Cureus
Part of SPRINGER NATURE

Open Access Original Article

Diagnostic Performance of CA-125, the Risk of Malignancy Index (RMI), IOTA Simple Rules, the ADNEX Model, and the Ovarian-Adnexal Reporting and Data System (O-RADS) in the Preoperative Evaluation of Adnexal Masses: A Prospective Cohort Study in Syria

Rafif Hadakie [1], Muhammad Nazir Yasmineh [1]

1. Obstetrics and Gynecology, University Hospital of Obstetrics and Gynecology, Damascus, SYR

Corresponding author: Rafif Hadakie, rafif.hadaki@gmail.com

Review began 11/10/2026
Review ended 11/24/2026
Published 11/25/2026

© **Copyright** 2026
Hadakie et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.120251

Abstract

Background: Preoperative discrimination between benign and malignant adnexal masses is essential to optimize management and referral. No single test is universally adopted. This study evaluated the diagnostic accuracy of cancer antigen 125 (CA-125), the Risk of Malignancy Index 1 (RMI 1), the International Ovarian Tumor Analysis (IOTA) Simple Rules (SR), IOTA ADNEX, and the Ovarian-Adnexal Reporting and Data System (O-RADS) in Syrian women.

Methodology: This prospective study included 75 patients scheduled for surgical excision of adnexal masses at the University Hospital of Obstetrics and Gynecology in Damascus. Patients underwent preoperative evaluation using serum CA-125 levels, RMI 1, IOTA SR, IOTA ADNEX, and O-RADS. Histopathology served as the reference standard. Diagnostic metrics were evaluated, and pairwise receiver operating characteristic (ROC) curve comparisons were performed using DeLong's test with Holm-Bonferroni adjustment.

Results: Histopathology confirmed 51 (68.0%) benign and 24 (32.0%) malignant lesions (mean age, 44.17 ± 13.89 years). CA-125 ⩾35 U/mL yielded a sensitivity of 58.3% and a specificity of 76.5%; at ⩾100 U/mL, sensitivity was 45.8% and specificity was 96.1%. RMI 1 ⩾25 yielded a sensitivity of 91.7% and a specificity of 60.8%, whereas RMI 1 ⩾200 yielded a sensitivity of 54.2% and a specificity of 92.2%. IOTA Simple Rules among classifiable cases (n = 62) yielded a sensitivity of 100.0% and a specificity of 90.0%; 13/75 cases (17.3%) were inconclusive. O-RADS ⩾4 yielded a sensitivity of 100.0%, a specificity of 72.5%, and an accuracy of 81.3%. At the prespecified ADNEX risk threshold of ⩾10%, sensitivity was 100.0%, specificity was 70.6%, and accuracy was 80.0%. In the primary paired DeLong analysis of 62 classifiable cases, no pairwise area under the curve (AUC) difference remained significant after Holm-Bonferroni adjustment (all adjusted p ⩾ 0.072). In the sensitivity analysis of all 75 cases, ADNEX had a higher AUC than CA-125, RMI 1, and binary IOTA Simple Rules, but not O-RADS.

Conclusions: All evaluated methods showed useful discriminatory performance, but the primary paired analysis did not demonstrate statistically significant AUC differences after correction. The handling of inconclusive IOTA Simple Rules results affected secondary comparisons; these findings require confirmation in larger external cohorts.

Categories: Obstetrics/Gynecology
Keywords: adnex, adnexal mass, cancer antigen 125, diagnostic accuracy, iota, o-rads, ovarian cancer, risk of malignancy index, rmi, ultrasound

Introduction

Ovarian cancer is a major gynecological malignancy and remains a leading cause of cancer-related mortality among women worldwide. Globally, approximately 324,603 new cases and 206,956 deaths from ovarian cancer were estimated in 2022, reflecting its substantial global disease burden [1]. Although it ranks as the third most common gynecological cancer globally, behind cervical and uterine corpus cancers, it accounts for more deaths than any other malignancy of the female reproductive system due to late-stage presentation and the lack of effective screening strategies [2]. Historically labeled a "silent killer," ovarian cancer is associated with nonspecific abdominal or pelvic symptoms in most patients, making early diagnosis challenging. Consequently, a substantial proportion of patients are diagnosed at an advanced stage, at which surgical management is complex and the overall prognosis remains poor [3].

Adnexal masses represent a frequent diagnostic challenge in routine gynecological practice. The principal clinical goal is to distinguish benign lesions from malignant tumors before surgical intervention, as this

differentiation dictates the surgical approach, subspecialty referral to gynecologic oncologists, and perioperative planning. Preoperative assessment relies on the integration of clinical evaluation, serum biomarkers, and diagnostic imaging features [4].

Serum cancer antigen 125 (CA-125) is the most widely used biomarker for evaluating suspected ovarian malignancy. However, its diagnostic specificity is limited because its levels may be elevated in various benign gynecological conditions (e.g., endometriosis and leiomyomas) and non-gynecological inflammatory processes [5]. To improve diagnostic accuracy, the Risk of Malignancy Index 1 (RMI 1) integrates serum CA-125 levels, ultrasound (US) features, and menopausal status to estimate the probability of malignancy, serving as a traditional triage index [6].

Although advanced imaging modalities such as magnetic resonance imaging (MRI) and computed tomography (CT) play an adjunctive role in staging and complex cases, pelvic ultrasonography remains the primary first-line method for evaluating adnexal pathology [7]. Recent advances in gynecological US have led to standardized diagnostic algorithms that provide structured risk stratification. Prominent among these are the International Ovarian Tumor Analysis (IOTA) Simple Rules (IOTA SR), the American College of Radiology Ovarian-Adnexal Reporting and Data System (ACR O-RADS US), and the IOTA ADNEX model.

The IOTA SR utilize 10 standardized US features (five benign features (B-rules) and five malignant features (M-rules)) to classify masses as benign, malignant, or inconclusive (when both or neither rule applies) [8]. The IOTA ADNEX model provides an individualized risk calculation that distinguishes among benign, borderline, stage I, stage II-IV, and metastatic malignancies using three clinical and six US variables [9]. Meanwhile, the ACR O-RADS US system stratifies adnexal lesions into five risk categories, ranging from O-RADS 1 (physiologic) to O-RADS 5 (high risk, ≥50%), to guide standardized clinical management [10].

Although several systematic reviews have examined these diagnostic tools independently, prospective head-to-head comparisons evaluating all five approaches within the same patient cohort remain limited [11-14]. Furthermore, in resource-limited healthcare environments such as Syria, identifying accurate, noninvasive, and cost-effective diagnostic tools is a critical clinical priority [15]. Evaluating whether structured US assessment, either alone or alongside accessible markers such as CA-125, provides sufficient diagnostic efficacy is of considerable practical value for clinical decision-making.

Therefore, the primary objective of this study was to perform a prospective head-to-head comparison within the same Syrian patient cohort to assess the diagnostic performance of serum CA-125, RMI 1, IOTA SR, ACR O-RADS US, and the IOTA ADNEX model for the preoperative discrimination of benign and malignant adnexal masses, using histopathological examination as the reference standard.

The secondary objectives of this study were to (1) estimate sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), overall accuracy, and area under the receiver operating characteristic curve (AUC) for each diagnostic strategy at prespecified thresholds; (2) perform pairwise comparisons of AUCs across the five diagnostic strategies; (3) evaluate the impact of alternative cutoffs for CA-125, RMI 1, O-RADS US, and ADNEX on diagnostic performance; (4) conduct a prespecified sensitivity analysis to assess how reclassifying inconclusive IOTA SR results as high risk affects diagnostic performance; and (5) explore cohort-specific cutoffs for CA-125, RMI 1, and the IOTA ADNEX model using Youden's index and receiver operating characteristic (ROC)-based methods, explicitly treating these analyses as exploratory and subject to external validation.

Materials And Methods

Study design and patient population

This prospective observational cohort study was conducted at the University Hospital of Obstetrics and Gynecology, Faculty of Medicine, Damascus University, from May 2024 to June 2026. The study included 75 consecutive female patients presenting with an adnexal mass who underwent preoperative ultrasonographic assessment. All preoperative US examinations were performed by a specialist in obstetrics and gynecology trained in IOTA and O-RADS criteria, under the direct supervision of a senior consultant gynecologist, using a Philips US system (Philips Healthcare, Andover, MA) equipped with transvaginal (5-9 MHz) and transabdominal (3.5-5 MHz) transducers. A standardized scanning protocol was followed: transvaginal sonography was performed routinely, whereas transabdominal sonography was used for large pelvic-abdominal masses or when transvaginal imaging was not feasible (e.g., in virginal patients). Because this was a prospective preoperative study, US examiners were blinded to the final histopathological results and to serum CA-125 and RMI 1 values at the time of scanning. All patients subsequently underwent surgical excision through laparoscopy or laparotomy, followed by definitive histopathological examination.

Eligible participants were women with a sonographically detected adnexal mass for which surgical intervention was indicated and for whom complete postoperative histopathological data were available. Pregnant patients, those with incomplete medical records, and those without definitive histopathological verification following surgery were excluded. A detailed medical history was obtained at enrollment, followed by a comprehensive general and pelvic clinical examination.

Written informed consent was obtained from all participants before enrollment. Additionally, the study was conducted in accordance with the principles of the Declaration of Helsinki.

Clinical evaluation and serum CA-125

Preoperative venous blood samples were collected before surgery. Serum CA-125 levels were measured using a chemiluminescent immunoassay (CLIA) on a Mindray automated analyzer (Mindray Bio-Medical Electronics Co., Ltd., Shenzhen, China), following the manufacturer's standard calibration and quality-control protocols, with a reference cutoff of 35 U/mL. A serum CA-125 level of ⩾35 U/mL was used as the primary positive threshold for suspected malignancy, consistent with conventional clinical practice. Additional exploratory thresholds of ⩾100 U/mL and ⩾200 U/mL were evaluated to assess the trade-off between sensitivity and specificity.

RMI 1

RMI 1 was calculated for each patient using the standard formula:

$$\mathrm{RMI}_1 = M \times U \times \mathrm{CA\text{-}125}$$

In this formula, M represents menopausal status, U represents the ultrasonographic score, and CA-125 represents the serum CA-125 level expressed in U/mL. The menopausal status score was assigned as 1 for premenopausal women and 3 for postmenopausal women.

The ultrasonographic score was determined according to five features: multilocular cysts, solid areas, intra-abdominal metastases, ascites, and bilateral lesions. A score of $U = 0$ was assigned when none of these features was present, $U = 1$ when one feature was present, and $U = 3$ when two or more features were present [16].

An RMI 1 score of ⩾200 was designated as the primary positive threshold for suspected malignancy, in accordance with conventional clinical guidelines. An additional exploratory threshold of ⩾25 was evaluated to investigate its potential to improve sensitivity.

Ultrasonographic evaluation and diagnostic models

All participants underwent standardized preoperative pelvic ultrasonography using high-resolution US equipment with transvaginal transducers operating at 5-9 MHz and/or transabdominal transducers operating at 2-5 MHz. Each adnexal mass was prospectively assessed using three structured US-based risk assessment approaches: IOTA SR, ACR O-RADS US, and the IOTA ADNEX model.

IOTA SR

Adnexal lesions were evaluated according to the 10 standardized IOTA SR features, comprising five benign (B) features and five malignant (M) features. The benign features were a unilocular cyst (B1), the presence of solid components with a largest diameter of less than 7 mm (B2), the presence of acoustic shadows (B3), a smooth multilocular tumor with a maximum diameter of less than 100 mm (B4), and the absence of detectable color Doppler flow (color score 1; B5).

The malignant features were an irregular solid tumor (M1), the presence of ascites (M2), at least four papillary projections (M3), an irregular multilocular-solid tumor with a maximum diameter of at least 100 mm (M4), and very strong color Doppler flow (color score 4; M5).

Lesions were classified as benign when only benign features were present and as malignant when only malignant features were present. Lesions were considered inconclusive when both benign and malignant features were present or when neither set of features was identified [8].

For the primary binary analysis, inconclusive cases were excluded, resulting in an evaluable cohort of 62 patients. In a prespecified sensitivity analysis, all 13 inconclusive cases were classified as positive (high risk) to assess a conservative clinical triage strategy intended to minimize missed malignant lesions.

ACR O-RADS US System

Adnexal masses were classified according to the ACR O-RADS US system. Category 1 represented normal ovarian morphology or a physiological finding. Category 2 indicated an almost certainly benign lesion, with a malignancy risk of less than 1%. Category 3 indicated a low risk of malignancy, ranging from 1% to less than 10%. Category 4 indicated an intermediate risk of malignancy, ranging from 10% to less than 50%, whereas Category 5 indicated a high risk of malignancy, defined as a risk of at least 50% [10].

For binary diagnostic performance analyses, three positivity thresholds were evaluated: O-RADS ⩾3, O-

RADS ⩾4, and O-RADS ⩾5.

IOTA ADNEX Model

The IOTA ADNEX model was prospectively applied to each lesion. The model incorporates nine variables, including three clinical or laboratory variables (patient age, serum CA-125 level, and center type (oncology referral center vs. non-oncology center)) and six ultrasonographic variables: maximum lesion diameter, maximum diameter of the solid component, presence of more than 10 cyst locules, number of papillary projections (0, 1, 2, 3, or more than 3), presence or absence of acoustic shadows, and presence or absence of ascites [9].

A predicted malignancy risk of ⩾10% was prespecified as the primary cutoff for classifying a lesion as positive (high risk). In addition, ROC curve analysis was performed to evaluate the model's continuous discriminatory performance across the full range of predicted malignancy probabilities.

Predicted malignancy probabilities from the IOTA ADNEX model were calculated using the official online IOTA ADNEX calculator, selecting "oncology center" for the center-type parameter to reflect our tertiary referral hospital setting.

Histopathological examination

Histopathological examination of the surgically resected specimens served as the definitive reference standard. The specimens were routinely processed and independently evaluated by two experienced pathologists who were blinded to the preoperative US findings, serum CA-125 levels, and model risk scores. Malignant ovarian tumors were staged according to the International Federation of Gynecology and Obstetrics (FIGO) staging system [17]. For binary diagnostic accuracy analyses, borderline ovarian tumors were classified within the malignant group.

Statistical analysis

The normality of continuous variables was assessed using the one-sample Kolmogorov-Smirnov test. Normally distributed variables were summarized as means and standard deviations, whereas non-normally distributed variables were expressed as medians and interquartile ranges. Categorical variables were reported as frequencies and percentages. Between-group comparisons of continuous variables were performed using the independent-samples t-test or the Mann-Whitney U test, as appropriate for the distribution of the data. Categorical variables were compared using the chi-square test or Fisher's exact test, as appropriate. There were no missing data for any index test or reference standard variable among the 75 enrolled patients.

For each diagnostic modality and prespecified cutoff, 2 × 2 contingency tables were constructed using histopathological findings as the reference standard. Sensitivity, specificity, PPV, NPV, and overall diagnostic accuracy were calculated. Continuous discriminatory performance was assessed using ROC curves and the area under the curve (AUC), and 95% confidence intervals (95% CIs) were calculated for all diagnostic performance metrics. For ACR O-RADS US, the ROC curve and AUC were generated across the ordered O-RADS categories (Categories 3, 4, and 5), whereas binary diagnostic performance metrics were evaluated at prespecified cutoffs (primarily O-RADS ⩾4 as positive vs. <4 as negative, alongside secondary thresholds of ⩾3 and ⩾5). Cohort-specific optimal cutoffs for CA-125, RMI 1, and the IOTA ADNEX model were explored using ROC/Youden index analysis and were explicitly treated as exploratory.

Pairwise comparisons of correlated ROC curves were performed using DeLong's test. The primary pairwise analysis included the 62 patients with classifiable IOTA SR results. All five diagnostic methods had complete results for these same patients, allowing direct comparisons within a common study population. In a prespecified sensitivity analysis, all 13 inconclusive IOTA SR cases were classified as positive (high risk), and pairwise comparisons were repeated in the full cohort of 75 patients. All five diagnostic methods had complete results for the same 75 patients in this analysis.

Within each analysis, P values for the 10 pairwise comparisons were adjusted using the Holm-Bonferroni method. All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant. Statistical analyses were performed using IBM SPSS Statistics for Windows, version 27 (IBM Corp., Armonk, NY).

Data availability

The de-identified dataset supporting the findings of this study is available from the corresponding author upon reasonable request, subject to institutional approval.

Results

Study population characteristics

A total of 75 women presenting with adnexal masses who underwent surgical management and had definitive histopathological diagnoses were included in the study. Histopathological examination identified 51 (68.0%) benign lesions and 24 (32.0%) malignant lesions.

The mean age of the study population was 44.17 ± 13.89 years, with a range of 15-76 years. Patients with malignant lesions were significantly older than those with benign lesions (50.54 ± 13.61 vs. 41.18 ± 13.11 years, respectively; $t(73) = 2.851$, $P = 0.006$). Premenopausal women accounted for 43 (57.3%) participants. Postmenopausal status was more frequent among patients with malignant lesions than among those with benign lesions (13, 54.2% vs. 19, 37.3%, respectively); however, this difference was not statistically significant ($P = 0.213$).

Pelvic or lower abdominal pain was the most frequently reported clinical presentation, occurring in 51 (68.0%) patients, followed by pelvic heaviness in 20 (26.7%). Abdominal enlargement was significantly more frequent among patients with malignant lesions than among those with benign lesions (5, 20.8% vs. 2, 3.9%, respectively; $P = 0.028$). Clinical symptoms were not mutually exclusive. Detailed demographic and clinical characteristics of the study population are presented in Table *1*.

Variable	Benign masses	Malignant	Test statistic	*P*-value
Age (years)	41.18 ± 13.11	50.54 ± 13.61	$t(73) = 2.851$	0.006 ŧ
Menopausal status				
Premenopausal	32 (62.7%)	11 (45.8%)	-	0.213 *ƒ*
Postmenopausal	19 (37.3%)	13 (54.2%)		
Symptoms				
Pelvic/abdominal pain	36 (70.6%)	15 (62.5%)	$\chi^2(3, n = 75) = 35.699$	0.15 €
Pelvic heaviness	13 (25.5%)	7 (29.2%)		
Uterine bleeding	10 (19.6%)	4 (16.7%)		
Abdominal enlargement	2 (3.9%)	5 (20.8%)		

TABLE 1: Study population characteristics.

Note: The symbols indicate the statistical tests used: ŧ represents the t-test, *ƒ* represents Fisher's exact test, and € represents the chi-square test.

Pathological characteristics

The mean maximum lesion diameter in the overall cohort was 118.23 ± 60.44 mm. No statistically significant difference was observed between benign and malignant lesions, with mean diameters of 115.53 ± 60.45 mm and 123.96 ± 61.31 mm, respectively ($U = 552$, $Z = -0.684$, $P = 0.494$). In contrast, the maximum diameter of the solid component was significantly greater in malignant lesions than in benign lesions (59.08 ± 41.79 mm vs. 14.41 ± 23.68 mm, respectively; $U = 146.5$, $Z = -5.426$, $P < 0.001$).

Several sonographic characteristics were significantly associated with malignancy. Ascites was detected in 12 (50%) malignant lesions compared with 4 (7.8%) benign lesions ($P < 0.001$). Marked vascularity, defined as a color Doppler score of 4, was observed in 9 (37.5%) malignant lesions and none of the benign lesions ($P < 0.001$). Papillary projections were identified in 10 (41.7%) malignant lesions compared with 4 (7.8%) benign lesions ($P = 0.001$).

Other sonographic features did not differ significantly between the benign and malignant groups. These included the presence of more than 10 cyst locules ($P = 0.717$), bilateral involvement ($P = 0.455$), and acoustic shadowing ($P = 0.168$).

Risk classification by the evaluated diagnostic systems

According to the IOTA SR, 36 (48%) lesions were classified as low risk, 26 (34.7%) as high risk, and 13 (17.3%) as inconclusive. All 36 (100%) low-risk lesions were confirmed to be benign on histopathological examination. Among the 26 high-risk lesions, 22 (84.6%) were malignant. Of the 13 (17.3%) inconclusive lesions, 11 (84.6%) were benign. For the primary diagnostic performance analysis, the 13 inconclusive cases were excluded, leaving 62 (82.7%) lesions with definitive low- or high-risk classifications. Within this subset, all 22 (100%) malignant lesions were classified as high risk, whereas 4 of the 40 benign lesions (10% of

benign cases, or 6.5% of the 62-lesion subset) were also classified as high risk.

According to the ACR O-RADS US system, 18 (24%) lesions were classified as category 2, 19 (25.3%) as category 3, 24 (32%) as category 4, and 14 (18.7%) as category 5. All lesions classified as categories 2 (18/18; 100%) and 3 (19/19; 100%) were benign on histopathological examination. Among the 24 category 4 lesions, 12 (50%) were malignant. Similarly, 12 of the 14 (85.7%) category 5 lesions were malignant.

The IOTA ADNEX model classified 51 (68%) lesions as benign, 4 (5.3%) as borderline tumors, 3 (4%) as stage I invasive malignancies, and 17 (22.7%) as stage II-IV invasive malignancies (totaling 24, 32%, malignant cases at the ⩾10% cutoff). A predicted malignancy risk of ⩾10% was prespecified as the primary cutoff for binary classification, with a predicted risk of less than 10% considered negative. As a secondary exploratory analysis, ROC curve analysis was performed to identify a cohort-specific threshold that optimized discrimination between benign and malignant lesions.

Regarding RMI 1, 33 (44%) lesions had scores below 25, whereas 17 (23%) had scores above 200. Serum CA-125 levels were below 35 U/mL in approximately 49 (65%) patients. The distribution of risk classifications across all evaluated diagnostic systems according to histopathological outcome is presented in Table 2.

Diagnostic system	Total cohort, *n* (%)	Benign (*n* = 51), *n* (%)	Malignant (*n* = 24), *n* (%)
IOTA Simple Rules			
Low-risk (class 1)	36 (48%)	36 (70.6%)	0 (0%)
High-risk (class 2)	26 (34.7%)	4 (7.8%)	22 (84.6%)
Inconclusive (class 3)	13 (17.3%)	11 (21.6%)	2 (8.3%)
O-RADS classification			
Category 2	18 (24%)	18 (35.3%)	0 (0%)
Category 3	19 (25.3%)	19 (37.3%)	0 (0%)
Category 4	24 (32%)	12 (23.5%)	12 (50%)
Category 5	14 (18.7%)	2 (3.9%)	12 (50%)
IOTA ADNEX model			
Benign	51 (68%)	49 (96.1%)	2 (8.3%)
Borderline tumor	4 (5.3%)	0 (0%)	4 (16.7%)
Stage I Invasive	3 (4%)	0 (0%)	3 (12.5%)
Stage II-IV Invasive	17 (22.7%)	2 (3.9%)	15 (62.5%)
RMI 1			
<25	33 (44%)	31 (60.8%)	2 (8.3%)
25-200	25 (33.3%)	16 (31.4%)	9 (37.5%)
>200	17 (22.7%)	4 (7.8%)	13 (54.2%)
Serum CA-125 (U/mL)			
<35	49 (65.3%)	39 (76.5%)	10 (41.7%)
35-200	16 (21.3%)	11 (21.6%)	5 (20.8%)
>200	10 (13.3%)	1 (2%)	9 (37.5%)

TABLE 2: Distribution of risk classifications across all evaluation systems according to histopathological outcomes.

IOTA, International Ovarian Tumor Analysis; O-RADS, Ovarian-Adnexal Reporting and Data System; CA-125, cancer antigen 125; RMI 1, Risk of Malignancy Index 1

Diagnostic performance and ROC curve analysis of the studied evaluation system

Serum CA-125

Serum CA-125 demonstrated an AUC of 0.788 (95% CI, 0.676-0.901), indicating discriminatory ability between benign and malignant adnexal lesions. At the conventional threshold of ⩾35 U/mL, sensitivity was 58.33%, specificity was 76.47%, PPV was 53.85%, NPV was 79.59%, and overall diagnostic accuracy was 70.67%.

Increasing the threshold to ⩾100 U/mL reduced sensitivity to 45.83% but increased specificity to 96.08%. At this threshold, PPV was 84.62%, NPV was 79.03%, and diagnostic accuracy was 80.00%. At the higher threshold of ⩾200 U/mL, sensitivity decreased further to 37.50%, whereas specificity increased to 98.04%. The corresponding PPV, NPV, and diagnostic accuracy were 90.00%, 76.92%, and 78.67%, respectively.

A separate analysis identified an optimal threshold of 25.53 U/mL, with a sensitivity of 79.2%, specificity of 62.7%, PPV of 50.0%, and NPV of 86.5%. The diagnostic accuracy reported for this threshold was 68.0%.

Overall, increasing the CA-125 threshold improved specificity and PPV at the expense of sensitivity. The conventional threshold of ⩾35 U/mL was retained as the primary cutoff for comparative diagnostic analysis.

RMI 1

RMI 1 demonstrated an AUC of 0.868 (95% CI, 0.787-0.948; $P < 0.001$), indicating good discriminatory performance between benign and malignant lesions.

At the exploratory threshold of ⩾25, sensitivity was 91.67%, specificity was 60.78%, PPV was 52.38%, NPV was 93.94%, and overall diagnostic accuracy was 70.67%. At the conventional threshold of ⩾200, sensitivity decreased to 54.17%, whereas specificity increased to 92.16%. The corresponding PPV, NPV, and diagnostic accuracy were 76.47%, 81.03%, and 80.00%, respectively.

An additional analysis using the Youden index identified an optimal cutoff of 67.5. At this threshold, sensitivity was 87.5%, specificity was 76.5%, PPV was 63.6%, NPV was 92.9%, and overall diagnostic accuracy was 80.0%.

These findings demonstrate that the selected RMI 1 threshold substantially influenced diagnostic performance. The lower threshold of ⩾25 provided greater sensitivity, whereas the conventional threshold of ⩾200 improved specificity. The Youden index-derived threshold of 67.5 provided an intermediate balance between sensitivity and specificity. The conventional threshold of ⩾200 was retained as the primary cutoff for comparative analysis.

IOTA Simple Rules

In the primary analysis, which excluded the 13 (17.3%) inconclusive cases, the IOTA SR were evaluated in 62 (82.7%) patients. High-risk classification yielded a sensitivity of 100%, specificity of 90%, PPV of 84.62%, NPV of 100%, and overall diagnostic accuracy of 93.55%. All malignant lesions in this analysis were classified as high risk, whereas 90% of benign lesions were correctly classified as low risk.

In the prespecified sensitivity analysis, all 13 (17.3%) inconclusive cases were classified as positive for malignancy, and the analysis was repeated in the full cohort of 75 patients. Sensitivity remained 100%, whereas specificity decreased to 70.59%. The corresponding PPV, NPV, and overall diagnostic accuracy were 61.54%, 100%, and 80.00%, respectively.

Thus, classifying inconclusive cases as positive maintained complete sensitivity but reduced specificity and overall diagnostic accuracy compared with the primary analysis, in which inconclusive cases were excluded.

ACR O-RADS US System

The diagnostic performance of the ACR O-RADS US system varied according to the selected positivity threshold. When category ⩾3 was used as the positive threshold, sensitivity was 100%, specificity was 35.29%, PPV was 42.11%, NPV was 100%, and overall accuracy was 56%.

Using category ⩾4 yielded a sensitivity of 100%, specificity of 72.55%, PPV of 63.16%, NPV of 100%, and overall accuracy of 81.33%. The AUC reported for this threshold was 0.912.

At the more restrictive threshold of category ⩾5, sensitivity decreased to 50%, whereas specificity increased

to 96.08%. The corresponding PPV, NPV, and accuracy were 85.71%, 80.33%, and 81.33%, respectively.

Among the evaluated thresholds, category ⩾4 provided a balance between sensitivity and specificity, whereas category ⩾5 offered greater specificity at the expense of sensitivity. Accordingly, O-RADS ⩾4 was retained as the primary threshold for comparative diagnostic analysis.

IOTA ADNEX Model

Using the prespecified malignancy risk threshold of ⩾10%, the IOTA ADNEX model demonstrated a sensitivity of 100%, specificity of 70.59%, PPV of 61.54%, NPV of 100%, and overall accuracy of 80%.

Continuous ROC curve analysis yielded an AUC of 0.972 (95% CI, 0.925-1.000), indicating strong discriminatory performance between benign and malignant lesions. In an exploratory analysis, the cohort-specific threshold of ⩾49.7% yielded a sensitivity of 100%, specificity of 96.08%, PPV of 92.31%, NPV of 100%, and overall accuracy of 97.33%.

Compared with the prespecified 10% threshold, the exploratory threshold of 49.7% was associated with higher specificity, PPV, and overall accuracy while maintaining 100% sensitivity in this cohort. Nevertheless, because the alternative threshold was derived from the same dataset, its performance should be interpreted as exploratory and requires external validation before general clinical application.

Pairwise comparison of ROC curves and diagnostic accuracy

In the primary paired analysis, restricted to the 62 cases with classifiable IOTA SR results (22 malignant and 40 benign), the AUCs were 0.818 for CA-125, 0.910 for RMI 1, 0.950 for IOTA SR, 0.953 for O-RADS, and 0.966 for ADNEX. None of the 10 pairwise DeLong comparisons remained statistically significant after Holm-Bonferroni adjustment (all adjusted $P \geqslant 0.072$).

In the sensitivity analysis of all 75 cases, with all 13 inconclusive IOTA SR results classified as high risk, the AUCs were 0.788 for CA-125, 0.868 for RMI 1, 0.853 for IOTA SR, 0.912 for O-RADS, and 0.972 for ADNEX. After adjustment, ADNEX had a significantly higher AUC than CA-125 (adjusted $P = 0.009$), RMI 1 (adjusted $P = 0.041$), and the binary IOTA SR classification (adjusted $P = 0.002$), but not O-RADS (adjusted $P = 0.277$).

Discussion

Preoperative discrimination between benign and malignant adnexal masses is essential for appropriate surgical planning and referral to gynecologic oncology. This prospective study compared CA-125, RMI 1, IOTA SR, O-RADS, and the ADNEX model in 75 women in Syria, providing comparative diagnostic data from a setting where resource availability may influence clinical decision-making.

At the prespecified thresholds, CA-125 ⩾35 U/mL yielded a sensitivity of 58.3% and specificity of 76.5%, whereas RMI 1 ⩾200 yielded a sensitivity of 54.2% and specificity of 92.2%. O-RADS ⩾4 and ADNEX ⩾10% both achieved 100% sensitivity, with specificities of 72.5% and 70.6%, respectively. These findings suggest that structured US-based assessment may provide useful diagnostic information beyond CA-125 alone. Similar patterns have been reported in external validation studies evaluating ADNEX and RMI, although diagnostic performance varies across populations and study settings [11,12].

The primary paired DeLong analysis included 62 (82.7%) patients with definitive IOTA SR classifications. The AUCs were 0.818 for CA-125, 0.910 for RMI 1, 0.950 for IOTA SR, 0.953 for O-RADS, and 0.966 for ADNEX. Although ADNEX had the highest numerical AUC, none of the 10 pairwise comparisons remained statistically significant after Holm-Bonferroni correction (all adjusted *P*-values ⩾0.072). Therefore, the primary analysis did not establish statistically significant superiority of one method over another. This finding is consistent with comparative evidence suggesting broadly similar diagnostic performance among established US-based approaches [12,14].

IOTA SR yielded inconclusive classifications in 13 of 75 patients (17.3%). In the sensitivity analysis, these cases were classified as high risk, and all 75 patients (100%) were included. ADNEX had a significantly higher AUC than CA-125 (adjusted $P = 0.009$), RMI 1 (adjusted $P = 0.041$), and the binary IOTA SR classification (adjusted $P = 0.002$), but not O-RADS (adjusted $P = 0.277$). These secondary findings should be interpreted cautiously because reclassifying inconclusive IOTA results changes the classification strategy, and the binary IOTA classification differs from the continuous or ordinal scores used for the other methods. The inconclusive rate also highlights a practical limitation of IOTA SR, particularly when expert second-line assessment is not readily available [8,18].

At the prespecified ADNEX risk threshold of 10%, sensitivity was 100% and specificity was 70.6%. A post hoc ROC-derived threshold of 49.7% increased specificity to 96.1% and accuracy to 97.3%. However, this threshold was selected and evaluated in the same small cohort and may therefore reflect overfitting. It should be considered exploratory rather than a clinically validated cutoff and requires independent external

validation. Previous validation studies have reported good diagnostic discrimination with ADNEX, although performance varies across populations and clinical settings [9,12].

Several limitations should be acknowledged. The sample size was small, with only 24 (32%) malignant tumors, limiting statistical power and the precision of diagnostic estimates. Restricting enrollment to patients who underwent surgery introduces selection and spectrum bias, as the findings may not generalize to all women presenting with adnexal masses, particularly those managed conservatively. The prevalence of malignancy in this surgical cohort may also affect predictive values and the applicability of the results to other clinical settings. In addition, US-based evaluation is operator-dependent, inter-rater reliability was not formally assessed, and subtle observer effects cannot be entirely excluded. Furthermore, data-derived thresholds, including CA-125 (25.53 U/mL), RMI 1 (67.5), and ADNEX (49.7%), were identified within the same cohort and are exploratory and hypothesis-generating. These thresholds carry a risk of overfitting and require independent validation. Comparisons between ordinal or continuous metrics (such as O-RADS) and binary classifications also warrant caution. Finally, the study was conducted at a single center, reflecting local equipment, clinical pathways, and operator context, which may limit generalizability to other settings.

Conclusions

In this prospective surgical cohort, US-based approaches, including IOTA SR, O-RADS, and the ADNEX model, demonstrated promising diagnostic performance in distinguishing benign from malignant adnexal masses, although statistically significant pairwise AUC differences were not found in the primary analysis after multiplicity adjustment. Because downstream clinical outcomes and incremental-value models were not assessed, these results reflect diagnostic accuracy in a surgical population rather than overall clinical utility. Inconclusive IOTA SR findings underscore the need for suitable second-line evaluation when lesions cannot be classified. Selection of an approach should be guided by clinical context, examiner expertise, and available resources, and larger multicenter studies are warranted to confirm these findings, especially in resource-constrained settings.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Rafif Hadakie, Muhammad Nazir Yasmineh

Acquisition, analysis, or interpretation of data: Rafif Hadakie

Drafting of the manuscript: Rafif Hadakie

Critical review of the manuscript for important intellectual content: Rafif Hadakie, Muhammad Nazir Yasmineh

Supervision: Muhammad Nazir Yasmineh

Disclosures

Human subjects: Informed consent for treatment and open access publication was obtained or waived by all participants in this study. Biomedical Research Ethics Committee (BMREC), Damascus University issued approval MD-210926-882, Session No. 36, September 21, 2026. Informed consent was obtained from all individual participants included in the study before evaluation and management. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors would like to acknowledge the clinical and nursing staff of the University Hospital of Obstetrics and Gynecology in Damascus for their administrative and technical support during data collection. The datasets used and analyzed during the current study are available from the corresponding author, Rafif Hadaki, upon reasonable request. The data are stored as de-identified participant data to ensure patient confidentiality. The authors used ChatGPT (GPT-6, OpenAI, San Francisco, CA) solely to improve English language, grammar, and readability. The tool was not used for data analysis, interpretation of results, or formulation of scientific conclusions. The authors take full responsibility for the content of the manuscript.

References

1. Ferlay J, Ervik M, Lam F, et al.: Global Cancer Observatory: Cancer Today. International Agency for Research on Cancer, Lyon, France; 2024.
2. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024, 74:229-63. 10.3322/caac.21834
3. Ceilley E, Jagsi R, Goldberg S, Kachnic L, Powell S, Taghian A: The management of ductal carcinoma in situ in North America and Europe. Results of a survey. Cancer. 2004, 101:1958-67. 10.1002/cncr.20580
4. American College of Obstetricians and Gynecologists: Practice Bulletin No. 174: Evaluation and management of adnexal masses. Obstet Gynecol. 2016, 128:e210-26. 10.1097/AOG.0000000000001768
5. Meden H, Fattahi-Meibodi A: CA 125 in benign gynecological conditions . Int J Biol Markers. 1998, 13:231-7. 10.1177/172460089801300411
6. Huwidi A, Abobrege A, Assidi M, Buhmeida A, Ermiah E: Diagnostic value of risk of malignancy index in the clinical evaluation of ovarian mass. Mol Clin Oncol. 2022, 17:118. 10.3892/mco.2022.2551
7. Sadowski EA, Rockall A, Thomassin-Naggara I, et al.: Adnexal lesion imaging: past, present, and future. Radiology. 2023, 307:e223281. 10.1148/radiol.223281
8. 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. 10.1136/bmj.c6839
9. Van Calster B, Van Hoorde K, Valentin L, 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. 10.1136/bmj.g5920
10. 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, 294:168-85. 10.1148/radiol.2019191150
11. Barreñada L, Ledger A, Kotlarz A, et al.: Head-to-head comparison of the RMI and ADNEX models to estimate the risk of ovarian malignancy: a systematic review and meta-analysis of external validation studies. BMJ Open. 2025, 15:e104141. 10.1136/bmjopen-2025-104141
12. Almeida G, Bort M, Alcázar JL: Comparison of the diagnostic performance of Ovarian Adnexal Reporting Data System (O-RADS) with IOTA Simple Rules and ADNEX model for classifying adnexal masses: a head-to-head meta-analysis. J Clin Ultrasound. 2025, 53:1574-83. 10.1002/jcu.24048
13. Suryawanshi SV, Dwidmuthe KS, Savalkar S, Bhalerao A: Diagnostic efficacy of ultrasound-based International Ovarian Tumor Analysis Simple Rules and Assessment of the Different Neoplasias in the Adnexa model in malignancy prediction among women with adnexal masses: a systematic review. Cureus. 2024, 16:e67365. 10.7759/cureus.67365
14. Han J, Wen J, Hu W: Comparison of O-RADS with the ADNEX model and IOTA SR for risk stratification of adnexal lesions: a systematic review and meta-analysis. Front Oncol. 2024, 14:1354837. 10.3389/fonc.2024.1354837
15. Mariani G, Kasznia-Brown J, Paez D, et al.: Improving women's health in low-income and middle-income countries. Part II: the needs of diagnostic imaging. Nucl Med Commun. 2017, 38:1024-8. 10.1097/MNM.0000000000000752
16. Spagnol G, Marchetti M, Carollo M, 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, 16:10.3390/cancers16223790
17. Prat J: Staging classification for cancer of the ovary, fallopian tube, and peritoneum . Int J Gynecol Obstet. 2014, 124:1-5. 10.1016/j.ijgo.2013.10.001
18. Timmerman D, Planchamp F, Bourne T, et al.: ESGO/ISUOG/IOTA/ESGE consensus statement on preoperative diagnosis of ovarian tumors. Ultrasound Obstet Gynecol. 2021, 58:148-68. 10.1002/uog.23635