

Comparative Diagnostic Performance of IOTA Simple Rules, ADNEX, GI-RADS, and O-RADS for Characterization of Adnexal Masses Using Histopathology as the Reference Standard

Varsha Shetty¹, Riya Deb², Abinaya Thenmozhi Senthilnathan³, Abiramisree Sanniyasi Senthilnathan⁴

¹Assistant Professor, Department of Radiology, The Oxford Medical College, Hospital and Research Centre, Bangalore, Karnataka, India

²Senior Resident, Department of Obstetrics and Gynaecology, Jalpaiguri Government Medical College & Hospital, Jalpaiguri, West Bengal, India

³Duty Medical Officer, Department of Gastroenterology, Srinivasa Medical College and Hospital, Tiruchirappalli, Tamil Nadu, India

⁴Duty Medical Officer, Department of Emergency Medicine, A to Z Speciality Hospital, Thanjavur, Tamil Nadu, India

Received: 01-06-2026 / Revised: 15-07-2026 / Accepted: 21-08-2026

Corresponding author: Dr. Abiramisree Sanniyasi Senthilnathan

Conflict of interest: Nil

Abstract

Background: Accurate preoperative characterization of adnexal masses is essential for appropriate patient triage and management. Several ultrasound-based systems, including the International Ovarian Tumor Analysis Simple Rules (IOTA SR), Assessment of Different NEoplasias in the adneXa (ADNEX), Gynecologic Imaging Reporting and Data System (GI-RADS), and Ovarian-Adnexal Reporting and Data System (O-RADS), have been developed for malignancy risk stratification.

Objective: To evaluate and compare the diagnostic performance of IOTA SR, ADNEX, GI-RADS, and O-RADS in differentiating benign from malignant/borderline adnexal masses.

Methods: This analytical cross-sectional study included 193 adnexal masses evaluated by standardized ultrasonography and classified using IOTA SR, ADNEX, GI-RADS, and O-RADS. Histopathological diagnosis served as the reference standard. Sensitivity, specificity, predictive values, accuracy, likelihood ratios, area under the receiver operating characteristic curve (AUC), and agreement with histopathology were assessed.

Results: Of 193 masses, 140 (72.5%) were benign and 53 (27.5%) were malignant/borderline. Malignant/borderline lesions occurred in older patients and were associated with higher CA-125 levels, larger lesion size, solid components, ascites, irregular morphology, and stronger Doppler vascularity (all $p < 0.001$). O-RADS demonstrated the highest sensitivity (96.2%) and negative predictive value (98.4%), with specificity of 90.7% and AUC of 0.970. GI-RADS showed sensitivity of 92.5%, specificity of 92.9%, accuracy of 92.7%, and AUC of 0.953. ADNEX demonstrated sensitivity of 94.3%, specificity of 88.6%, and AUC of 0.944. Among conclusive IOTA SR assessments, sensitivity, specificity, and accuracy were 93.6%, 96.8%, and 96.0%, respectively.

Conclusion: All four ultrasound-based systems demonstrated excellent diagnostic performance. O-RADS provided the highest sensitivity, whereas conclusive IOTA SR showed the highest specificity and accuracy, supporting the use of standardized ultrasound risk-stratification systems for preoperative evaluation of adnexal masses.

Keywords: Adnexal mass, O-RADS, GI-RADS, IOTA Simple Rules, ADNEX model, ovarian malignancy.

DOI: 10.25258/ijcpr.18.9.194

This is an Open Access article that uses a funding model which does not charge readers or their institutions for access and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>) and the Budapest Open Access Initiative (<http://www.budapestopenaccessinitiative.org/read>), which permit unrestricted use, distribution, and reproduction in any medium, provided original work is properly credited.

Introduction

Adnexal masses constitute a heterogeneous group of ovarian, tubal, and paraovarian lesions ranging from physiological cysts and benign neoplasms to borderline tumours, primary ovarian malignancies, and metastatic disease. Their accurate preoperative

characterization is clinically important because management differs substantially between lesions suitable for observation or conservative surgery and those requiring referral to a gynaecologic oncology. The ESGO/ISUOG/IOTA/ESGE consensus

emphasizes ultrasonography as the principal imaging modality for the initial characterization of ovarian and other adnexal tumours and recommends structured assessment using validated ultrasound terminology and prediction models. [1] The clinical importance of appropriately identifying malignant adnexal lesions is underscored by the continuing global burden of ovarian cancer. According to GLOBOCAN 2022, ovarian cancer accounted for approximately 324,603 new cases and 206,956 deaths worldwide, corresponding to an age-standardized incidence rate of 6.7 and mortality rate of 4.0 per 100,000 women. Correct preoperative triage may also influence oncological outcomes; Vernooij et al. found that management in specialized hospitals and by gynaecologic oncologists was associated with improved outcomes, including an estimated 5–8-month median survival advantage among subgroups of women with advanced ovarian cancer. [2]

Clinical characteristics and serum biomarkers alone cannot reliably characterize every adnexal mass. In particular, CA-125 lacks disease specificity because it may be elevated in benign gynaecological conditions; Hirsch et al., in a meta-analysis involving 2,920 women, demonstrated that CA-125 ≥ 30 U/mL had only 52% sensitivity for endometriosis despite a specificity of 93%. [3] Ultrasound morphology therefore remains central to risk assessment, particularly through evaluation of lesion architecture, solid tissue, papillary projections, wall irregularity, acoustic shadows, ascites, and vascularity. To improve objectivity and reproducibility, several standardized ultrasound-based systems have been developed. External validation of earlier IOTA prediction models in 997 women demonstrated AUCs of 0.941–0.956, outperforming non-IOTA models, whose AUCs ranged from 0.839 to 0.928. [4] IOTA Simple Rules subsequently provided a clinically accessible morphology-based approach; Alcázar et al. independently validated the Simple Rules. [5] The ADNEX model further expanded risk prediction by estimating probabilities for benign tumours, borderline tumours, stage I cancer, stage II–IV cancer, and metastatic disease; Sayasneh et al. externally validated ADNEX in 610 women and reported an AUC of 0.937 with CA-125 and 0.925 without CA-125 for discrimination between benign and malignant masses. [6] GI-RADS translates sonographic morphology and vascularity into standardized malignancy-risk; a meta-analysis of 2,474 patients reported pooled sensitivity of 95%, specificity of 86%, and an AUC of 0.981. [7] O-RADS US was subsequently introduced by the American College of Radiology to provide a unified lexicon, defined risk categories, and corresponding management recommendations across the spectrum from normal findings to high-

risk lesions. [8] Given the conceptual and operational differences among these systems, direct comparison is clinically relevant. Against this background, the objective of the present study was to evaluate and compare diagnostic performance of ultrasound-based reporting systems IOTA SR, ADNEX, GIRADS, ORADS for discrimination between benign and malignant adnexal masses.

Materials and Methods

This was a single-centre, hospital-based, analytical cross-sectional study conducted in the Department of Radiology and Department of Obstetrics and Gynaecology of a tertiary teaching healthcare facility in South India over a period of two years (January 2024–December 2025). The study was approved by the Institutional Human Ethics Committee (IHEC). Each participant was provided with a Participant Information Sheet (PIS) translated into their local language. The information was also explained verbally to ensure clear understanding and voluntary agreement. Written informed consent was obtained prior to enrolling participants in the study. Patients with a sonographically detected adnexal mass who underwent ultrasound evaluation using IOTA Simple Rules, ADNEX, GI-RADS, and O-RADS and subsequently had a definitive reference diagnosis based on histopathological examination were included in the study. Patients with incomplete or technically inadequate ultrasound examinations, insufficient clinical or imaging data for application of the reporting systems, adnexal lesions that were not subsequently confirmed by histopathology, or those who had received prior treatment for the adnexal mass before imaging were excluded.

The sample size was calculated using the diagnostic performance reported by Basha et al., who evaluated O-RADS, GI-RADS, and IOTA Simple Rules for characterising adnexal masses and reported an O-RADS sensitivity of 96.8%, specificity of 92.8%, and a malignancy prevalence of 27.5% (178/647 masses). [9] Assuming a 95% confidence level ($Z = 1.96$) and an absolute precision of 5%, the required sample size based on sensitivity was calculated using $N = Z^2 \times Se \times (1-Se)/(d^2 \times P)$; the minimum required sample size was rounded off to 193 participants, after accounting for 10% non-response rate. The participants were enrolled using nonprobability sampling – purposive/convenience sampling technique. Data were collected using a predesigned structured case-record proforma after obtaining written informed consent. Baseline clinical information included age, presenting symptoms, parity, menstrual and menopausal status, relevant medical and surgical history, personal or family history of gynaecological malignancy, and serum cancer antigen 125 (CA-125) level where

applicable. Each participant underwent a pelvic ultrasonographic examination, primarily using the transvaginal approach, supplemented by transabdominal ultrasonography when required for large lesions or lesions extending beyond the field of view. Grey-scale and colour Doppler examinations were performed, and each adnexal mass was systematically evaluated for laterality, maximum lesion diameter, cystic or solid morphology, number of locules, wall and septal characteristics, presence and size of solid components, number of papillary projections, irregularity of the inner wall or solid component, acoustic shadows, ascites, peritoneal abnormalities, and vascularity. Colour Doppler vascularity was graded using the IOTA colour score as 1 (no detectable flow), 2 (minimal flow), 3 (moderate flow), or 4 (very strong flow). The recorded sonographic findings were subsequently classified using all four predefined ultrasound-based diagnostic systems. For the IOTA SR, five benign (B) and five malignant (M) features were assessed. The benign features comprised a unilocular cyst (B1), a solid component with a maximum diameter of <7 mm (B2), acoustic shadows (B3), a smooth multilocular tumour measuring <100 mm (B4), and absence of detectable Doppler flow corresponding to colour score 1 (B5). The malignant features comprised an irregular solid tumour (M1), ascites (M2), at least four papillary structures (M3), an irregular multilocular-solid tumour measuring ≥ 100 mm (M4), and very strong Doppler flow corresponding to colour score 4 (M5).

A lesion was classified as benign when one or more B features were present without an M feature and as malignant when one or more M features were present without a B feature; lesions demonstrating both B and M features and neither type of feature were recorded as inconclusive. For the ADNEX model, the required clinical and ultrasonographic variables were entered into the IOTA ADNEX model. [10] These comprised age in years, serum CA-125 concentration in U/mL, type of clinical centre, maximum lesion diameter in millimetres, proportion of solid tissue calculated as the maximum diameter of the largest solid component divided by the maximum lesion diameter, presence of more than 10 cyst locules, number of papillary projections (0, 1, 2, 3, or >3), presence of acoustic shadows, and presence of ascites. The model-generated probabilities of a benign tumour, borderline tumour, stage I primary invasive malignancy, stage II–IV primary invasive malignancy, and secondary metastatic tumour were recorded, together with the overall calculated risk of malignancy. Each lesion was also assigned a GI-RADS category according to its morphological appearance, pattern-recognition features, and Doppler vascularity: GI-RADS 1 represented a normal examination with no adnexal mass, GI-

RADS 2 indicated a very probably benign or functional lesion with an estimated malignancy risk of <1%, GI-RADS 3 represented a probably benign lesion with a risk of approximately 1–4%, GI-RADS 4 represented a probably malignant lesion with a risk of approximately 5–20%, and GI-RADS 5 represented a very probably malignant lesion with a risk >20%. [11] The same lesions were independently categorized according to the American College of Radiology O-RADS US v2022 using lesion morphology, size, wall and septal regularity, presence and number of papillary projections or other solid components, acoustic shadowing, colour Doppler vascularity, ascites, and peritoneal nodules. O-RADS 0 denoted an incomplete examination, O-RADS 1 a normal ovary, O-RADS 2 an almost certainly benign lesion with a malignancy risk <1%, O-RADS 3 a low-risk lesion with a risk of 1 to <10%, O-RADS 4 an intermediate-risk lesion with a risk of 10 to <50%, and O-RADS 5 a high-risk lesion with a risk $\geq 50\%$. [12] Following surgical management, the excised adnexal specimen was subjected to histopathological examination, and the final histopathological diagnosis was recorded as the reference standard for determining whether the mass was benign or malignant.

Statistical Analysis: Statistical analysis was performed using IBM SPSS Statistics for Windows, Version 27.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean $\pm$ standard deviation or median with interquartile range, depending on data distribution, and were compared using the independent-samples t-test or Mann–Whitney U test, as appropriate. Categorical variables were presented as frequencies and percentages and compared using the chi-square test or Fisher's exact test. The diagnostic performance of IOTA SR, ADNEX, GI-RADS, and O-RADS was assessed by calculating sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and likelihood ratios with 95% confidence intervals. Receiver operating characteristic curves were generated and the area under the curve was calculated with bootstrap confidence intervals based on 3,000 resamples. Agreement with histopathology was assessed using Cohen's kappa coefficient, and overall differences in classification correctness among the four systems were evaluated using Cochran's Q test. A two-tailed p value <0.05 was considered statistically significant.

Results

Among the 193 adnexal masses evaluated, 140 (72.5%) were benign and 53 (27.5%) were malignant or borderline on histopathology. Patients with malignant/borderline lesions were significantly older than those with benign lesions (53.0 ± 10.6 vs. 42.1 ± 11.8 years; $p < 0.001$), and a

greater proportion were postmenopausal (64.2% vs. 30.7%; $p < 0.001$). Serum CA-125 levels were markedly higher in the malignant/borderline group, with a median value of 241.6 U/mL compared with 22.2 U/mL in the benign group ($p < 0.001$), while CA-125 ≥ 35 U/mL was observed in 86.8% and 29.3%, respectively. On ultrasonography, malignant/borderline masses were significantly larger (108.5 ± 41.4 vs. 69.8 ± 29.4 mm; $p < 0.001$) and more frequently demonstrated solid components (73.6% vs. 17.9%), ascites (43.4% vs. 2.9%), very strong Doppler flow (50.9% vs. 5.7%), and irregular walls or solid components (69.8% vs. 12.9%), all of which were statistically significant. Acoustic shadowing was more frequently associated with benign lesions (30.0% vs. 11.3%; $p = 0.007$). Histopathologically, the commonest benign lesion was serous cystadenoma (17.6%), followed by endometrioma (15.0%) and mature cystic teratoma (12.4%), whereas high-grade serous carcinoma was the most frequent malignant lesion (9.8% of the total cohort), followed by borderline epithelial tumours (5.7%).

All four ultrasound-based reporting systems demonstrated strong discrimination between benign and malignant/borderline adnexal masses. Using O-RADS, 75.5% of malignant/borderline lesions were classified as O-RADS 5 and a further 20.8% as O-RADS 4, whereas 90.7% of benign lesions were categorized as O-RADS 2 or 3. Similarly, 69.8% of

malignant/borderline masses were classified as GI-RADS 5, while 92.9% of benign lesions were assigned to GI-RADS 2 or 3. At a 10% malignancy-risk threshold, the ADNEX model identified 94.3% of malignant/borderline lesions as high risk, compared with only 11.4% of benign lesions. IOTA Simple Rules classified 83.0% of malignant/borderline masses as malignant and 87.1% of benign masses as benign, with 20 lesions being inconclusive.

In diagnostic-performance analysis, O-RADS showed the highest sensitivity at 96.2% and an NPV of 98.4%, while GI-RADS demonstrated the highest complete-cohort accuracy among the fully classified systems at 92.7% and a specificity of 92.9%. ADNEX showed a sensitivity of 94.3%, specificity of 88.6%, and accuracy of 90.2%. Among the 173 lesions with conclusive IOTA SR findings, IOTA achieved the highest specificity (96.8%) and accuracy (96.0%). ROC analysis demonstrated excellent discrimination for all systems, with the highest AUC observed for O-RADS at 0.970, followed by GI-RADS at 0.953, IOTA SR at 0.947, and ADNEX at 0.944. Agreement with histopathology was substantial to almost perfect, with Cohen's κ ranging from 0.749 for IOTA SR to 0.824 for GI-RADS. However, the overall proportion of correctly classified lesions did not differ significantly among the four complete-cohort systems (Cochran's $Q = 2.12$, $p = 0.547$).

Table 1: Clinical characteristics according to histopathological diagnosis

	Overall (N=193)	Benign (n=140)	Malignant/borderline (n=53)	p value
Age, years	45.1 $\pm$ 12.5	42.1 $\pm$ 11.8	53.0 $\pm$ 10.6	<0.001
Parity, median (IQR)	2.0 (1.0–3.0)	2.0 (1.0–3.0)	2.0 (1.0–3.0)	0.324
Postmenopausal	77 (39.9%)	43 (30.7%)	34 (64.2%)	<0.001
Abdominal/pelvic pain	122 (63.2%)	86 (61.4%)	36 (67.9%)	0.404
Abdominal bloating/distension	52 (26.9%)	28 (20.0%)	24 (45.3%)	<0.001
Family history of gynaecologic malignancy	13 (6.7%)	8 (5.7%)	5 (9.4%)	0.357
Previous pelvic surgery	43 (22.3%)	30 (21.4%)	13 (24.5%)	0.644
Serum CA-125, U/mL, median (IQR)	35.5 (16.4–113.5)	22.2 (14.8–42.0)	241.6 (151.6–364.8)	<0.001
CA-125 ≥ 35 U/mL	87 (45.1%)	41 (29.3%)	46 (86.8%)	<0.001

Table 2: Ultrasound characteristics according to histopathological diagnosis

	Overall (N=193)	Benign (n=140)	Malignant/borderline (n=53)	p value
Maximum lesion diameter, mm	80.4 $\pm$ 37.3	69.8 $\pm$ 29.4	108.5 $\pm$ 41.4	<0.001
Bilateral adnexal masses	29 (15.0%)	15 (10.7%)	14 (26.4%)	0.006
Solid component present	64 (33.2%)	25 (17.9%)	39 (73.6%)	<0.001
≥ 4 papillary projections	27 (14.0%)	6 (4.3%)	21 (39.6%)	<0.001
> 10 locules	25 (13.0%)	9 (6.4%)	16 (30.2%)	<0.001
Acoustic shadows	48 (24.9%)	42 (30.0%)	6 (11.3%)	0.007
Ascites	27 (14.0%)	4 (2.9%)	23 (43.4%)	<0.001
Very strong Doppler flow (colour score 4)	35 (18.1%)	8 (5.7%)	27 (50.9%)	<0.001
Irregular wall/solid component	55 (28.5%)	18 (12.9%)	37 (69.8%)	<0.001

Table 3: Histopathological spectrum of adnexal masses

Histopathological diagnosis	Diagnostic group	n	% of total
Serous cystadenoma	Benign	34	17.6%
Endometrioma		29	15.0%
Mature cystic teratoma		24	12.4%
Mucinous cystadenoma		21	10.9%
Functional/simple cyst		15	7.8%
Fibroma/thecoma		10	5.2%
Other benign		7	3.6%
High-grade serous carcinoma	Malignant/borderline	19	9.8%
Borderline epithelial tumour		11	5.7%
Mucinous carcinoma		6	3.1%
Endometrioid carcinoma		5	2.6%
Clear-cell carcinoma		4	2.1%
Metastatic tumour		4	2.1%
Other malignant		4	2.1%

Table 4: Distribution of IOTA SR, ADNEX, GI-RADS and O-RADS classifications by histopathology

		Benign (n=140)	Malignant/borderline (n=53)	p value
IOTA Simple Rules	Benign	122 (87.1%)	3 (5.7%)	<0.001
	Inconclusive	14 (10.0%)	6 (11.3%)	
	Malignant	4 (2.9%)	44 (83.0%)	
ADNEX (10% cut-off)	<10% malignancy risk	124 (88.6%)	3 (5.7%)	<0.001
	≥10% malignancy risk	16 (11.4%)	50 (94.3%)	
GI-RADS	GI-RADS 2	62 (44.3%)	1 (1.9%)	<0.001
	GI-RADS 3	68 (48.6%)	3 (5.7%)	
	GI-RADS 4	8 (5.7%)	12 (22.6%)	
	GI-RADS 5	2 (1.4%)	37 (69.8%)	
O-RADS	O-RADS 2	75 (53.6%)	0 (0.0%)	<0.001
	O-RADS 3	52 (37.1%)	2 (3.8%)	
	O-RADS 4	10 (7.1%)	11 (20.8%)	
	O-RADS 5	3 (2.1%)	40 (75.5%)	

Table 5: Diagnostic performance of the ultrasound-based reporting systems

	TP/FN/TN/FP	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)	LR+	LR-
O-RADS (4–5 positive)	51/2/127/13	96.2% (87.2–99.0)	90.7% (84.8–94.5)	79.7% (68.3–87.7)	98.4% (94.5–99.6)	92.2% (87.6–95.2)	10.36	0.04
GI-RADS (4–5 positive)	49/4/130/10	92.5% (82.1–97.0)	92.9% (87.4–96.1)	83.1% (71.5–90.5)	97.0% (92.6–98.8)	92.7% (88.2–95.6)	12.94	0.08
ADNEX (risk ≥10%)	50/3/124/16	94.3% (84.6–98.1)	88.6% (82.2–92.8)	75.8% (64.2–84.5)	97.6% (93.3–99.2)	90.2% (85.1–93.6)	8.25	0.06
IOTA SR, all lesions*	50/3/122/18	94.3% (84.6–98.1)	87.1% (80.6–91.7)	73.5% (62.0–82.6)	97.6% (93.2–99.2)	89.1% (83.9–92.8)	7.34	0.06
IOTA SR, conclusive (n=173)	44/3/122/4	93.6% (82.8–97.8)	96.8% (92.1–98.8)	91.7% (80.4–96.7)	97.6% (93.2–99.2)	96.0% (91.9–98.0)	29.49	0.07

*For the complete-cohort IOTA analysis, inconclusive Simple Rules results were treated as test-positive because such lesions would require additional expert assessment or referral rather than reassurance

Table 6: ROC discrimination and agreement with histopathology

	AUC (95% bootstrap CI)	Cohen κ vs histopathology
O-RADS	0.970 (0.945–0.988)	0.817
GI-RADS	0.953 (0.912–0.984)	0.824
ADNEX	0.944 (0.911–0.972)	0.770
IOTA SR	0.947 (0.907–0.981)	0.749

Bootstrap confidence intervals were based on 3,000 resamples. For complete-cohort dichotomized predictions, Cochran's Q test showed no significant difference in overall correctness among the four systems ($Q=2.12$, $p=0.547$).

Discussion

The present study demonstrated that ultrasound-based risk-stratification systems provided excellent discrimination between benign and malignant/borderline adnexal masses. Of the 193 masses evaluated, 53 (27.5%) were malignant or borderline, a prevalence closely comparable to the 27.5% malignancy rate reported by Basha et al., who evaluated 647 adnexal masses, including 178 malignant lesions. [9] The similarity in prevalence is relevant because predictive values and apparent diagnostic performance can vary substantially according to the proportion of malignant disease within the evaluated population. Patients with malignant/borderline masses in the present study were significantly older than those with benign lesions (53.0 ± 10.6 vs. 42.1 ± 11.8 years; $p<0.001$), and postmenopausal status was also considerably more frequent among malignant/borderline lesions (64.2% vs. 30.7%). This association is biologically and diagnostically plausible and has long been incorporated into ovarian-malignancy risk assessment. Malkasian et al., in 158 women undergoing surgery for pelvic masses, similarly demonstrated that patients with malignancy were significantly older and more frequently postmenopausal than women with benign disease. [13] An Indian tertiary-centre study by Chopra et al. also found a significant association of ovarian malignancy with increasing age and menopausal status. [14]

Serum CA-125 showed marked separation between diagnostic groups, with a median concentration of 241.6 U/mL in malignant/borderline masses compared with 22.2 U/mL in benign masses, while 86.8% versus 29.3%, respectively, had CA-125 ≥ 35 U/mL. These findings support the clinical value of CA-125 as an adjunct rather than a stand-alone discriminator. Malkasian et al. reported CA-125 >65 U/mL in 75% of malignant pelvic masses but in only 8% of benign lesions. [13] More recent evidence also indicates that CA-125 alone is inferior to multivariable ultrasound models: in 581 women, Spagnol et al. reported sensitivity and specificity of 60% and 80% for CA-125 compared with 82% and 91% for ADNEX, with corresponding AUCs of 0.76 and 0.92. [15]

The morphological differences observed on ultrasound further supported the biological distinction between benign and malignant lesions. Malignant/borderline masses were larger (108.5 ± 41.4 vs. 69.8 ± 29.4 mm) and more commonly contained solid components (73.6%), ascites (43.4%), very strong Doppler vascularity (50.9%), and irregular walls or solid components (69.8%). These findings closely correspond to the IOTA malignant features originally described by Timmerman et al., namely an irregular solid tumour, ascites, at least four papillary projections, an irregular multilocular-solid lesion ≥ 100 mm, and very strong colour Doppler flow. [16] In their prospective study, these Simple Rules produced a sensitivity of 95% and specificity of 91%. Conversely, acoustic shadowing was significantly more frequent in benign masses in the present study (30.0% vs. 11.3%; $p=0.007$). Acoustic shadows are specifically included among the IOTA benign features, and more recent O-RADS research has reinforced their benign predictive value. Wu et al. evaluated 1,061 adnexal masses and showed that incorporating acoustic shadowing as a benign feature increased O-RADS specificity by 5.4%; the modified approach achieved 94.4% sensitivity, 81.6% specificity and an NPV of 98.5%. [17]

Among the four evaluated systems, O-RADS achieved the highest sensitivity (96.2%) and NPV (98.4%), with an AUC of 0.970. This result closely mirrors Basha et al., who reported significantly greater sensitivity for O-RADS than for GI-RADS and IOTA Simple Rules (96.8% vs. 92.7% and 92.1%, respectively), while specificity remained comparable at 92.8%, 93.6% and 93.2%. [9] The present O-RADS specificity of 90.7% and LR[−] of 0.04 therefore fall within an expected high-performance range. A head-to-head meta-analysis subsequently reported pooled O-RADS sensitivity of 95.1%, specificity of 88.8%, LR⁺ of 8.5, LR[−] of 0.06 and AUC of 0.97, remarkably similar to the current AUC of 0.970. [18] The distribution of malignancy across O-RADS was also clinically coherent. In the present cohort, no malignant/borderline lesion was classified as O-RADS 2, only 3.8% were O-RADS 3, whereas 20.8% and 75.5% were categorized as O-RADS 4 and 5, respectively. Hack et al. similarly reported malignancy rates of 0%, 3%, 35% and 78% for O-

RADS categories 2, 3, 4 and 5, respectively, in an external validation cohort. [19] Their use of O-RADS ≥ 4 yielded 99% sensitivity and 70% specificity, demonstrating the intended emphasis of O-RADS on avoiding missed malignancy. GI-RADS provided a slightly different diagnostic balance. In the present analysis, GI-RADS achieved sensitivity of 92.5%, specificity of 92.9%, NPV of 97.0%, accuracy of 92.7% and AUC of 0.953. Amor et al., in the original GI-RADS proposal involving 187 masses, reported sensitivity of 92%, specificity of 97%, PPV of 85%, NPV of 99% and accuracy of 96%. [20]

In a subsequent prospective multicentre investigation of 432 tumours, GI-RADS showed even higher sensitivity of 99.1%, although specificity was lower at 85.9%. [11] The current results therefore demonstrate an intermediate and well-balanced operating profile between those two published experiences. At the predefined $\geq 10\%$ threshold, the ADNEX model classified 94.3% of malignant/borderline lesions as high risk and achieved sensitivity of 94.3%, specificity of 88.6%, NPV of 97.6% and AUC of 0.944. The original prospective multicentre study by Van Calster et al. developed ADNEX using age, CA-125, centre type and six ultrasound predictors and reported a validation AUC of 0.943, [10] almost identical to the AUC of 0.944 observed in the present study. At the same 10% malignancy threshold, however, their sensitivity was 96.5% and specificity 71.3%, suggesting that the present cohort achieved greater specificity with only a small reduction in sensitivity. IOTA Simple Rules classified 83.0% of malignant/borderline masses as malignant and 87.1% of benign masses as benign, while 20 of 193 lesions (10.4%) were inconclusive. Among the 173 conclusive lesions, sensitivity was 93.6%, specificity 96.8% and accuracy 96.0%. These values closely reproduce the characteristic high-specificity performance of the IOTA approach. In the prospective IOTA validation involving 1,938 women, the rules were conclusive in 77% of masses and yielded sensitivity of 92% and specificity of 96%. [21] A South Indian study involving 205 tumours found the rules applicable in 89.3% and reported sensitivity of 92.8% and specificity of 92.9%. [22]

Although O-RADS produced the numerically highest AUC (0.970), followed by GI-RADS (0.953), IOTA SR (0.947) and ADNEX (0.944), the absence of a significant difference in overall correctness among the four systems (Cochran's $Q=2.12$; $p=0.547$) indicates that all provided broadly comparable classification performance within this cohort. Similar convergence has been observed externally: one comparative study of 445 adnexal masses reported AUCs of 0.941 for O-RADS and 0.925 for ADNEX, with sensitivities of

93.3% and 94.4% and specificities of 90.2% and 90.6%, respectively. [23] Thus, the principal distinction among these systems appears less related to whether they discriminate malignancy effectively and more to how they balance sensitivity, specificity, probability-based risk estimation, and management-oriented categorization. The present study had several limitations. Being a single-centre, hospital-based study, the findings may not be fully generalizable to populations with different referral patterns, disease prevalence, operator expertise, or healthcare settings. The sample size was relatively modest, particularly for individual malignant and borderline histopathological subtypes, which limited the precision of subgroup-specific diagnostic estimates.

Ultrasound interpretation is inherently operator dependent, and despite the use of standardized reporting criteria, interobserver variability could have influenced lesion characterization and subsequent assignment of IOTA SR, ADNEX, GI-RADS, and O-RADS categories. The study also included only adnexal masses with a definitive histopathological diagnosis, which may have introduced verification or selection bias by excluding lesions managed conservatively without surgery. In addition, the four systems differ in their intended design and output – feature-based classification, probabilistic modelling, and categorical risk stratification – which may limit direct equivalence when comparing diagnostic performance using a single dichotomous benign-versus-malignant reference outcome.

Conclusion

The present study demonstrated that IOTA Simple Rules, ADNEX, GI-RADS, and O-RADS were all effective ultrasound-based systems for differentiating benign from malignant or borderline adnexal masses. O-RADS provided the highest sensitivity and negative predictive value, supporting its usefulness for minimizing missed malignancies, while GI-RADS showed a favourable balance of sensitivity, specificity, and overall accuracy.

The ADNEX model also demonstrated excellent discrimination and offered the additional advantage of quantitative malignancy-risk estimation, whereas IOTA Simple Rules achieved particularly high specificity and accuracy among lesions with conclusive findings. Overall, the comparable diagnostic performance of all four systems suggests that standardized ultrasound-based risk stratification can substantially improve the preoperative characterization of adnexal masses and support appropriate referral, counselling, and management decisions.

References

1. Timmerman D, Planchamp F, Bourne T, Landolfo C, du Bois A, Chiva L, et al. ESGO/ISUOG/IOTA/ESGE Consensus Statement on pre-operative diagnosis of ovarian tumors. *Int J Gynecol Cancer*. 2021;31(7):961-82.
2. Vernooij F, Heintz P, Witteveen E, van der Graaf Y. The outcomes of ovarian cancer treatment are better when provided by gynecologic oncologists and in specialized hospitals: a systematic review. *Gynecol Oncol*. 2007;105(3):801-12.
3. Hirsch M, Duffy J, Davis CJ, Nieves Plana M, Khan KS. Diagnostic accuracy of cancer antigen 125 for endometriosis: a systematic review and meta-analysis. *Bjog*. 2016;123(11):1761-8.
4. Van Holsbeke C, Van Calster B, Bourne T, Ajossa S, Testa AC, Guerriero S, et al. External validation of diagnostic models to estimate the risk of malignancy in adnexal masses. *Clin Cancer Res*. 2012;18(3):815-25.
5. Alcázar JL, Pascual MA, Graupera B, Aubá M, Errasti T, Olartecoechea B, et al. External validation of IOTA simple descriptors and simple rules for classifying adnexal masses. *Ultrasound Obstet Gynecol*. 2016;48(3):397-402.
6. Sayasneh A, Ferrara L, De Cock B, Saso S, Al-Memar M, Johnson S, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model: a multicentre external validation study. *Br J Cancer*. 2016;115(5):542-8.
7. Guo W, Zou X, Xu H, Zhang T, Zhao Y, Gao L, et al. The diagnostic performance of the Gynecologic Imaging Reporting and Data System (GI-RADS) in adnexal masses. *Ann Transl Med*. 2021;9(5):398.
8. Andreotti RF, Timmerman D, Strachowski LM, Froyman W, Benacerraf BR, Bennett GL, 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.
9. Basha MAA, Metwally MI, Gamil SA, Khater HM, Aly SA, El Sammak AA, 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.
10. Van Calster B, Van Hoorde K, Valentin L, Testa AC, Fischerova D, Van Holsbeke C, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. *Bmj*. 2014;349:g5920.
11. Amor F, Alcázar JL, Vaccaro H, León M, Iturra A. GI-RADS reporting system for ultrasound evaluation of adnexal masses in clinical practice: a prospective multicenter study. *Ultrasound Obstet Gynecol*. 2011;38(4):450-5.
12. Strachowski LM, Jha P, Phillips CH, Blanchette Porter MM, Froyman W, Glanc P, et al. O-RADS US v2022: An Update from the American College of Radiology's Ovarian-Adnexal Reporting and Data System US Committee. *Radiology*. 2023;308(3):e230685.
13. Malkasian GD, Jr., Knapp RC, Lavin PT, Zurawski VR, Jr., Podratz KC, Stanhope CR, et al. Preoperative evaluation of serum CA 125 levels in premenopausal and postmenopausal patients with pelvic masses: discrimination of benign from malignant disease. *Am J Obstet Gynecol*. 1988;159(2):341-6.
14. Chopra S, Vaishya R, Kaur J. An Evaluation of the Applicability of the Risk of Malignancy Index for Adnexal Masses to Patients Seen at a Tertiary Hospital in Chandigarh, India. *J Obstet Gynaecol India*. 2015;65(6):405-10.
15. Spagnol G, Marchetti M, Carollo M, Bigardi S, Tripepi M, Facchetti E, et al. Clinical Utility and Diagnostic Accuracy of ROMA, RMI, ADNEX, HE4, and CA125 in the Prediction of Malignancy in Adnexal Masses. *Cancers (Basel)*. 2024;16(22).
16. Timmerman D, Testa AC, Bourne T, Ameye L, Jurkovic D, Van Holsbeke C, et al. Simple ultrasound-based rules for the diagnosis of ovarian cancer. *Ultrasound Obstet Gynecol*. 2008;31(6):681-90.
17. Wu C, Wu L, Shang J, Xie H, Peng R. Diagnostic Performance of O-RADS US (Version 2019 and Version 2022) Incorporating Acoustic Shadowing by Junior Radiologists: Analyzing 1061 Adnexal Masses. *J Ultrasound Med*. 2025;44(5):845-55.
18. Perez M, Meseguer A, Vara J, Vilches JC, Brunel I, Lozano M, et al. GI-RADS versus O-RADS in the differential diagnosis of adnexal masses: a systematic review and head-to-head meta-analysis. *Ultrasonography*. 2024;43(6):438-47.
19. Hack K, Gandhi N, Bouchard-Fortier G, Chawla TP, Ferguson SE, Li S, et al. External Validation of O-RADS US Risk Stratification and Management System. *Radiology*. 2022;304(1):114-20.
20. Amor F, Vaccaro H, Alcázar JL, León M, Craig JM, Martinez J. Gynecologic imaging reporting and data system: a new proposal for classifying adnexal masses on the basis of sonographic findings. *J Ultrasound Med*. 2009;28(3):285-91.

21. Timmerman D, Ameye L, Fischerova D, Epstein E, Melis GB, Guerriero S, et al. Simple ultrasound rules to distinguish between benign and malignant adnexal masses before surgery: prospective validation by IOTA group. *Bmj*. 2010;341:c6839.
22. Shetty J, Saradha A, Pandey D, Bhat R, Pratap K, Bharatnur S. IOTA Simple Ultrasound Rules for Triage of Adnexal Mass: Experience from South India. *J Obstet Gynaecol India*. 2019;69(4):356-62.
23. Wang R, Yang Z. Evaluating the risk of malignancy in adnexal masses: validation of O-RADS and comparison with ADNEX model, SA, and RMI. *Ginekol Pol*. 2023;94(10):799-806.