SONICTM had slightly lower sensitivity (91.5% vs 96.7%, respectively), its specificity was significantly higher than that of human physicians (92.0% vs 79.2%, respectively). Additionally, AI-SONICTM exhibited high diagnostic value even in studies on rare malignant tumors of thyroid cancer, such as follicular carcinoma and medullary carcinoma.¹⁵

Moreover, other thyroid artificial intelligence-assisted diagnosis systems, whether still in the development and testing phase or widely used in the market (such as Afirma, ThyroSeq, Rosetta GX reveal, and Thyramid),^{16,17} most studies have also shown that they have high diagnostic value. In another review, it was also suggested that in almost all studies involving thyroid artificial intelligence-assisted diagnosis, artificial intelligence achieved better results than radiologists with less than 5 years of experience. However, a recent meta-analysis on the application of artificial intelligence in thyroid diseases stated that current challenges of available artificial intelligence applications include lack of prospective and multi-center validation and utility studies,¹⁸ small and low-diversity training datasets, data source differences, lack of interpretability, unclear clinical impact, insufficient stakeholder participation, and difficulties in using different research environments for external use. These factors may limit its future adoption.

The pathological types of thyroid malignant nodules in this study only included the most common papillary thyroid carcinoma (PTC). On one hand, the aim was to utilize artificial intelligence to address the most prevalent issues in thyroid ultrasound screening and improve work efficiency. On the other hand, this was due to the fact that the largest proportion of pathological types in the dataset used to train the AI was also PTC, and data volume is closely related to the effectiveness of AI-assisted diagnosis systems. The application of these systems in other rare cancers will require additional dataset support in the future. Similarly, for guidelines involving thyroid nodule risk stratification, they are effective tools for diagnosing PTC.¹⁷ The combination of ultrasound and thyroid risk stratification systems yields a much higher accuracy rate compared to using ultrasound alone. However, these methods may not perform as well in less common pathological types.^{18,19}

In this study, AACE/ACE/AME had the highest sensitivity and negative predictive value, while C-TIRADS and KSThR had higher specificity. In the study by Yang et al, papillary carcinoma accounted for 71% of all malignant tumors.¹¹ The Korean Thyroid Association and Korean Society of Thyroid Radiology (KTA/KSThR), National Comprehensive Cancer Network (NCCN), and American Thyroid Association (ATA) guidelines had higher sensitivity than other guidelines. In the study by EJ Ha et al, papillary carcinoma accounted for 85.5% of all malignant tumors. The Korean Thyroid Association and Korean Society of Thyroid Radiology (KTA/KSThR, 94.5%), National Comprehensive Cancer Network (NCCN, 92.6%), and American Thyroid Association (ATA, 89.6%) guidelines had higher sensitivity than other guidelines.²⁰ In another meta-analysis,²¹ involving 49,661 patients, the most accurate risk category threshold was AACE/ACE/AME system's grade 3 (high risk), ACR TI-RADS's TR5 (highly suspicious), EU-TIRADS's EU-TIRDS 5 (high risk), Kwak TIRADS's 4c (moderate concern but not a typical malignancy), K-TIRADS grade 5 (highly suspicious), and ATA system's highly suspicious. At these thresholds, the system's sensitivity was 64–77% and specificity was 82–90%, with ACR TI-RADS having the highest sensitivity and specificity, followed by K-TIRADS. These differences may be attributed to the higher proportion of malignant nodules included in this study, as all nodules were surgically removed and had definitive postoperative pathological results.

The ATA system, consistent with the study by Sparano et al,²² exhibited the highest accuracy in our study. However, it excluded 39 to 56 nodules that had a high risk of malignancy based on previous studies, which could pose challenges

for less experienced physicians.²³ The overall diagnostic performance of different-year physicians in our study was similar, with higher-year physicians exhibiting only slightly higher diagnostic efficiency than lower-year physicians. This may be attributed to the fact that prior to image interpretation, all physicians were provided with a uniform guideline-based reading and nodule classification training, leading them to rely more on the guidelines and risk stratification systems rather than their individual experiences.

Physicians of different seniorities adopt various approaches in diagnosis, yet they maintain a remarkable consistency. Such consistency may stem from classification disagreements caused by non-classifiable nodules. A study involving seven radiologists and endocrinologists from three major thyroid centers jointly conducted a consistency analysis on 100 ultrasound images.²⁴ The study indicates that among single-center observations, there is moderate consistency among observers; however, in multi-center studies, consistency performance is poorer. Specifically, the ATA consistency is the lowest at 0.3, while the consistencies for AACE, ACR-TIRADS, and EU-TIRADS are 0.44, 0.42, and 0.39, respectively. In a study by Koh et al²⁵ four physicians of varying seniorities were involved. The findings reveal that when utilizing TIRADS, the experienced physician group demonstrates higher consistency compared to the less experienced group; however, when applying the ATA guidelines, the less experienced group exhibits higher consistency. Overall, this study is a single-center investigation with only three participants who collectively explored various guidelines, potentially explaining the higher consistency observed in this study compared to others. This phenomenon suggests that through unified training and practice, consistency among observers can potentially be enhanced, thereby improving the repeatability of risk stratification system applications.

The consistency exhibited by AI-SONICTM is nearly perfect, primarily attributed to two major factors. Firstly, compared to human physicians, AI-SONICTM captures nodule features at the pixel level, achieving a level of refinement beyond human capabilities and reducing inter-observer variations due to its objectivity. Secondly, the limited availability of ultrasound images selected in retrospective studies, possibly due to the fact that physicians did not collect the images themselves, contributes to minimal variability. In our study, we aimed to simulate real clinical scenarios by allowing physicians to independently select ultrasound images. In this process, different sections of the same nodule can lead to varying diagnoses, an issue that dynamic AI may effectively address. Hongxia Luo et al²⁶ developed and validated a deep learning method based on ultrasound videos (utilizing a Cascade Region-based Convolutional Network, R-CNN) for the automatic detection and segmentation of the thyroid and its surrounding tissue. Separately, Wang et al²⁷ employed Dynamic AI for comprehensive scanning of the thyroid, achieving real-time positioning, real-time tracing, and real-time diagnosis during examinations, with an accuracy rate of 92.21% for malignant nodules and 83.20% for benign nodules.

Attempts to explain diagnostic accuracy through physician-observed baselines or ultrasound features have been largely unsuccessful. This study has found that AI-SONICTM demonstrates a 15.4% increase in diagnostic accuracy for nodules containing cystic components compared to solid nodules. However, no significant correlation was observed between other ultrasound features and diagnostic accuracy. It remains unclear whether this is related to the machine learning from a larger sample of cystic and solid nodules. Notably, features such as pathological type, nodule location, nodule volume, combined diffuse lesions, texture, echo, boundary, margin, aspect ratio, microcalcification, macrocalcification, and thyroid extension have minimal impact on the diagnostic accuracy of AI-SONICTM. In contrast to the findings of the current study, Liu et al²⁸ research revealed that the agreement between AI-SONICTM diagnosis and pathological diagnosis is moderate in a diffuse background ($\kappa=0.417$) and nearly perfect in a non-diffuse background ($\kappa=0.81$). Additionally, in non-diffuse backgrounds, AI-SONICTM demonstrates significantly higher sensitivity (96.2% vs 73.4%, $P < 0.001$), specificity (82.9% vs 71.2%, $P = 0.007$), and negative predictive value (90.3% vs 53.3%, $P < 0.001$). Whether this situation is related to the type of diffuse lesion or the degree of destruction it causes to the thyroid parenchyma requires further investigation to determine.

Regretfully, this study only observed that the AI showed abnormal nodules' edge recognition when these sonograms appeared, perhaps because its training set was less likely to include these nodules. But we have no way of knowing exactly why and how the technological black-box-like AI detection system abnormal nodules. Abnormal nodules were rare and not enough to be analysed in our study. And the physician's re-sketched results were mostly in agreement with the original results identified by AI in terms of their benign-malignant tendency, with very rare instances of inconsistent

benign and malignant tendencies. Therefore, this study did not analyze the specially abnormal nodules that were manually outlined by the physicians, which should be studied in further study with large scale samples.

The limitations of this study are mainly reflected in the following two aspects: (1) This study employed a retrospective research design, where the images selected by each physician were obtained from those collected by different physicians during their routine practice. Consequently, it was not possible to adequately simulate the impact of variations in image acquisition among physicians of different experience levels on AI usage. This limitation may impose some constraints on the interpretation and extrapolation of the results. (2) The pathological types of thyroid malignant nodules included in this study were limited to PTC (papillary thyroid carcinoma), which does not comprehensively represent the diverse pathological types that may be encountered in practical applications. However, our research team is actively engaged in studying rare cancers and is committed to developing new artificial intelligence systems. To expand the scope of AI application and enhance diagnostic accuracy, we will continue to collect more data and samples of different types of malignant nodules.

Conclusion

AI-SONICTM and the six thyroid nodule ultrasound risk stratification systems showed high diagnostic performance for papillary thyroid carcinoma, and showed strong or moderate interobserver agreement with examiners with different experiences. AI-SONICTM may have higher accuracy for nodules with cystic components.

Data Sharing Statement

The data related with this study was achievable from the corresponding author.

Ethics Approval and Consent to Participate

This study was proceeded in compliance with the Declaration of Helsinki, and was approved by the Ethics Committee of Fujian Provincial Hospital (No. K2019-01-051). Each participant has given written informed consent for this study.

Disclosure

The authors declare that they have no competing interests.

References

1. Mulita F, Anjum F. Thyroid Adenoma. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2024. PMID: 32965923.
2. Houten PV, Netea-Maier RT, Smit JW. Differentiated thyroid carcinoma: an update. *Best Pract Res Clin Endocrinol Metab*. 2023;37(1):101687. doi:10.1016/j.beem.2022.101687
3. Galuppini F, Censi S, Merante Boschini I, et al. Papillary thyroid carcinoma: molecular distinction by MicroRNA profiling. *Front Endocrinol*. 2022;13:834075. doi:10.3389/fendo.2022.834075
4. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2016;26(1):1–133. doi:10.1089/thy.2015.0020
5. Gharib H, Papini E, Garber JR, et al. American Association of clinical endocrinologists, American college of endocrinology, and associazione Medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules–2016 update. *Endocrine Pract*. 2016;22(5):622–639.
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 Radio*. 2017;14(5):587–595. doi:10.1016/j.jacr.2017.01.046
7. 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. doi:10.1159/000478927
8. Zhou J, Yin L, Wei X, et al. 2020 Chinese guidelines for ultrasound malignancy risk stratification of thyroid nodules: the C-TIRADS. *Endocrine*. 2020;70(2):256–279. doi:10.1007/s12020-020-02441-y
9. Lee JY, Baek JH, Ha EJ, et al. 2020 imaging guidelines for thyroid nodules and differentiated thyroid cancer: Korean society of thyroid radiology. *Korean J Radiol*. 2021;22(5):840–860. doi:10.3348/kjr.2020.0578
10. Raza MA, Aziz S. Transformative potential of artificial intelligence in pharmacy practice. *Saudi Pharm J*. 2023;31(9):101706. doi:10.1016/j.jsps.2023.101706
11. Yang L, Lin N, Wang M, Chen G. Diagnostic efficiency of existing guidelines and the AI-SONICTM artificial intelligence for ultrasound-based risk assessment of thyroid nodules. *Front Endocrinol*. 2023;14:1116550. doi:10.3389/fendo.2023.1116550
12. Guo FQ, Zhao JQ, Liu S. Application of artificial intelligence automatic detection system in preoperative ultrasonic diagnosis of thyroid nodules. *Acad J Second Military Med Univ*. 2019;40(11):1183–1189.

13. Peng S, Liu Y, Lv W, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health*. 2021;3(4):e250–e259. doi:10.1016/S2589-7500(21)00041-8
14. Guo F, Chang W, Zhao J, Xu L, Zheng X, Guo J. Assessment of the statistical optimization strategies and clinical evaluation of an artificial intelligence-based automated diagnostic system for thyroid nodule screening. *Quant Imaging Med Surg*. 2023;13(2):695–706. doi:10.21037/qims-22-85
15. Wang Y, Xu L, Lu W, et al. Clinical evaluation of malignancy diagnosis of rare thyroid carcinomas by an artificial intelligent automatic diagnosis system. *Endocrine*. 2023;80(1):93–99.
16. Toro-Tobon D, Llor-Torres R, Duran M, et al. Artificial intelligence in thyroidology: a narrative review of the current applications, associated challenges, and future directions. *Thyroid*. 2023;33(8):903–917. doi:10.1089/thy.2023.0132
17. Trimboli P, Durante C. Ultrasound risk stratification systems for thyroid nodule: between lights and shadows, we are moving towards a new era. *Endocrine*. 2020;69(1):1–4. doi:10.1007/s12020-020-02196-6
18. Ferrarazzo G, Camponovo C, Deandrea M, Piccardo A, Scappaticcio L, Trimboli P. Suboptimal accuracy of ultrasound and ultrasound-based risk stratification systems in detecting medullary thyroid carcinoma should not be overlooked. Findings from a systematic review with meta-analysis. *Clin Endocrinol*. 2022;97(5):532–540. doi:10.1111/cen.14739
19. Yang J, Sun Y, Li X, et al. Diagnostic performance of six ultrasound-based risk stratification systems in thyroid follicular neoplasm: a retrospective multi-center study. *Front Oncol*. 2022;12:1013410. doi:10.3389/fonc.2022.1013410
20. Ha EJ, Baek JH, Kim KW, et al. Comparative efficacy of radiofrequency and laser ablation for the treatment of benign thyroid nodules: systematic review including traditional pooling and Bayesian network meta-analysis. *J Clin Endocrinol Metab*. 2015;100(5):1903–1911. doi:10.1210/jc.2014-4077
21. Kim DH, Kim SW, Basurrah MA, Lee J, Hwang SH. Diagnostic performance of six ultrasound risk stratification systems for thyroid nodules: a systematic review and network meta-analysis. *AJR Am J Roentgenol*. 2023;220(6):791–803. doi:10.2214/AJR.22.28556
22. Sparano C, Verdiani V, Pupilli C, et al. Choosing the best algorithm among five thyroid nodule ultrasound scores: from performance to cytology sparing—a single-center retrospective study in a large cohort. *Eur Radiol*. 2021;31:5689–5698. doi:10.1007/s00330-021-07703-5
23. Peng JY, Pan FS, Wang W, et al. Malignancy risk stratification and FNA recommendations for thyroid nodules: a comparison of ACR TI-RADS, AACE/ACE/AME and ATA guidelines. *Am J Otolaryngol*. 2020;41(6):102625. doi:10.1016/j.amjoto.2020.102625
24. 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(2):237–242. doi:10.1089/thy.2019.0360
25. Koh J, Kim SY, Lee HS, et al. Diagnostic performances and interobserver agreement according to observer experience: a comparison study using three guidelines for management of thyroid nodules. *Acta Radiol*. 2018;59(8):917–923. doi:10.1177/0284185117744001
26. Luo H, Ma L, Wu X, et al. Deep learning-based ultrasonic dynamic video detection and segmentation of thyroid gland and its surrounding cervical soft tissues. *Med Phys*. 2022;49(1):382–392. doi:10.1002/mp.15332
27. Wang B, Wan Z, Li C, et al. Identification of benign and malignant thyroid nodules based on dynamic AI ultrasound intelligent auxiliary diagnosis system. *Front Endocrinol*. 2022;13:1018321. doi:10.3389/fendo.2022.1018321
28. Liu T, Wu C, Wang G, Jia Y, Zhu Y, Nie F. Clinical value of artificial intelligence-based computer-aided diagnosis system versus contrast-enhanced ultrasound for differentiation of benign from malignant thyroid nodules in different backgrounds. *J Ultrasound Med*. 2023;42(8):1757–1766. doi:10.1002/jum.16195

Therapeutics and Clinical Risk Management

Dovepress

Publish your work in this journal

Therapeutics and Clinical Risk Management is an international, peer-reviewed journal of clinical therapeutics and risk management, focusing on concise rapid reporting of clinical studies in all therapeutic areas, outcomes, safety, and programs for the effective, safe, and sustained use of medicines. This journal is indexed on PubMed Central, CAS, EMBASE, Scopus and the Elsevier Bibliographic databases. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/therapeutics-and-clinical-risk-management-journal>