

IOTA Logistic回归模型2对良性与非良性附件肿物的鉴别诊断价值

贾寒冰¹, 梁晓彦¹, 周英杰¹, 王紫¹, 周晓静², 田彩^{1*}

(1.河北省生殖健康医院功能科,河北石家庄 050000;2.河北医科大学第二医院妇七科,河北石家庄 050000)

[摘要] 目的 评价 IOTA Logistic 回归模型 2 (IOTA Logistic regression model 2, IOTA·LR2) 在附件肿物良性与非良性(恶性+交界性)鉴别诊断中的效能,并比较其在绝经前与绝经后人群中的表现。方法 回顾性纳入河北医科大学第二医院 2019 年 7 月—2020 年 6 月经手术及病理确诊的卵巢/附件肿物患者 198 例(其中非良性 66 例),收集术前 IOTA·LR2 输出的风险值(LR2%),以术后病理诊断为金标准。以 LR2%≥10% 判定为非良性,计算敏感度、特异度、阳性预测值(positive predictive value, PPV)、阴性预测值(negative predictive value, NPV)及似然比,并绘制受试者工作特征(receiver operating characteristic, ROC)曲线评价判别能力;分别在绝经前与绝经后亚组进行分析。结果 IOTA·LR2 的敏感度为 90.91%,特异度为 85.61%,曲线下面积(area under the curve, AUC)为 0.917(95%CI:0.875~0.960)。Kappa 值为 0.729(P<0.001),提示与病理诊断具有较强一致性。亚组分析显示,绝经前患者敏感度 85.00%、特异度 89.42%(Kappa=0.641; AUC=0.893, 95%CI:0.799~0.986);绝经后患者敏感度 93.48%、特异度 71.43%(Kappa=0.673; AUC=0.847, 95%CI:0.745~0.948)。总体 PPV 为 75.95%,NPV 为 94.96%。结论 IOTA·LR2 诊断模型对附件肿物良性与非良性的鉴别诊断具有较高诊断价值,敏感度较高、漏诊风险较低;但在绝经后人群中特异度相对下降,临床上对“阳性(≥10%)”结果应结合肿瘤标志物及其他影像学征象综合判断,以减少假阳性导致的过度干预风险。

[关键词] 子宫附件; IOTA·LR2; 诊断效能 doi:10.3969/j.issn.1007-3205.2026.05.004

[中图分类号] R322.65 **[文献标志码]** A **[文章编号]** 1007-3205(2026)05-0519-08

Diagnostic performance of the IOTA Logistic regression model 2 for differentiating benign from non-benign adnexal masses

JIA Han-bing¹, LIANG Xiao-yan¹, ZHOU Ying-jie¹, WANG Zi¹, Zhou Xiao-jing², TIAN Cai^{1*}

(1. Department of Functional Medicine, Hebei Provincial Reproductive Health Hospital, Shijiazhuang 050000, China; 2. Seven Section of Department of Gynaecology, the Second Hospital of Hebei Medical University, Shijiazhuang 050000, China)

[Abstract] **Objective** To evaluate the diagnostic performance of the IOTA Logistic regression model 2 (IOTA-LR2) for differentiating benign from non-benign (malignant + borderline) adnexal masses, and to compare its performance between premenopausal and postmenopausal women. **Methods** This retrospective study included 198 patients with ovarian/adnexal masses confirmed by surgery and pathology at the Second Hospital of Hebei Medical University from July 2019 to June 2020 (including 66 cases of non-benign masses). Preoperative IOTA-LR2 risk estimates (LR2%) were collected, with postoperative pathology as the reference standard. Using a predefined cutoff of LR2% ≥10% to classify a mass as non-benign, we calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and likelihood ratios. Receiver operating characteristic (ROC) analysis was performed to assess discrimination, and subgroup analyses were conducted in premenopausal and postmenopausal women. **Results** IOTA-LR2 achieved a sensitivity of 90.91% and a specificity of 85.61%, with an area under the curve (AUC) of 0.917 (95%CI: 0.875—0.960). The kappa statistic was 0.729 (P<0.001), indicating

[收稿日期] 2026-02-09

[基金项目] 河北省医学科学研究课题计划项目(20170109)

[作者简介] 贾寒冰(1990-),女,河北邢台人,河北省生殖健康医院主治医师,医学硕士,从事超声诊断研究。

*通信作者。E-mail: 275067072@qq.com



strong agreement with the pathological diagnosis. In subgroup analyses, the premenopausal group showed a sensitivity of 85.00% and a specificity of 89.42% (Kappa=0.641; AUC=0.893, 95%CI: 0.799—0.986), whereas the postmenopausal group showed a sensitivity of 93.48% and a specificity of 71.43% (Kappa=0.673; AUC=0.847, 95%CI: 0.745—0.948). The overall PPV was 75.95% and the NPV was 94.96%. **Conclusion** The IOTA-LR2 diagnostic model demonstrates good performance for differentiating benign from non-benign adnexal masses, with high sensitivity and a low risk of missed diagnosis; however, specificity is relatively lower in postmenopausal women. Therefore, in clinical practice, positive ($\geq 10\%$) results should be interpreted in conjunction with serum biomarkers and other imaging findings to reduce unnecessary interventions due to false positives.

[Key words] adnexa uteri; IOTA-LR2; diagnostic performance

卵巢恶性肿瘤在全球女性癌症死因中位居第五^[1-2],其晚期患者5年生存率不足30%^[3],是全球公共卫生的重大负担。尽管在过去几十年中,卵巢恶性肿瘤患者的病死率显著下降,但总体发病率仍呈上升趋势^[4-5]。卵巢恶性肿瘤的早期诊断对患者的预后非常重要^[6],早期卵巢恶性肿瘤(I期)患者5年存活率可达90%以上^[7-9],而晚期卵巢恶性肿瘤(III、IV期)仅13%~42%^[10-12]。因此,术前正确评估附件肿物的良恶性有助于帮助临床医生选择合适的治疗方式^[13],从而减少临床医生延误病情或过度治疗的可能,避免患者承担不必要的费用。成本低和易获取是超声检查的优点^[14-15],它可以精确测量卵巢的大小、形态、血流状况以及血管分布,对于早期识别卵巢病变有很大帮助,但检查医师的经验水平在很大程度上影响了模型的诊断准确性^[16-17]。IOTA工作组是一个专注于用超声检查对附件病变进行评估的国际多中心协作团队^[18-19],为术前能够更准确的评估附件肿物的恶性程度,使临床医师能够制定更适合患者的治疗和手术策略,IOTA开发并验证了多种基于超声的诊断模型,IOTA Logistic回归模型2(IOTA Logistic regression model 2,IOTA·LR2)是IOTA基于大样本前瞻性研究开发的模型^[20],这些诊断模型需要在不同种族人群中进行更广泛的验证与应用。本研究的目的在于利用IOTA·LR2模型对国人的附件肿块进行良恶性判断,并以术后病理结果作为依据对其诊断表现进行验证,研究其在预测卵巢良恶性肿瘤中的诊断效率和可靠性,从而探讨IOTA·LR2在预测附件肿物性质中的临床应用价值。

1 资料与方法

1.1 一般资料 本研究选取2019年7月—2020年6月在河北医科大学第二医院因附件包块住院并接受手术的女性患者198例。年龄为9~79岁,平

均(41.11±17.25)岁,绝经前女性124例,绝经后女性74例。卵巢肿瘤病理分型:恶性肿瘤66例,包括卵巢浆液性囊腺癌38例,输卵管浆液性癌6例,卵巢颗粒细胞瘤5例,卵巢交界性浆液性肿瘤4例,卵巢透明细胞癌3例,卵巢子宫内膜样腺癌3例,卵巢交界性浆一黏液性肿瘤2例,阑尾黏液腺癌累及双侧输卵管2例,卵巢黏液性癌1例,卵巢无性细胞瘤1例,肉瘤1例;良性肿瘤132例,包括卵巢畸胎瘤38例,输卵管系膜囊肿21例,卵巢黄体囊肿伴出血12例,卵巢卵泡膜一纤维瘤组肿瘤12例,输卵管积液12例,卵巢子宫内膜异位囊肿10例,卵巢浆液性囊腺瘤8例,卵巢单纯囊肿5例,卵巢黏液性囊腺瘤4例,输卵管卵巢脓肿4例,卵巢浆液性囊腺纤维瘤2例,卵巢浆一黏液性囊腺瘤2例,卵巢间叶源性肿瘤1例,卵巢浆液性囊腺瘤伴子宫内膜异位囊肿1例。组织病理学结果分析显示浆液性腺癌是最常见的恶性肿瘤类型,而成熟畸胎瘤在良性肿瘤中占比最高。收集患者的完整临床信息,包括年龄、月经状况(≥ 50 岁且曾行子宫切除者视为绝经后)、既往病史、手术资料、术后病理报告及家族史等。

本研究已获得河北医科大学第二医院伦理委员会批准(批准文号:2022-R763),所有研究对象均签署了书面知情同意书。

1.2 纳入标准与排除标准 纳入标准:①术前行彩色多普勒超声检查发现附件肿物;②末次超声检查后 ≤ 120 d于河北医科大学第二医院行手术治疗(为减少肿物形态及血流征象随时间变化可能引入的偏倚);③术后病理结果明确(以病理为金标准);④关键临床与超声资料完整(包括病理诊断、绝经状态、LR2%及模型所需变量等)。排除标准:①生理性囊肿;②妊娠合并附件肿物;③既往附件恶性肿瘤病史;④病理结果不明确;⑤图像质量差或缺乏可识别诊断征象者。

1.3 结局定义与分组 本研究以术后组织病理学

诊断为参考标准,将附件肿物分为2组:良性组(病理报告诊断为良性病变)与非良性组(病理报告提示为恶性或交界性肿瘤),所有病理诊断均由2位高年资病理医师独立复核确认。超声评估于检查过程中完成,检查医师对患者术后病理结果保持未知。绝经状态的判定标准如下:≥50岁且既往行子宫切除术者视为绝经后,其余患者根据临床月经史进行判定。

1.4 超声检查及IOTA·LR2模型评估 患者超声图像存储和采集于河北医科大学第二医院超声工作系统。图像采集使用5~9 MHz腔内超声探头,如果肿块太大而不能用经腔内超声观察,则使用1~5 MHz曲面探头经腹采集超声图像。对子宫、附件及盆腔进行系统评估,详细观察附件肿块的位置,按照超声图像特征进行分析诊断。超声检查由具有7年或7年以上妇科超声检查经验的超声医师进行。

采用IOTA·LR2模型对每个附件肿块进行标准化超声特征评估,为减少诊断偏倚,所有检查均在未知患者最终组织病理学结果的情况下进行。LR2模型包括以下6项参数:①患者年龄(岁);②是否存在腹腔积液;③是否出现带血流的乳头状突起;④实性部分的最大直径(若>50 mm,则记录为50 mm,风险值不随直径增加而上升);⑤囊肿内壁是否不规则;⑥是否存在声影。将上述指标输入LR2软件可计算风险值,反映肿物良恶性可能性(%)。以≥10%作为恶性风险的截断点,<10%判定为良性,≥10%判定为可疑恶性^[18]。追踪随访患者术后病理诊断结果,并与术前LR2评估结果进行对比分析。

评估IOTA·LR2诊断模型在附件肿物良性与非良性(恶性+交界性)鉴别中的诊断效能。主要评价指标包括敏感度、特异度、阳性预测值(positive predictive value, PPV)、阴性预测值(negative predictive value, NPV)、阳性似然比(positive likelihood ratio, +LR)、阴性似然比(negative likelihood ratio, -LR)及总体准确率,均基于2×2列联表计算并报告95%置信区间(95%CI)。其中,PPV与NPV的95%CI采用Clopper-Pearson精确法按二项分布计算:PPV以[真阳性(true positive, TP)+假阳性(false positive, FP)]为参数,NPV以[真阴性(true negative, TN)+假阴性(false negative, FN)]为参数。总体准确率=(TP+TN)/总例数。

1.5 六变量回归分析 以病理结果为金标准,将

“非良性(恶性+交界性)”定义为1、“良性”为0,将IOTA·LR2的6个预测变量(年龄、腹水、带血流的乳头状突起、实性部分最大径、囊壁不规则、声影)同时纳入多因素Logistic回归模型。按绝经状态进行亚组分析并分别报告各组诊断效能指标;结合患病率差异对预测值进行解释。纳入病例关键信息完整(病理诊断、绝经状态、LR2%及六变量),未因缺失值剔除病例。

1.6 阈值与Youden指数 采用IOTA·LR2推荐阈值10%作为非良性判定标准(LR2%≥0%判为非良性)进行二分类诊断效能评价。基于LR2连续风险值(LR2%)绘制受试者工作特征(receiver operating characteristic, ROC)曲线并计算曲线下面积(area under the curve, AUC)。采用Youden指数最大化原则(J=敏感度+特异度-1)确定总体及按绝经状态分层的最佳阈值,并对绝经后人群进行多阈值敏感性分析,以评估阈值调整对敏感度与特异度的权衡。

1.7 样本量与估计精度 本研究为回顾性诊断效能评价研究,样本量主要影响敏感度、特异度及AUC估计的精度。敏感度的分母为非良性病例数($n=66$),特异度的分母为良性病例数($n=132$)。在本研究样本量下,敏感度90.91%的95%CI: 0.814~0.966,特异度85.61%的95%CI: 0.784~0.912,区间宽度处于可接受范围,能够满足诊断效能评价的精度要求;AUC为0.917(95%CI: 0.875~0.960),提示判别能力估计较为稳定。另针对六变量多因素Logistic回归分析,本研究非良性事件数为66、模型包含6个预测变量,满足每变量约11个事件($EPV\approx 11$)的经验要求。单个比例(如敏感度/特异度)的样本量亦可按二项分布近似估算: $n=1.96^2\times P(1-P)/d^2$,其中P为预期比例、d为允许误差(半宽度);本研究现有样本量对应的95%CI半宽度为8%~10%,提示样本量可满足分析需要。

1.8 统计学方法 应用R4.3.2统计软件分析数据。连续型变量经正态性检验后,符合正态分布的计量资料以($\bar{x}\pm s$)表示,比较采用独立样本t检验;不符合正态分布者以[M(QR)]表示,比较采用秩和检验。计数变量以例数(%)表示,比较采用 χ^2 检验或Fisher精确检验。采用Kappa分析模型预测与病理结果的一致性。采用Clopper-Pearson精确法计算PPV与NPV的二项分布情况。绘制受试者工作特征(receiver operating characteristic, ROC)曲线并计算曲线下面积

积 (area under the curve, AUC), 以综合评价模型判别能力。P<0.05为差异有统计学意义。

2 结 果

2.1 不同病理性质卵巢肿瘤患者的一般临床特征比较 在纳入的198例附件肿物患者中, 132例病理诊断为良性, 66例诊断为非良性。与良性组相

比, 非良性组年龄较大, 绝经后比例较高, 差异均有统计学意义 (均P<0.001)。单因素比较结果显示, 非良性组腹水、伴血流的乳头状突起及囊壁不规则的发生比例均高于良性组, 实性部分最大径也高于良性组, 差异均有统计学意义 (均P<0.05); 声影在良性组中更常见, 差异有统计学意义 (P<0.001), 见表1。

表1 良性与非良性附件肿物患者一般资料及IOTA·LR2模型变量比较(按病理诊断分类)

Table 1 Comparison of baseline characteristics and IOTA-LR2 model variables between patients with benign and non-benign adnexal masses(classified by pathological diagnosis)

组别	例数	年龄($\bar{x}\pm s$,岁)	绝经状态(例数,%)		腹水(例数,%)		实性部分最大径 [M(QR),mm]
			绝经前	绝经后	有	无	
良性组	132	35.4±16.4	104(78.8)	28(21.2)	7(5.3)	125(94.7)	0.0(36.0)
非良性组	66	52.5±12.9	20(30.3)	46(69.7)	35(53.0)	31(47.0)	41.5(27.5)
t/ χ^2 /Z值		7.380	44.192		59.971		5.346
P值		<0.001	<0.001		<0.001		<0.001
组别	例数	血流乳头状突起(例数,%)		囊壁不规则(例数,%)		声影(例数,%)	
		有	无	有	无	有	无
良性组	132	1(0.8)	131(99.2)	15(11.4)	117(88.6)	28(21.2)	104(78.8)
非良性组	66	5(7.6)	61(92.4)	38(57.6)	28(42.4)	0(0.0)	66(100.0)
t/ χ^2 /Z值		—		47.935		—	
P值		0.016*		<0.001		<0.001*	

*Fisher精确检验

2.2 IOTA·LR2模型总体诊断效能 本研究整体患者的非良性患病率为33.3% (66/198), 绝经后亚组非良性患病率为62.2% (46/74)。鉴于预测值 (PPV/NPV) 受患病率影响, 上述差异可能对预测性能产生影响。以LR2%≥10%判为非良性, IOTA·LR2模型预测为良性119例、非良性79例。以病理结果为金标准, 模型对卵巢肿瘤良恶性的判别显示出较高诊断效能: 敏感度90.91% (60/66), 特异度85.61% (113/132), PPV 75.95% (60/79), NPV 94.96% (113/119); +LR 6.32, -LR 0.11, 总体准确率87.37% (173/198), 约登指数J=0.765。Kappa值为0.729 (P<0.001), 提示IOTA·LR2模型与病理诊断结果具有良好一致性, 见表2。

2.3 绝经前患者亚组分析 本研究共纳入124例绝经前卵巢肿瘤患者, 其中病理结果为良性104例 (83.9%)、非良性20例 (16.1%), 以LR2%≥10%判为非良性, 模型预测为良性96例、非良性28例。以病理结果为金标准, IOTA·LR2模型对绝经前卵巢肿瘤良恶性判断具有较高诊断效能: 敏感度85.00% (17/20), 特异度89.42% (93/104), PPV 60.71% (17/28), NPV 96.88% (93/96); +LR 8.05, -LR 0.17,

表2 IOTA·LR2与病理诊断对照表

Table 2 Comparison of IOTA-LR2 and pathological diagnosis

临床实际	IOTA·LR2预测		总计
	非良性	良性	
非良性	60	6	66
良性	19	113	132
总计	79	119	198
χ^2 值	107.425		
P值	<0.001		
Kappa值(95%CI)	0.729(0.588~0.871)		
PPV(95%CI)	75.95%(0.650~0.849)		
NPV(95%CI)	94.96%(0.894~0.981)		

总体准确率88.71% (110/124), 约登指数J=0.744。Kappa值为0.641 (P<0.001), 表明IOTA·LR2模型在绝经前人群中具有中等一致性 (表3)。

2.4 绝经后患者亚组分析 本研究纳入74例绝经后卵巢肿瘤患者, 病理诊断结果为良性28例 (37.8%) 和非良性46例 (62.2%)。以LR2%≥10%判为非良性, 模型预测为良性23例、非良性51例。IOTA·LR2模型在绝经后卵巢肿瘤良恶性鉴别中表现出较高诊断效能: 敏感度

表3 绝经前IOTA·LR2与病理诊断对照表

Table 3 Comparison of IOTA-LR2 and pathological diagnosis in premenopausal women

临床实际	IOTA·LR2预测		总计
	非良性	良性	
非良性	17	3	20
良性	11	93	104
总计	28	96	124
χ^2 值	53.146		
P值	<0.001		
Kappa值(95%CI)	0.641(0.416~0.865)		
PPV(95%CI)	60.71%(0.406~0.785)		
NPV(95%CI)	96.88%(0.911~0.994)		

93.48% (43/46), 特异度 71.43% (20/28), PPV 84.31% (43/51), NPV 86.96% (20/23); +LR 3.27, -LR 0.09, 总体准确率 84.46% (63/74), 约登指数 J=0.649。Kappa 值为 0.673 ($P<0.001$), 提示 IOTA·LR2 模型预测与病理诊断具有中等一致性, 见表 4。

表4 绝经后IOTA·LR2与病理诊断对照表

Table 4 Comparison of IOTA-LR2 and pathological diagnosis in postmenopausal women

临床实际	IOTA·LR2预测		总计
	非良性	良性	
非良性	43	3	46
良性	8	20	28
总计	51	23	74
χ^2 值	34.232		
P值	<0.001		
Kappa值(95%CI)	0.673(0.441~0.904)		
PPV(95%CI)	84.31%(0.714~0.930)		
NPV(95%CI)	86.96%(0.664~0.972)		

2.5 ROC曲线分析及阈值探索 基于IOTA·LR2模型输出的连续风险值(LR2%)绘制ROC曲线并计算AUC(DeLong法), 总体ROC分析显示AUC为0.917(95%CI: 0.875~0.960), 提示模型具有较高的鉴别能力。采用约登指数最大化原则探索最佳截断点, 得到总体最佳阈值为10.86%(0.1086), 该点对应敏感度90.91%、特异度85.61%。按绝经状态分层分析: 绝经前组AUC为0.893(95%CI: 0.799~0.986), 最佳截断值为0.111, 对应敏感度85.00%、特异度89.42%; 绝经后组AUC为0.847(95%CI: 0.745~0.948), 最佳截断值为0.106, 对应敏感度93.48%、特异度71.43%。总体及分层的最佳

截断值均与预设10%阈值接近, 提示采用10%作为统一阈值具有可比性与可操作性。见图1。

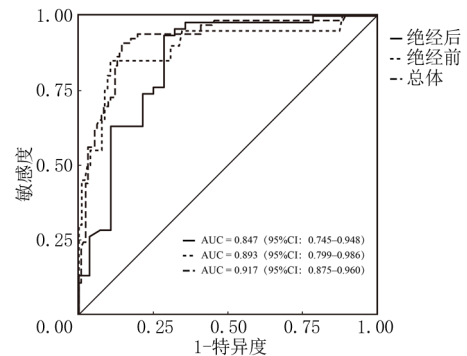


图1 IOTA·LR2模型预测附件肿物良恶性的ROC曲线(总体及按绝经状态分层)

Figure 1 ROC curves of IOTA-LR2 model for predicting benign and malignant adnexal masses (overall and stratified by menopausal status)

为进一步评估绝经后人群阈值调整对特异度提升的影响, 本研究对不同阈值进行了敏感性分析, 结果显示: 当阈值由0.10提高至0.12~0.25时, 绝经后组特异度仍为71.43%, 而敏感度由93.48%下降至91.30%; 当阈值进一步提高至0.30/0.40/0.50时, 敏感度分别下降至82.61%/76.09%/67.39%, 而特异度仅为71.43%/75.00%/78.57%。见表5。提示在本研究样本中, 绝经后人群单纯上调阈值对提升特异度的收益有限, 且可能以敏感度下降为代价。

2.6 IOTA·LR2六变量在中国人群中的独立预测能力 将IOTA·LR2的6个预测变量[年龄(连续变量)、腹水(无=0, 有=1)、带血流的乳头状突起(无=0, 有=1)、实性部分最大径(连续变量)、囊壁不规则(否=0, 是=1)、声影(无=0, 有=1)]同时纳入多因素Logistic回归模型, 以“非良性(恶性+交界性)/良性”为结局。结果显示: 年龄($OR=1.067$, 95%CI: 1.032~1.108, $P<0.001$)、腹水($OR=7.932$, 95%CI: 2.463~28.862, $P<0.001$)、带血流的乳头状突起($OR=26.567$, 95%CI: 2.557~645.321, $P=0.012$)、实性部分最大径($OR=1.041$, 95%CI: 1.017~1.068, $P=0.001$)及囊壁不规则($OR=16.665$, 95%CI: 5.610~59.509, $P<0.001$)均为非良性的独立预测因素。声影在本队列中未显示独立预测价值($P=0.991$)。见表6。基于该六变量模型的预测概率绘制ROC曲线, 模型AUC为0.950(95%CI: 0.919~0.982), 提示其在中国人群中具有良好的判别能力。

表5 绝经后不同阈值敏感性分析
Table 5 Sensitivity analysis of different thresholds after menopause

阈值	敏感度	特异度	Youden	PPV	NPV	PLR	NLR	准确率	TP	FP	TN	FN
0.05	0.978	0.429	0.407	0.737	0.923	1.712	0.050	0.770	45	16	12	1
0.08	0.957	0.643	0.599	0.814	0.900	2.678	0.067	0.838	44	10	18	2
0.10	0.935	0.714	0.649	0.843	0.869	3.271	0.091	0.851	43	8	20	3
0.12	0.913	0.714	0.627	0.840	0.833	3.195	0.121	0.838	42	8	20	4
0.15	0.913	0.714	0.627	0.840	0.833	3.195	0.121	0.838	42	8	20	4
0.20	0.913	0.714	0.627	0.840	0.833	3.195	0.121	0.838	42	8	20	4
0.25	0.913	0.714	0.627	0.840	0.833	3.196	0.122	0.838	42	8	20	4
0.30	0.826	0.714	0.540	0.826	0.714	2.891	0.244	0.784	38	8	20	8
0.40	0.761	0.750	0.511	0.833	0.656	3.044	0.319	0.757	35	7	21	11
0.50	0.674	0.786	0.459	0.838	0.595	3.145	0.415	0.716	31	6	22	15

表6 IOTA·LR2六个预测变量的多因素Logistic回归结果
Table 6 Multivariate Logistic regression results of six predictors of IOTA-LR2

变量	回归系数	标准误	Wald χ^2 值	P值	OR值	95%CI
年龄	0.065	0.018	12.799	<0.001	1.067	1.032~1.108
腹水	2.071	0.628	10.880	0.001	7.932	2.463~28.862
带血流乳头状突起	3.280	1.411	5.403	0.012	26.567	2.557~645.321
实性部分最大径	0.040	0.012	10.363	0.001	1.041	1.017~1.068
囊壁不规则	2.813	0.602	21.807	<0.001	16.665	5.610~59.509
声影	—	—	—	0.991	0.000	0.000~1.45×10 ³¹

注：声影变量在本队列中事件分布稀疏，置信区间跨度较大，需谨慎解读

3 讨论

在女性生殖系统肿瘤中，卵巢/附件恶性肿瘤病死率较高，早期识别有助于改善预后^[21-23]。超声检查是附件肿物术前评估的首选影像手段^[24-25]，但诊断表现受操作者经验与征象提取一致性影响^[26]。本研究在我国单中心队列中验证了IOTA·LR2诊断模型对附件肿物良性与非良性（恶性+交界性）的鉴别能力：在LR2%≥10%阈值下，总体敏感度为90.91%、特异度为85.61%，AUC为0.917（95%CI：0.875~0.960），Kappa值为0.729（P<0.001），提示模型具有较好的判别能力与一致性，支持其作为术前风险评估的辅助工具。

本研究采用IOTA·LR2推荐阈值10%进行二分类判读。分层分析显示，绝经后亚组敏感度较高（93.48%），但特异度下降（71.43%），约登指数相对降低，提示在绝经后人群中固定阈值10%可能带来更多假阳性。考虑到绝经后人群非良性患病率较高，且部分良性病变亦可呈现复杂回声或血流征象，LR2风险评分可能被上调，从而降低特异度。基于本研究获得的LR2连续风险值，进一步进行了ROC分析并采用约登指数最大

化探索最佳截断点，同时对绝经后亚组开展多阈值敏感性分析。结果提示：绝经后最佳阈值与10%阈值非常接近，且在一定阈值范围内单纯上调阈值对特异度提升的收益有限，而在更高截断值下敏感度可能下降。因而在临床实践中，对绝经后人群“阳性（≥10%）”结果更应强调与肿瘤标志物及其他影像学征象联合解读与分层管理，以减少假阳性导致的过度手术风险；未来仍需在更大样本、多中心前瞻性队列中进一步评估并优化绝经后人群的阈值与综合诊断策略，以在敏感度与特异度之间取得更佳平衡。

为评估IOTA·LR2模型变量在中国人群中的适用性与稳健性，本研究将LR2的六个预测变量纳入多因素Logistic回归模型进行验证。结果显示，年龄、腹水、带血流乳头状突起、实性部分最大径及囊壁不规则在本队列中仍为非良性的独立预测因素，效应方向与原模型所依据的临床病理学机制一致；六变量模型的判别能力良好（AUC=0.950，95%CI：0.919~0.982），提示IOTA·LR2的变量结构在中国人群中具有一定可迁移性。需要指出的是，个别征象（如声影）在队列中分布稀疏，回归参数不稳定，因此相关OR值需谨慎解读；仍需更大样本、多中心及前瞻性

队列开展外部验证与校准评估,以进一步确认模型在不同机构与操作者条件下的稳健性与泛化能力。

从临床应用角度,似然比可用于衡量检测结果对疾病概率的改变程度^[27-28]。本研究在 $LR2\% \geq 10\%$ 阈值下总体+LR为6.32、-LR为0.11,提示阴性结果对降低非良性可能性具有较强提示意义,模型更适合作为高敏感度的筛查/排除工具。绝经后亚组+LR相对降低,提示阳性结果的“确诊价值”相对受限,更需要联合血清学指标与影像学综合判断。

值得进一步关注误分类的临床意义。尽管模型总体NPV为94.96%,漏诊风险相对较低,但特异度在绝经后人群中下降意味着假阳性仍需重点管理。在本研究中假阴性为6例,提示以10%为阈值时总体漏诊风险较低;但假阳性为19例,提示仍存在一定比例的“过度警报”。这类假阳性可能与部分良性病变在灰阶或血流表现上呈现“类恶性”征象有关^[29](如复杂囊实性结构、局灶性血流信号、囊壁/分隔不规则等),尤其在绝经后人群中更易出现复杂表现^[12],从而触发LR2风险评分上调。因而临床应用中,IOTA·LR2更适合作为高敏感度的筛查/排除工具:当 $LR2\% < 10\%$ 时可支持低风险管理;而当 $LR2\% \geq 10\%$ (尤其绝经后)时,应采用“二步法”降低误诊风险,即结合肿瘤标志物(如CA125、HE4)及其他影像学信息(专家复核超声、必要时MRI),并综合临床背景进行二次分层,以减少假阳性导致的不必要手术。未来研究可进一步对误分类病例开展影像复核与病理类型归因分析,总结典型“假阳性谱”,并探索更符合我国人群的综合分层策略,以在维持高敏感度的同时提升特异度。

本研究仍存在一定局限性。首先,本研究为单中心回顾性研究,纳入对象均为住院并接受手术且获得病理结果者,可能存在选择偏倚与谱偏倚,研究结论对门诊随访或保守管理人群的外推需谨慎。其次,关于诊断时间窗,本研究设置末次超声至手术 ≤ 120 d的纳入标准以降低时间窗影响,但同意120 d跨度可能偏长,肿物形态、内部结构及血流信号在此期间可能发生变化,从而导致超声征象与最终病理状态不匹配并影响模型特异度与一致性;受回顾性数据限制,本研究未完整保留末次超声与手术的具体日期,无法提供诊断时间间隔的中位数/IQR分布,也无法按时间窗

分层开展敏感性验证。未来研究需在前瞻性或多中心队列中完整记录超声与手术日期,设置更严格的时间窗(如 ≤ 30 d或 ≤ 60 d)并进行分层分析。再次,本研究分层后部分亚组样本量较小:绝经后亚组良性病例数有限,绝经前亚组非良性病例数亦较少(19~20例),导致分层敏感度/特异度及其置信区间稳定性下降。除此之外,超声检查由经验较丰富的专科医师完成,模型表现可能高于基层或非专科环境;本研究未进行多观察者一致性评估,无法量化IOTA术语与征象提取的观察者一致性。另外,本研究排除了图像质量不足以提取关键征象的病例,该排除标准有助于保证IOTA术语记录的可重复性,但可能引入选择偏倚,使模型在本研究中的诊断性能(尤其特异度)较真实临床环境偏高。未来真实世界研究应纳入不同图像质量水平病例,评估图像质量对模型表现的影响,并探索质量分层或最低图像质量门槛的应用策略。综上,仍需更大样本、多中心前瞻性研究,并结合血清学指标及其他IOTA模型进行综合验证与校准评估。

[参考文献]

- [1] Tang W, Cai Q, Huang K, et al. The global burden of polycystic ovary syndrome, endometriosis, uterine fibroids, cervical cancer, uterine cancer, and ovarian cancer from 1990 to 2021[J]. BMC Public Health, 2025, 25(1):1774.
- [2] Sun P, Yu C, Yin L, et al. Global, regional, and national burden of female cancers in women of child-bearing age, 1990–2021: Analysis of data from the global burden of disease study 2021[J]. EClinicalMedicine, 2024, 74:102713.
- [3] Wei J, Wang M, Wu Y. A disulfidptosis-related lncRNAs cluster to forecast the prognosis and immune landscapes of ovarian cancer[J]. Front Genet, 2024, 15:1397011.
- [4] Merritt MA, Abe SK, Islam MR, et al. Reproductive factors and risk of epithelial ovarian cancer: Results from the Asia Cohort Consortium[J]. Br J Cancer, 2025, 132(4):361-370.
- [5] Jiang B. Analysis of global ovarian cancer disease burden and its changing trend from 1990 to 2021 [J]. BMC Womens Health, 2025, 25(1):352.
- [6] Li L, Zhang J, Song N, et al. Multigene germline and somatic testing for epithelial ovarian cancer in China[J]. NPJ Precision Oncol, 2025, 9(1):281.
- [7] Li Y, Dou J, Fu Y, et al. Targeted immunotherapies and nanomedicines for ovarian cancer: The way forward[J]. NPJ Prec Oncol, 2026, 10(1):80.
- [8] Zelba H, Kyzirakos C, Kayser S, et al. Case report: Long-term survival of a patient with cerebral metastasized ovarian carcinoma treated with a personalized peptide vaccine and anti-PD-1 Therapy[J]. Vaccines (Basel), 2024, 12(4):397.

- [9] Jara-Rosales S, González-Stegmaier R, Rotarou ES, et al. Risk factors for ovarian cancer in south america; A literature review[J]. *J Pers Med*, 2024, 14(9):992.
- [10] Szewczuk W, Szewczuk O, Czajkowski K, et al. Methylation of the selected TP53 introns in advanced-stage ovarian carcinomas[J]. *J Cancer*, 2024, 15(13):4040-4046.
- [11] Wang G, Liu X, You Y, et al. Development and clinical validation of a seven-gene signature based on tumor stem cell-related genes to predict ovarian cancer prognosis[J]. *J Ovarian Res*, 2024, 17(1):58.
- [12] Kicman A, Gacuta E, Kulesza M, et al. Diagnostic utility of selected matrix metalloproteinases (MMP-2, MMP-3, MMP-11, MMP-26), HE4, CA125 and ROMA algorithm in diagnosis of ovarian cancer[J]. *Int J Mol Sci*, 2024, 25(11):6265.
- [13] Pan R, Zhang S, Zhang X, et al. Clinical value of ACR O-RADS combined with CA125 in the risk stratification of adnexal masses[J]. *Front Oncol*, 2024, 14:1369900.
- [14] Pirri C, Pirri N, Macchi V, et al. Ultrasound imaging of thoracolumbar fascia: A systematic review [J]. *Medicina (Kaunas)*, 2024, 60(7):1090.
- [15] Crisan D, Tarnowietzki E, Bernhard L, et al. Rationale for using high-frequency ultrasound as a routine examination in skin cancer surgery: A practical approach [J]. *J Clin Med*, 2024, 13(7):2152.
- [16] Sharma A, Burnell M, Gentry-Maharaj A, et al. Factors affecting visualization of postmenopausal ovaries: Descriptive study from the multicenter United Kingdom Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) [J]. *Ultrasound Obstet Gynecol*, 2013, 42(4):472-477.
- [17] Koutras A, Perros P, Prokopakis I, et al. Advantages and limitations of ultrasound as a screening test for ovarian cancer [J]. *Diagnostics (Basel)*, 2023, 13(12):2078.
- [18] Geysels A, Garofalo G, Timmerman S, et al. ADNEX-AI: Automated extraction of ultrasound predictors for interpretable ovarian cancer risk stratification[J]. *NPJ Precis Oncol*, 2025, 10(1):18.
- [19] Yasmin, Parveen N, Gupta B, et al. Comparison of ADNEX model with GI-RADS ultrasound scoring system in evaluation of adnexal mass[J]. *J Obstet Gynaecol India*, 2025, 75(Suppl 1):86-92.
- [20] Kougioumtsidou A, Karavida A, Mamopoulos A, et al. Performance of International Ovarian Tumor Analysis (IOTA) predictive models in preoperative discrimination between benign and malignant adnexal lesions: Preliminary outcomes in a tertiary care hospital in Greece [J]. *Arch Gynecol Obstet*, 2025, 311(1):113-122.
- [21] Englisz A, Smycz-Kubańska M, Mielczarek-Palacz A. Sensitivity and specificity of selected biomarkers and their combinations in the diagnosis of ovarian cancer [J]. *Diagnostics (Basel)*, 2024, 14(9):949.
- [22] Lahelma M, Rauhamaa H, Isomeri O, et al. Real-life treatment patterns and time to next treatment among patients with ovarian cancer in the pre-PARP inhibitor era: The OCRWE-Finland study[J]. *Acta Oncol*, 2024, 63:772-782.
- [23] Xiao L, Li H, Jin Y. Automated early ovarian cancer detection system based on bioinformatics [J]. *Sci Rep*, 2024, 14(1):22887.
- [24] 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 [J]. *Front Oncol*, 2024, 14:1354837.
- [25] Avesani G, Panico C, Nougaret S, et al. ESR essentials: Characterisation and staging of adnexal masses with MRI and CT-practice recommendations by ESUR [J]. *Eur Radiol*, 2024, 34(12):7673-7689.
- [26] Du Y, Guo W, Xiao Y, et al. Ultrasound-based deep learning radiomics model for differentiating benign, borderline, and malignant ovarian tumours: A multi-class classification exploratory study[J]. *BMC Med Imaging*, 2024, 24(1):89.
- [27] Cuesta G, Puerta-Alcalde P, Vergara A, et al. Reply to Paranhos-Baccalâet al. Comment on "Cuesta et al. An assessment of a new rapid multiplex pcr assay for the diagnosis of meningoencephalitis. *Diagnostics* 2024, 14, 802" [J]. *Diagnostics (Basel)*, 2024, 14(17):1885.
- [28] Elkahwagy DMAS, Kiriacos CJ, Mansour M. Logistic regression and other statistical tools in diagnostic biomarker studies[J]. *Clin Transl Oncol*, 2024, 26(9):2172-2180.
- [29] Wei H, Wang Y, Li J, et al. Diagnosis of benign and malignant peripheral lung lesions based on a feature model constructed by the random forest algorithm for grayscale and contrast-enhanced ultrasound[J]. *Front Oncol*, 2024, 14:1352028.

(本文编辑:王聪)