Yang *et al. BMC Medical Imaging* (2025) 25:297
https://doi.org/10.1186/s12880-025-01845-4

BMC Medical Imaging

RESEARCH Open Access

O-RADS US versus IOTA simple rules in the diagnosis of benign and malignant adnexal masses: a prospective study

Ya Yang[1], Hongyan Wang[1*], Na Su[1], Luying Gao[1], Yang Gu[1], Siman Cai[1], Qing Dai[1], Jianchu Li[1] and Yuxin Jiang[1]

Abstract

Background Although many studies have validated the diagnostic performance of Ovarian-Adnexal Reporting and Data Systems ultrasound (O-RADS US), most have been observed by experienced sonologists, and relatively few by junior sonologists. The purpose of this study was to compare the diagnostic performance of the O-RADS US and the International Ovarian Tumor Analysis (IOTA) Simple Rules (SRs) in senior and junior sonologists to determine a more suitable assessment model for general clinical use.

Methods We prospectively recruited 228 patients diagnosed with adnexal masses (AMs). Two senior sonologists acquired images and evaluated them following the O-RADS US and IOTA guidelines, and two junior sonologists reviewed and analyzed images and evaluated them following the same guidelines. In this research, pathological findings were used as the reference standard. Comparisons of categorical variables were made using the chi-square test, and comparisons of continuous variables were made using the two independent-samples t-test. The diagnostic performance of the models was compared by analyzing the receiver operating characteristic (ROC) curve. The kappa value (κ) was used to compare the interobserver agreement between the senior and junior sonologists and the agreement between each ultrasound method and the reference standard.

Results Of 228 AMs, 176 were benign and 52 malignant. The junior adjusted O-RADS US (> O-RADS 4a represents malignancy) had the highest diagnostic validity, with a sensitivity, specificity, and accuracy of 94.23%, 87.5%, and 89.04%, respectively, and ROC curve of 0.959 (95% CI, 0.924–0.980). Both junior unadjusted (> O-RADS 3 represents malignancy) and adjusted O-RADS US had significantly higher diagnostic performance than the junior SRs (AUC 0.951 and 0.959 vs. 0.840, $P = 0.0003$, 0.0001, respectively). Interobserver agreement between senior and junior sonologists using O-RADS US was moderate (κ = 0.465), and interobserver agreement between senior and junior sonologists using SRs, unadjusted, and adjusted O-RADS US was good (κ = 0.618, 0.657, and 0.718, respectively). The junior unadjusted O-RADS US, adjusted O-RADS US, and SRs showed good agreement with the pathological results (κ = 0.648, 0.724, 0.716, respectively).

Ya Yang: First author

*Correspondence:
Hongyan Wang
whychina@126.com

Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Conclusions When assisting sonologists in AM diagnosis, the O-RADS US, especially the adjusted O-RADS US, had higher diagnostic performance than the SRs, and it would be more suitable for general clinical application.

Keywords Ultrasound, Adnexal masses, Ovarian cancer, Ovarian-adnexal reporting and data systems ultrasound (O-RADS US), Simple rules

Introduction

Ovarian cancer remains the leading cause of death from gynecologic malignancies, and as a safe, noninvasive, and affordable method, transvaginal ultrasonography (TVS) remains one of the main screening modalities for ovarian cancer [1–3]. With the advancement of ultrasound (US) technology, its application in female pelvic masses is becoming increasingly widespread, especially in recent years [4]. Studies have shown that, for borderline tumors, US is more sensitive (91%) than CA125 (55%) [5]. And in postmenopausal patients with elevated CA125 levels, US can effectively distinguish between patients with an increased cancer risk index and those with a non-increased risk index [6, 7].

According to statistics, approximately 63% of ovarian cancer patients are already in stage IV at the time of diagnosis [8, 9]. Compared with stage I patients with a higher 5-year survival rate (92.1%), the 5-year survival rate of patients in this stage is only 17% [2, 10]. Therefore, early identification of ovarian cancer has become both a daunting but rewarding task. However, the accuracy of US diagnosis relies heavily on the experience of the sonologist, and identifying early-stage ovarian cancer characterized by a lack of specific clinical symptoms is a great challenge for inexperienced junior sonologists. Studies showed that US can be an effective tool for the early detection of recurrent ovarian cancer if the examination is performed by an experienced sonologist [11, 12]. Therefore, there is an urgent need to improve the ability of junior sonologists to diagnose ovarian tumors, and this may be an effective measure to improve the overall survival rate of ovarian cancer patients.

To improve the consistency and accuracy of US reporting, several structured reports and guidelines have been established for the evaluation of ovarian-adnexal masses [13]. One such model is the Simple Rules (SRs) proposed by the International Ovarian Tumour Analysis (IOTA) group, which is now generally accepted in clinical practice. These rules, proposed in 2008, include five B features for benign tumors and five M features for malignant tumors, and studies have shown that, when combined with sonologists' subjective assessments, they still have high sensitivity and specificity [14].

To optimize the prognosis of ovarian cancer while reducing unnecessary surgery in patients with low-grade malignancy risk tumors, the American College of Radiology (ACR) officially released consensus guidelines for the US risk stratification and management system of Ovarian-Adnexal Reporting and Data Systems ultrasound (O-RADS US) in 2020 [15]. The guideline classifies ovarian-adnexal masses into 6 categories, which include normal to highly malignant risk categories, and the guidelines define each category in detail so that sonologists have rules to follow in the process of diagnosis.

At present, many studies [13, 16] have validated the diagnostic performance of the O-RADS US and/or compared it with the US diagnostic classification systems for adnexal masses (AMs), such as the IOTA SRs and GI-RADS, but most of the studies have been observed by experienced sonologists, and relatively few have been performed by junior sonologists. The aim of this study was to compare the diagnostic performance of the IOTA SRs and the O-RADS US model to determine a more suitable assessment model for general clinical use.

Methods

This prospective study was approved by the ethics review committee of the Peking Union Medical College Hospital (PUMCH). All patients were informed of the procedure and provided written informed consent prior to the examination.

Study population

This prospective study was conducted on 239 patients diagnosed with suspected AMs between June 2021 and August 2022 at PUMCH. These AMs were first detected by clinical palpation and later confirmed by ultrasound or MRI. Patients underwent surgery if their AMs met the criteria for surgical treatment or if they had a strong desire for surgery due to dysmenorrhoea or other reasons. All patients were enrolled consecutively, and all US examinations were completed preoperatively. The inclusion criteria for the study were patients hospitalized for surgery for primary adnexal masses. The exclusion criteria were as follows: (1) undetermined specific pathological type of the lesion ($n=5$); (2) poor image quality (For example, inappropriate scale adjustment, blurred images, etc.) ($n=6$). If a patient had multiple lesions at the same time, we included only the lesion with the highest O-RADS US category, or the largest if the O-RADS US categories were the same. Finally, we included 228 lesions from 228 patients. The flow chart of the study was shown in Fig. 1. Before starting the examination, the patient's age, body mass index (BMI), age at menarche, and clinical symptoms (abdominal distension, abdominal pain, abdominal mass, vaginal bleeding or drainage,

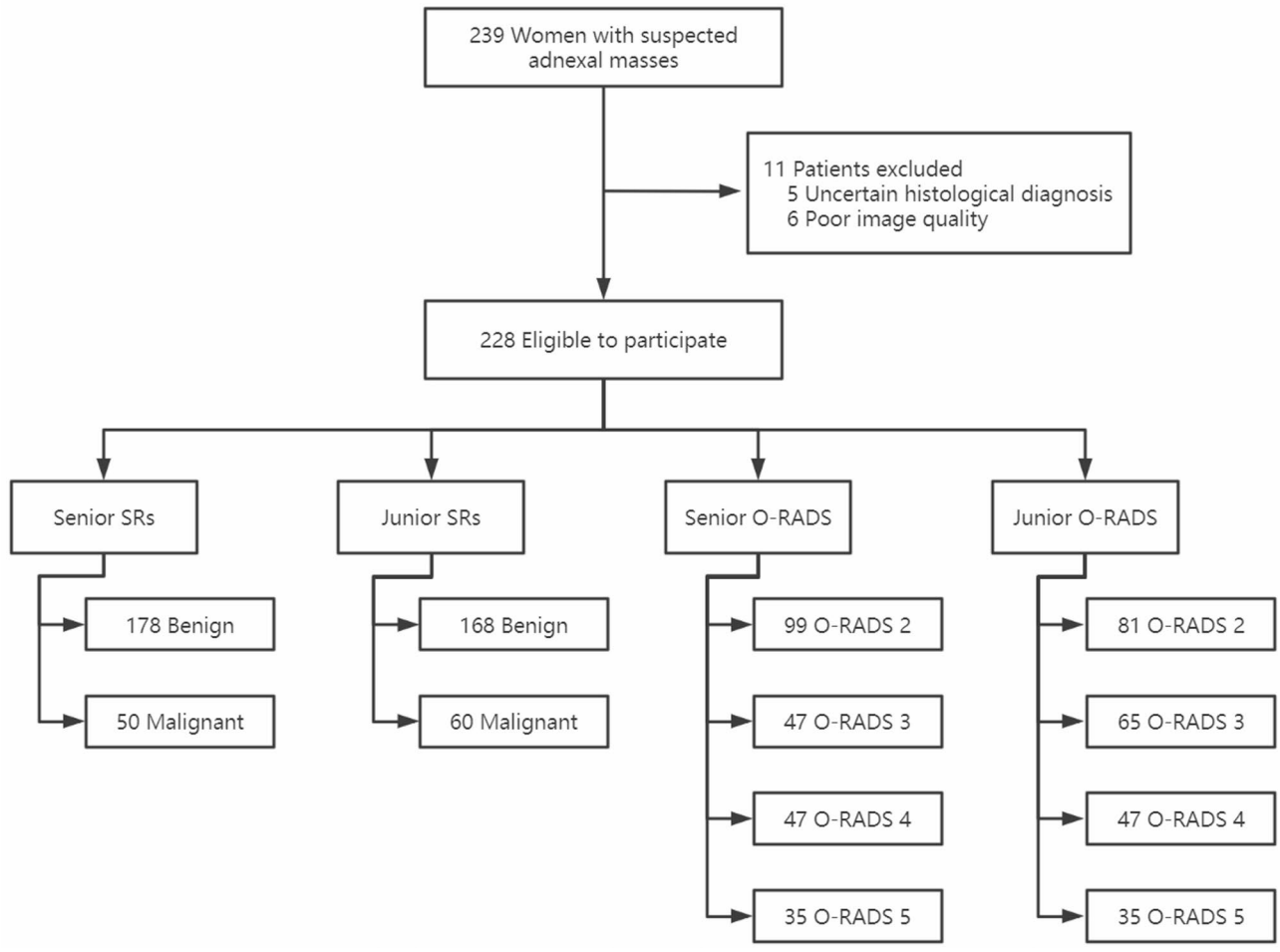

Fig. 1 Flow chart of study population selection. SRs, Simple Rules; O-RADS, Ovarian-Adnexal Reporting and Data Systems

menstrual abnormalities, and unexplained weight loss) were recorded in detail.

Image acquisition and analysis

The US machines used in our study were Nuewa R9 (Mindray Medical). All US images in the study were acquired and interpreted by two senior sonologists with at least 6 years of experience in ovarian-adnexal US in PUMCH. Before participating in this study, all sonologists received theoretical training on the O-RADS US lexicon terms and the risk stratification and management system, which was organized by experienced gynecological sonologists from PUMCH.

Depending on the patient's condition, we performed transabdominal, transvaginal or combined transabdominal and transvaginal US examinations. During the examination, if an ovarian mass was detected, the sonologist was required to perform a thorough evaluation of the mass and to retain separate images (both with and without measurement marker images) in the largest long axis of the lesion and its vertical section and to record the size of the mass. In addition, the section of the lesion with the most abundant blood flow needed to be retained. At the end of the examination, the two senior sonologists jointly provided the SRs and O-RADS US assessments. All images were saved in the picture archiving and communication systems (PACS) of PUMCH.

Then, the US images of all subjects were processed in an anonymized manner and then submitted to two junior sonologist with 2 years and 3 years US experience respectively, neither of whom participated in the image acquisition process (They received training on the IOTA SRs and O-RADS US classification systems and passed the appropriate examinations prior to the image evaluations). During the examination and evaluation, the patient information that was available to the senior and junior sonologists was the patient's age, clinical symptoms, CA125 level, past history, and family history. The two junior sonologists read the images independently and gave their assessments, and for inconsistent assessments, the final unanimous decision was made after discussion between the two sonologists.

The criteria used in the O-RADS US classification of the lesions were the O-RADS US guidelines issued by the

ACR [15]. As mentioned in the guidelines, O-RADS category 4 includes the following four subcategories [15]: (1) multilocular cysts without solid components; (2) unilocular cysts with solid components; (3) multilocular cysts with solid components; and (4) smooth solid masses. As mentioned in some of the studies [17], multilocular cysts without solid components (subcategory 1 above) and smooth solid masses (subcategory 4 above) in the O-RADS 4 category were classified as low-risk O-RADS 4a, and the remaining unilocular or multilocular cysts with solid components (subcategory 2&3 above) were classified as high-risk O-RADS 4b. In the present study, we utilized this classification method to reclassify lesions and define them as adjusted O-RADS. In this study, we calculated the cut-off values of O-RADS US before and after adjustment separately.

At the same time, the lesions were also classified into Begin (B) group and Malignant (M) group according to the SRs proposed by the IOTA Group [18]. Lesions classified as inconclusive by the SRs were classified into group B or M after a subjective assessment by the sonologists, and this classification was based on their own experience.

Reference standards

The postoperative pathological findings of the patients were used as the gold standard for diagnosis, and because borderline tumors have the same intervention as malignant tumors in clinical practice, they were also classified as malignant tumors in the study process [17].

Table 1 Pathological types of the 228 adnexal masses

Type of pathology	No. (%)
Benign adnexal masses	176 (77.19)
Endometriosis cysts	54 (30.68)
Mucinous cystadenoma	16 (9.09)
Serous cystadenoma	31 (17.61)
Mature cystic teratoma	40 (22.73)
Ovarian fibroma/follicular membrane fibroma	3 (1.7)
Salpingitis/hydrosalpinx/tubal cyst	12 (6.82)
Corpus luteum cysts/simple cysts > 3 cm or ovarian hemorrhagic cysts	14 (7.95)
Round/broad ligament myoma	3 (1.7)
Follicular Membranous Cell Tumor	2 (1.14)
Brenner's tumor	1 (0.57)
Malignant adnexal masses	52 (22.81)
Mucinous cystadenocarcinoma	6 (11.54)
Serous cystadenocarcinoma	20 (38.46)
Borderline mucinous cystadenoma	2 (3.85)
Borderline serous cystadenoma	5 (9.62)
Granular cell tumor	1 (1.92)
Endometrioid carcinoma	3 (5.77)
Krukenberg's tumor	5 (9.62)
Clear cell carcinoma	7 (13.46)
Immature cystic teratoma	2 (3.85)
Mixed neuroendocrine carcinoma	1 (1.92)

No., Number

Data analysis

We analyzed the study data using SPSS version 25.0 (IBM Corporation, Armonk, NY) and Medcalc version 20.0.22 (MedCalc Software, Ostend, Belgium) software. Continuous variables were expressed as the means ± standard deviation, and categorical variables were expressed as the numbers and percentages. Comparisons of categorical variables were made using the chi-square test, and comparisons of continuous variables were made using the two independent samples t test. The receiver operating characteristic (ROC) curve was applied to calculate and compare the AUCs and to determine the optimal cutoff value. Comparison of AUC values between different US classification systems was performed by DeLong's test, calculated with the help of MedCalc 20.0.22 software. All tests were two-tailed, and $P<0.05$ indicated a statistically significant difference.

Interobserver agreement was calculated using Cohen's Kappa, calculated with the help of SPSS version 25.0 software. The kappa value (κ) was used to compare the interobserver agreement between the senior and junior sonologists and the agreement between each US classification method and the gold standard pathological diagnosis. Kappa values of 0.0-0.20 indicated poor agreement, 0.21–0.40 indicated fair agreement, 0.41–0.60 indicated moderate agreement, 0.61–0.80 indicated good agreement, and 0.81-1.00 indicated very good agreement.

Results

Patient characteristics and lesion condition

During this study, 228 patients diagnosed with AMs were recruited. The flow chart of the study population selection process was shown in Fig. 1. Among the 228 AMs included, there were 176 benign lesions (77.19%) and 52 malignant lesions (22.81%). The specific pathological types were detailed in Table 1.

The mean age of these patients was 40.52 ± 13.10 years (range, 16–77 years), and the mean age of patients with malignant lesions (47.67 ± 14.80 years) was significantly higher than that of the patients with benign lesions (38.40 ± 11.79 years) ($P<0.001$).

Table 2 listed the clinical characteristics of the patients and the characteristics associated with the lesions. The maximum diameter of the malignant lesions (10.53 ± 4.84 cm) was significantly larger than that of the benign lesions (7.29 ± 3.19 cm) ($P<0.001$), and the type of lesions and the blood flow score were associated with the benignity and malignancy of the tumors ($P<0.001$).

Table 2 Clinical characteristics and lesions of the patients

	Benign, *n* (%)	Malignant, *n* (%)	Total	*P* value
Age (years)	38.40 ± 11.79	47.67 ± 14.80		< 0.001
Age at menarche (years)	13.56 ± 1.42	14.15 ± 2.10		0.061
BMI	22.52 ± 3.54	22.38 ± 3.25		0.795
With or without clinical symptoms				< 0.001
No	91 (89.22)	11 (10.78)	102	
Yes	85 (67.46)	41 (32.54)	126	
Maximum diameter of the lesion (cm)	7.29 ± 3.19	10.53 ± 4.84		< 0.001
Type of lesion				< 0.001
Unilocular cysts	71 (98.61)	1 (1.39)	72	
Multilocular cysts	58 (93.55)	4 (6.45)	62	
Cystic lesions with a solid component	28 (58.33)	20 (41.67)	48	
Masses with a solid or predominant solid component	19 (41.30)	27 (58.70)	46	
Blood flow				< 0.001
1	83 (100)	0	83	
2	83 (80.58)	20 (19.42)	103	
3	9 (31.03)	20 (68.97)	29	
4	1 (7.69)	12 (92.31)	13	

Having clinical symptoms means that the patient had abdominal distension, abdominal pain, abdominal mass, vaginal bleeding or fluid discharge, abnormal menstruation, and unexplained weight loss

Data in parentheses are percentages

Classification results using the two US classification systems

The final diagnostic results of the experienced sonologists were shown in Table 3. Of the 228 lesions included in the study, 99 were classified as O-RADS 2, 47 as O-RADS 3, 47 as O-RADS 4, and 35 as O-RADS 5, and the malignancy rates were 0%, 6.38%, 40.43%, and 85.71%, respectively, with statistically significant differences ($P < 0.001$). By combining the SRs with subjective assessment, 178 of the 228 lesions were included in group B, and 50 were included in group M. The malignancy rates were 3.93% and 90%, respectively, with a statistically significant difference ($P < 0.001$). The final diagnostic results of the inexperienced sonologists were shown in Fig. 2.

The interobserver agreement of the two US classification systems between the senior and junior sonologists

The interobserver agreement between the senior and junior sonologists was as follows (see Additional files S1 and S2): SRs was good (κ = 0.618), O-RADS US was moderate (κ = 0.465), unadjusted O-RADS US was good (κ = 0.657), and adjusted O-RADS US was good (κ = 0.718).

Comparison of the diagnostic validity of the two US classification systems

When > O-RADS 4a was used as a predictor of malignant tumors, 11 lesions were downgraded to the benign category, of which 1 malignant lesion was wrongly downgraded (Fig. 3), and 2 lesions diagnosed as malignant by SRs were accurately downgraded (Fig. 4).

The diagnostic validity and ROC curves of the two US classification systems were shown in Table 4 and Fig. 5, respectively. The ROC curves showed that the unadjusted O-RADS US classification system had a cut-off value of O-RADS 3 and the adjusted O-RADS US classification system had a cut-off value of O-RADS 4a. The unadjusted O-RADS US was dichotomised using the > O-RADS 3 represents malignancy and the adjusted O-RADS US was dichotomised using the > O-RADS 4a represents malignancy. There were statistically significant differences in the ROC curves among the junior unadjusted and adjusted O-RADS US and SRs ($P = 0.0003$ and $P = 0.0001$, respectively). Among them, the junior adjusted O-RADS US had the highest diagnostic validity, with a sensitivity, specificity, and accuracy of 94.23%, 87.50%, and 89.04%, respectively, and the AUC was 0.959 (95% CI, 0.924–0.980). Compared with the junior SRs and unadjusted O-RADS US, the difference in the AUC was 0.118 ($P = 0.0001$) and 0.008 ($P = 0.0295$),

Table 3 Results of the two US classification systems of the experienced sonologists

Category	Total (*n* = 228)	Benign (*n* = 176)	Malignant (*n* = 52)	Observed malignancy rate (%)	Guideline specified malignancy rate (%)	*P* value
O-RADS US						< 0.001
2	99 (43.42)	99 (56.25)	0 (0.00)	0.00	< 1	
3	47 (20.61)	44 (25.00)	3 (5.77)	6.38	1 - <10	
4	47 (20.61)	28 (15.91)	19 (36.54)	40.43	10-<50	
4a	16 (7.02)	15 (8.52)	1 (1.92)			
4b	31 (13.60)	13 (7.39)	18 (34.62)			
5	35 (15.35)	5 (2.84)	30 (57.69)	85.71	≥ 50	
SRs						< 0.001
B group	178(78.07)	171(97.16)	7(13.46)	3.93		
M group	50(21.93)	5(2.84)	45(86.54)	90.00		

O-RADS US, Ovarian-Adnexal Reporting and Data Systems ultrasound; SRs, Simple Rules; B, Begin; M, Malignant

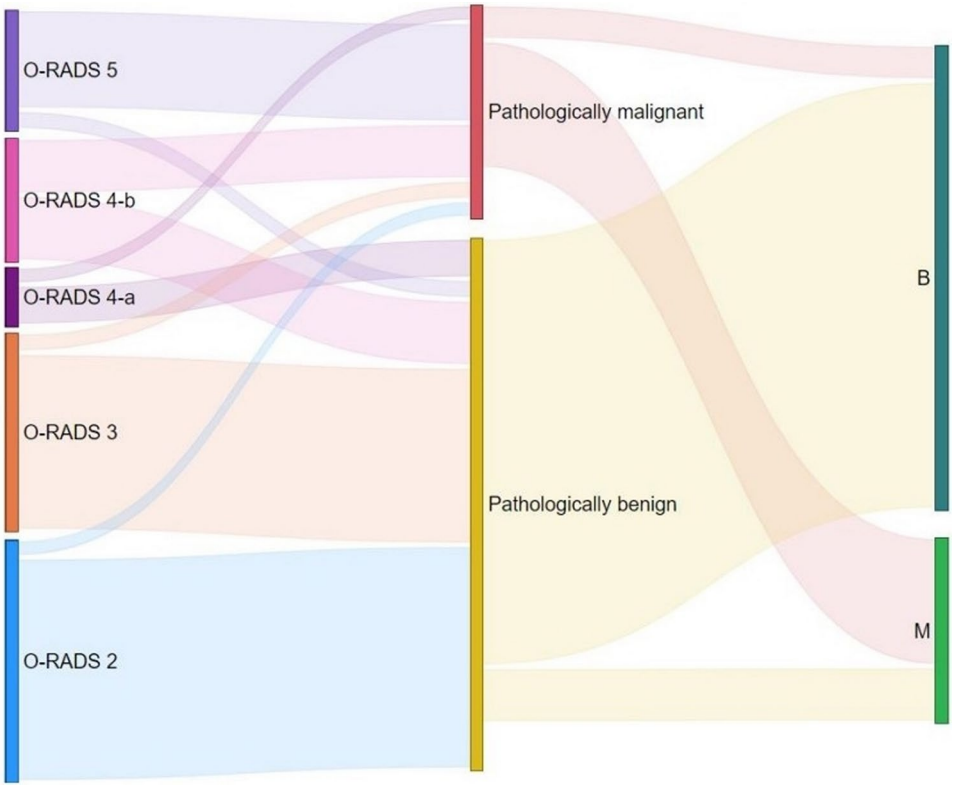

Fig. 2 Sankey diagram of the final diagnosis of the junior sonologists. O-RADS, Ovarian-Adnexal Reporting and Data Systems; SRs, Simple Rules; B, Begin; M, Malignant

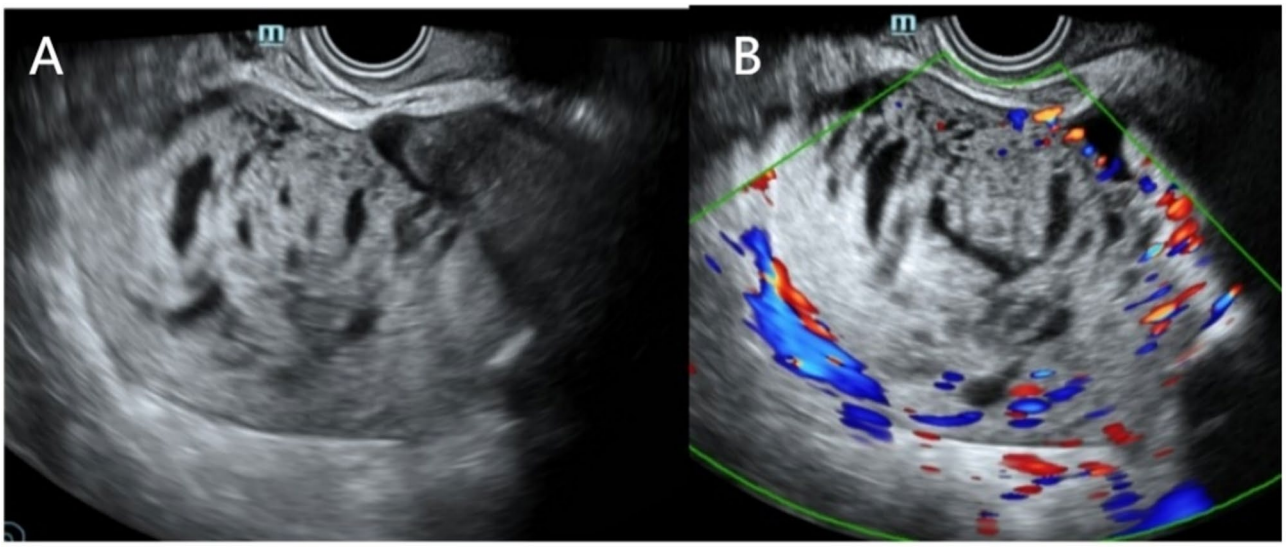

Fig. 3 Case of malignant lesion was wrongly downgraded. Pathology: Mucinous cystadenocarcinoma. **A** The B-mode US showed a regular mass with predominantly solid components, **B** Moderate amount of blood flow within the lesion (Color Score = 3). During the evaluation, the junior observers classified the lesion into O-RADS category 4, adjusted to O-RADS 4a, and the result of junior SRs was M. O-RADS, Ovarian-Adnexal Reporting and Data Systems; SRs, Simple Rules; B, Begin; M, Malignant

respectively. It was followed by the unadjusted O-RADS US, with a sensitivity, specificity, and accuracy of 96.15%, 81.82%, and 85.09%, respectively. And the AUC was 0.951 (95% CI, 0.914–0.975). The difference between the AUC of the junior unadjusted O-RADS US and SRs was 0.111 ($P = 0.0003$). The diagnostic validity of the SRs was slightly lower than the former two, with a sensitivity, specificity, and accuracy of 84.62%, 90.91%, and 89.47%, respectively.

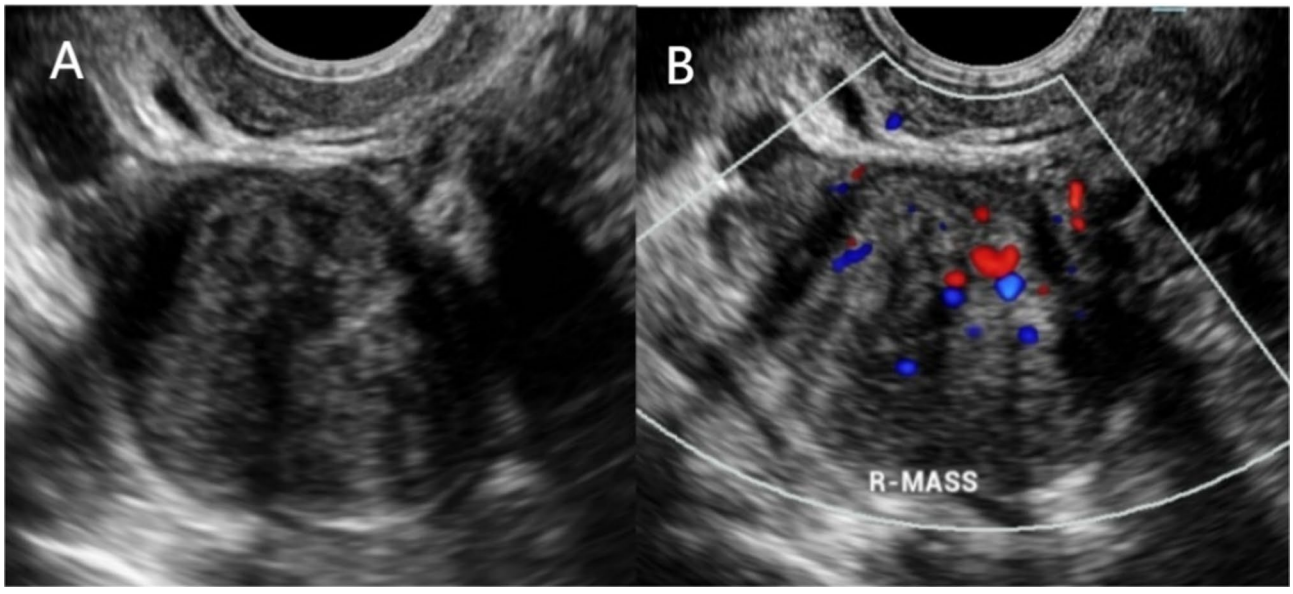

Fig. 4 Case of benign lesion was successfully downgraded. Pathology: Broad ligament leiomyoma. **A** The B-mode US showed a regular solid mass. **B** Moderate amount of blood flow within the lesion (Color Score = 3). During the evaluation, the junior observers classified the lesion as O-RADS category 4, adjusted to O-RADS category 4a, and the junior SRs was M. O-RADS, Ovarian-Adnexal Reporting and Data Systems; SRs, Simple Rules; B, Begin; M, Malignant

Table 4 Diagnostic validity of the two US classification systems

	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	AUC	*P* value
Senior SRs	86.54	97.16	94.74	90.00	96.07	0.918	0.0348[a]
Junior SRs	84.62	90.91	89.47	73.33	95.24	0.840	0.0001[b]
Junior unadjusted O-RADS US	96.15	81.82	85.09	60.98	98.63	0.951	0.0295[c]
Junior adjusted O-RADS US	94.23	87.5	89.04	69.01	98.09	0.959	0.1125[d]

P[a] represents the difference between the AUC of senior SRs and junior SRs, P[b] represents the difference between the AUC of junior SRs and junior adjusted O-RADS, and P[c] represents the difference between the AUC of junior unadjusted O-RADS US and junior adjusted O-RADS US. P[d] represents the difference between the AUC of senior SRs and junior adjusted O-RADS US

SRs, Simple Rules; O-RADS US, Ovarian-Adnexal Reporting and Data Systems ultrasound; adjusted O-RADS, > O-RADS 4a represents malignancy; unadjusted O-RADS, > O-RADS 3 represents malignancy

And the AUC was 0.878 (95% CI, 0.786–0.885). However, the junior unadjusted O-RADS US, adjusted O-RADS US, and SRs all had lower diagnostic accuracy than the senior SRs.

A comparison of the diagnostic agreement between the two US classification systems and the gold standard was shown in Table 5. The senior SRs showed very good diagnostic agreement with the pathological findings (κ = 0.848), and the junior unadjusted O-RADS US, adjusted O-RADS US, and SRs all showed good diagnostic agreement with the pathological findings (κ = 0.648, 0.724, and 0.716, respectively).

Discussion

The IOTA SRs have been widely validated and incorporated into international guidelines; at the same time, due to their simplicity of use, they are also very popular in clinical applications [19]. Considering that the O-RADS US classification system was recently proposed, the literature research and clinical application of the O-RADS US classification system are relatively limited [20]. As described in most studies, the best way to differentiate benign and malignant masses by US is a subjective assessment of the findings by an experienced sonologist [19, 21–23]. With the help of standardized US classification systems, the diagnostic accuracy of junior sonologists has been effectively improved [24]. The primary focus of this study was to evaluate the diagnostic performance of O-RADS US and IOTA SRs among junior sonologists, as their diagnostic accuracy is often lower compared to experienced sonologists. By assessing the performance of adjusted and unadjusted O-RADS US in this group, we aimed to identify a more suitable diagnostic model for

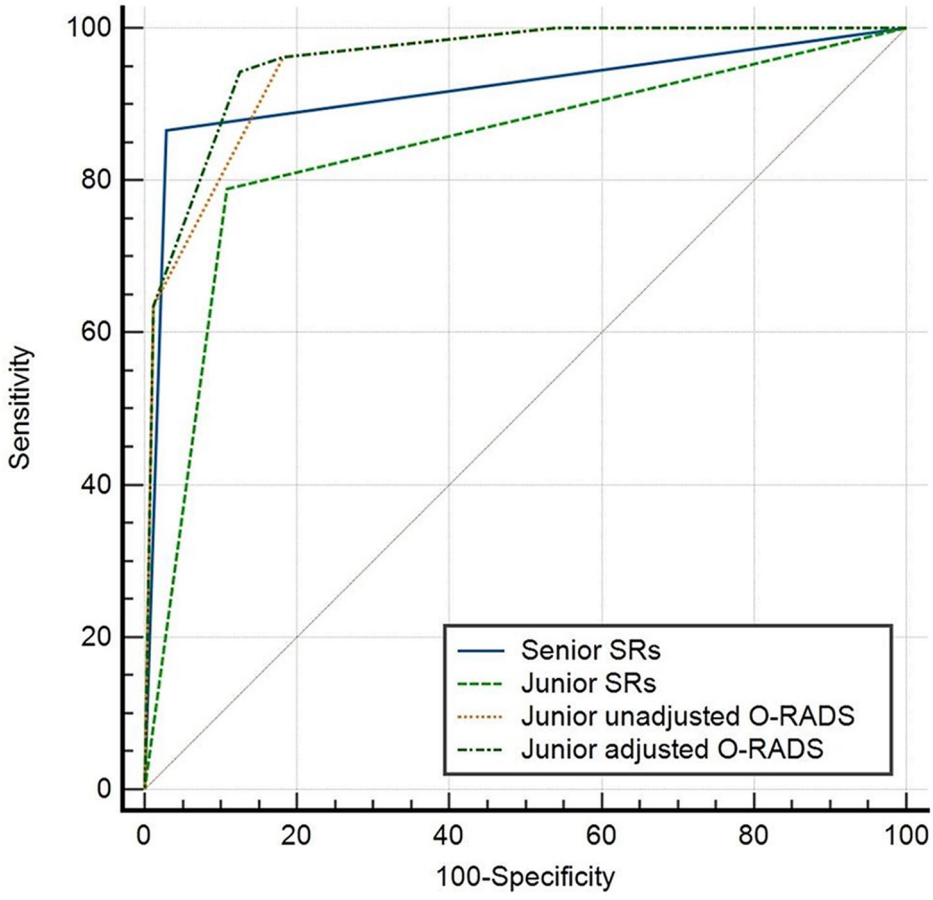

Fig. 5 ROC curves of the two US classification systems. SRs, Simple Rules; O-RADS, Ovarian-Adnexal Reporting and Data Systems; adjusted O-RADS, > O-RADS 4a represents malignancy; unadjusted O-RADS, > O-RADS 3 represents malignancy

Table 5 Comparison of the two US classification systems with the gold standard

	Total	Pathological benign	Pathological malignancy	K value
Senior SRs				0.848
B group	178	171	7	
M group	50	5	45	
Junior SRs				0.716
B group	168	160	8	
M group	60	16	44	
Junior unadjusted O-RADS US				0.648
> O-RADS 3	82	32	50	
≤O-RADS 3	146	144	2	
Junior adjusted O-RADS US				0.724
> O-RADS 4a	71	22	49	
≤O-RADS 4a	157	154	3	

SRs, Simple Rules; O-RADS US, Ovarian-Adnexal Reporting and Data Systems ultrasound; B, Begin; M, Malignant; adjusted O-RADS US, > O-RADS 4a represents malignancy; unadjusted O-RADS US, > O-RADS 3 represents malignancy

K value of 0.0-0.20 indicated poor agreement, 0.21–0.40 indicated fair agreement, 0.41–0.60 indicated moderate agreement, 0.61–0.80 indicated good agreement, and 0.81-1.00 indicated very good agreement

general clinical use, particularly in settings where experienced sonologists may not be available.

As in previous studies, when pathology was the reference standard, the malignancy rates for each category of O-RADS US in the present study were consistent with the guideline-defined malignancy rates [13, 15, 17]. As mentioned in the study by Cao, L et al., O-RADS category 4 recommended in the guidelines has a malignancy risk of 10%-50%, similar to the inconclusive category in the IOTA SRs [17]. Therefore, in the current research, we further divided O-RADS category 4 into two subcategories (O-RADS 4a and 4b), and subjective assessment was used to further classify the inclusive lesions based on the IOTA SRs. Compared with the unadjusted O-RADS US, the AUC of the adjusted O-RADS US was significantly improved ($P = 0.0295$). In our study, the junior SRs combined with subjective assessment and O-RADS US before and after adjustment all had high diagnostic accuracy; however, both the unadjusted and adjusted O-RADS US classification systems had significantly higher diagnostic performance than the SRs (AUC 0.951 and 0.959 vs. 0.840, $P = 0.0003$, 0.0001, respectively). Compared with

the SRs, the O-RADS US lexicon and classification system had detailed definitions for each category of lesions, and there was therefore relatively little dependence on the experience of the observers. This may be why the diagnostic performance of the O-RADS US system was higher than that of the SRs in assisting junior sonologists in diagnosing these lesions.

Compared with the junior SRs, the O-RADS US had a relatively higher sensitivity; coupled with its detailed and comprehensive description of the management of lesions, the O-RADS US system may have more advantages than the SRs for clinical diagnosis and management. However, O-RADS US was less specific than SRs, which may lead to clinical overtreatment of ovarian masses [13]. In our study, multilocular cysts and smooth-solid masses in senior O-RADS category 4 tended to appear benign. In senior O-RADS category 4 lesions, when we divided multilocular cysts and smooth solid masses into subcategory 4a and the rest of the lesions into subcategory 4b, we found that the malignancy rates of subtypes 4a and 4b were 1.92% and 34.62%, respectively. Furthermore, compared with the unadjusted O-RADS US, the diagnostic specificity of the adjusted O-RADS US was improved. Therefore, more studies are needed for further validation and revision before the O-RADS US system is formally applied in clinical practice.

In our research, the interobserver agreement between the senior and junior sonologists using the SRs was good, and using O-RADS US was moderate, and both were less than the interobserver agreement among experienced sonologists [25, 26]. During this research, we found that, when using the O-RADS US, there were large differences in the classification of O-RADS 2 and 3 categories between the senior and junior sonologists, which may be related to the typical benign lesions mentioned in the O-RADS US classification system. Because of their inexperience, junior sonologists were less able to identify typical benign lesions, and they were therefore more inclined to classify them according to unilocular, multilocular cysts, and solid masses, which may be why the κ was lower in the O-RADS US than in the SRs. However, either O-RADS 2 or 3 tended to be benign; thus, the interobserver agreement of the O-RADS US system was significantly improved after using a binary classification. When comparing the assessment results of the junior sonologists with the pathology, all showed good agreement, and junior adjusted O-RADS US was slightly higher than the other two. In conclusion, in this study, the O-RADS US classification system, especially the adapted O-RADS US, was more suitable for use by the inexperienced sonologists.

This study has the following limitations: (1) In the present study, the images used by the junior sonologists were obtained from senior sonologists, which may lead to an increase in the accuracy of the assessment, and it is hoped that subsequent studies can be designed for independent acquisition and assessment by junior sonologists. (2) The sample size included in our study was small and should be further expanded. (3) Only patients hospitalised for surgery for AMs were included in this study, while lesions with poor image quality (O-RADS category 0) and normal ovaries (O-RADS category 1) were excluded, meanwhile, patients with multiple AMs were included with only one AM in the highest O-RADS US category, which may lead to selection bias.

Conclusion

In conclusion, both the O-RADS US and IOTA SRs had high diagnostic value in assisting sonologists of different seniority in making a diagnosis. Multilocular cysts and smooth-solid masses in O-RADS category 4 tended to appear benign. When > O-RADS 4a was used as a predictor of malignant tumors, the specificity was significantly improved without significantly reducing the sensitivity. Meanwhile, compared with unadjusted O-RADS US and SRs, adjusted O-RADS US was more consistent with pathological diagnosis, had higher diagnostic efficacy, and was more suitable for general clinical application.

Abbreviations

AMs	Adnexal masses
SRs	Simple Rules
US	Ultrasound
IOTA	International Ovarian Tumor Analysis
O-RADS US	Ovarian-Adnexal Reporting and Data Systems ultrasound
ACR	American College of Radiology
ROC	Receiver Operating Characteristic
AUC	Area under the curve
TVS	Transvaginal Ultrasonography
B	Benign
M	Malignant
adjusted O-RADS US	> O-RADS 4a represents malignancy
unadjusted O-RADS US	> O-RADS 3 represents malignancy

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01845-4.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Not applicable.

Author contributions

Ya Yang: Data curation, Manuscript writing and editing. Hongyan Wang: Data curation, Manuscript review and editing. Na Su: Data analysis, data interpretation, Manuscript review. Luying Gao: Data analysis, data interpretation, Manuscript review. Yang Gu: Data analysis, Manuscript review. Siman Cai: Manuscript review. Qing Dai: Conceptualization, Manuscript review. Jianchu Li: Conceptualization, Manuscript review. Yuxin Jiang: Conceptualization, Manuscript review.

Funding
This work was supported by the International Health Exchange and Cooperation Center [grant number: ihecc2020C20032], and the National High Level Hospital Clinical Research Funding [grant number: 2022-PUMCH-B-064].

Data availability
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
This study was approved by the ethics review committee of the Peking Union Medical College Hospital (PUMCH), and all methods in this study were carried out in accordance with relevant guidelines and regulations (declaration of helsinki). Informed consent was obtained from all subjects and/or their legal guardians prior to the ultrasound examination.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Ultrasound, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No.1 Shuai Fu Yuan, Dong Cheng District, Beijing 100730, China

Received: 29 January 2023 / Accepted: 24 July 2025
Published online: 28 July 2025

References

1. Gorski JW, Dietrich CS 3rd, Davis C, et al. Significance of pelvic fluid observed during ovarian cancer screening with transvaginal sonogram. Diagnostics (Basel). 2022;12(1):144.
2. Matulonis UA, Sood AK, Fallowfield L, et al. Ovarian cancer. Nat Rev Dis Primers. 2016;2:16061. https://doi.org/10.1038/nrdp.2016.61.
3. Torre LA, Trabert B, DeSantis CE, et al. Ovarian cancer statistics, 2018. CA Cancer J Clin. 2018;68(4):284–96. Epub 2018 May 29.
4. Rao A, Carter J. Ultrasound and ovarian cancer screening: is there a future? J Minim Invasive Gynecol. 2011 Jan-Feb;18(1):24–30.
5. Jacobs IJ, Menon U, Ryan A, et al. Ovarian cancer screening and mortality in the UK collaborative trial of ovarian cancer screening (UKCTOCS): a randomised controlled trial. Lancet. 2016;387(10022):945–56.
6. Menon U, Talaat A, Jeyarajah AR, et al. Ultrasound assessment of ovarian cancer risk in postmenopausal women with CA125 elevation. Br J Cancer. 1999;80(10):1644–7.
7. Guo B, Lian W, Liu S, et al. Comparison of diagnostic values between CA125 combined with CA199 and ultrasound combined with CT in ovarian cancer. Oncol Lett. 2019;17(6):5523–8.
8. Saorin A, Di Gregorio E, Miolo G, et al. Emerging role of metabolomics in ovarian cancer diagnosis. Metabolites. 2020;10(10):419.
9. Sehouli J, Grabowski JP. Surgery in recurrent ovarian cancer. Cancer. 2019;125(Suppl 24):4598–601.
10. Roett MA, Evans P. Ovarian cancer: an overview. Am Fam Physician. 2009;80(6):609–16.
11. Timmerman D, Schwärzler P, Collins WP, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol. 1999;13(1):11–6.
12. Rosati A, Alletti SG, Capozzi VA, et al. Role of ultrasound in the detection of recurrent ovarian cancer: a review of the literature. Gland Surg. 2020;9(4):1092–101.
13. Basha MAA, Metwally MI, Gamil SA, et al. Comparison of O-RADS, GI-RADS, and IOTA simple rules regarding malignancy rate, validity, and reliability for diagnosis of adnexal masses. Eur Radiol. 2021;31(2):674–84.
14. Timmerman D, Ameye L, Fischerova D, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. BMJ. 2010;341:c6839.
15. Andreotti RF, Timmerman D, Strachowski LM, et al. O-RADS US risk stratification and management system: A consensus guideline from the ACR Ovarian-Adnexal reporting and data system committee. Radiology. 2020;294(1):168–85.
16. Guo Y, Zhou S, Zhao B, et al. Ultrasound findings and O-RADS malignancy risk stratification of ovarian collision tumors. J Ultrasound Med. 2022;41(9):2325–31.
17. Cao L, Wei M, Liu Y, et al. Validation of American college of radiology Ovarian-Adnexal reporting and data system ultrasound (O-RADS US): analysis on 1054 adnexal masses. Gynecol Oncol. 2021;162(1):107–12.
18. Timmerman D, Testa AC, Bourne T, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. Ultrasound Obstet Gynecol. 2008;31(6):681–90.
19. Froyman W, Wynants L, Landolfo C, et al. Validation of the performance of international ovarian tumor analysis (IOTA) methods in the diagnosis of early stage ovarian cancer in a Non-Screening population. Diagnostics (Basel). 2017;7(2):32.
20. Pi Y, Wilson MP, Katlariwala P, et al. Diagnostic accuracy and inter-observer reliability of the O-RADS scoring system among staff radiologists in a North American academic clinical setting. Abdom Radiol (NY). 2021;46(10):4967–73.
21. Valentin L, Hagen B, Tingulstad S, et al. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. Ultrasound Obstet Gynecol. 2001;18(4):357–65.
22. Timmerman D. The use of mathematical models to evaluate pelvic masses; can they beat an expert operator? Best Pract Res Clin Obstet Gynaecol. 2004;18(1):91–104.
23. Meys EMJ, Kaijser J, Kruitwagen R F P M, et al. Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis. Eur J Cancer. 2016;58:17–29.
24. Guo Y, Zhao B, Zhou S, et al. A comparison of the diagnostic performance of the O-RADS, RMI4, IOTA LR2, and IOTA SR systems by senior and junior Doctors. Ultrasonography. 2022;41(3):511–8.
25. Lai HW, Lyu G, Kang Z, et al. Comparison of O-RADS, GI-RADS, and ADNEX for diagnosis of adnexal masses: an external validation study conducted by junior sonologists. J Ultrasound Med. 2022;41(6):1497–507.
26. Niemi RJ, Saarelainen SK, Luukkaala TH, et al. Reliability of preoperative evaluation of postmenopausal ovarian tumors. J Ovarian Res. 2017;10(1):15.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.