

ORIGINAL RESEARCH

Research on the diagnostic value of ultrasound combined with MRI features based on deep learning in distinguishing benign and malignant ovarian tumors

Ning Sun¹, Xiangli Yang¹, Lili Fan², Nan Zhang^{2,*}, Yan Xue^{3,*}

¹Department of Radiology, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, 030032 Taiyuan, Shanxi, China

²Department of Stomatology, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, 030032 Taiyuan, Shanxi, China

³Department of Obstetrics and Gynecology, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, 030032 Taiyuan, Shanxi, China

***Correspondence**

z_nan0628@163.com

(Nan Zhang);

xueyan@sxhqh.com.cn

(Yan Xue)

Abstract

Background: This study aimed to assess the diagnostic value of a deep learning (DL)-based multimodal approach that combines ultrasound (US) and magnetic resonance imaging (MRI) features in differentiating benign and malignant ovarian tumors, and to develop an intelligent auxiliary diagnostic tool. **Methods:** A total of 887 patients (665 benign, 222 malignant) with pathologically confirmed ovarian tumors from 2022 to 2024 were retrospectively enrolled. All patients underwent preoperative US and MRI within one week. A dual-channel DL model was constructed: the US branch (ResNet50) extracted 2D features from grayscale color Doppler flow imaging, while the MRI branch (3D ResNeXt101) extracted 3D features from T2-weighted imaging (T2WI), dynamic contrast-enhanced (DCE)-MRI, and apparent diffusion coefficient (ADC) maps. The extracted features were integrated using an attention-based fusion mechanism. With pathology as the gold standard, the diagnostic performance of the proposed model was compared with US alone, MRI alone, and the Assessment of Different NEoplasias in the adneXa (ADNEX) model. **Results:** In the test cohort ($n = 266$), the DL model showed a sensitivity of 92.73%, specificity of 98.58%, accuracy of 97.37%, and an area under the curve (AUC) of 0.957, which were significantly higher than those of US (AUC = 0.792, $z = 3.92$, $p < 0.001$), MRI (AUC = 0.844, $z = 2.76$, $p = 0.006$), and ADNEX (AUC = 0.885, $z = 2.07$, $p = 0.022$). **Conclusions:** The DL-based US-MRI multimodal fusion model significantly enhances the diagnostic accuracy for differentiating benign and malignant ovarian tumors, providing a promising intelligent auxiliary tool for early and precise diagnosis.

Keywords

Ovarian tumor; Deep learning; Ultrasonography; Magnetic resonance imaging; Benign and malignant differentiation; Diagnostic value; Multimodal fusion

1. Introduction

Ovarian tumors are one of the most common tumors of the female reproductive system, with malignant types (ovarian cancer) often referred to as the “king of gynecological cancers” due to their insidious onset and high aggressiveness [1]. Globally, more than 300,000 new cases of ovarian cancer are diagnosed each year, and its mortality rate ranks first among gynecological malignancies. The 5-year survival rate of Chinese patients remains below 50%, which is significantly lower than that of developed countries [2]. Clinical data show that approximately 70% of ovarian cancer patients are diagnosed at an advanced stage (III–IV), when tumor metastasis to the pelvic and abdominal cavity has already occurred, resulting in a sharp decline in the 5-year survival rate to below 30%. In contrast, early-stage (I) patients can achieve a 5-year survival rate exceeding 90% after standardized treatment [3, 4]. Therefore, early and accurate differentiation between benign

and malignant ovarian tumors is crucial for improving patient prognosis.

Currently, the preoperative diagnosis of ovarian tumors primarily relies on imaging examinations. Ultrasound (US), being radiation-free, real-time, convenient, and cost-effective, serves as a first-line screening modality [5]. By evaluating tumor characteristics, such as size, morphology, wall structure, septa, internal echogenicity, and vascularity, experienced physicians can make a preliminary assessment for most tumors [6, 7]. Magnetic Resonance Imaging (MRI), known for its superior soft tissue resolution and multi-sequence imaging capabilities, plays a vital complementary role in the differential diagnosis of indeterminate lesions, identification of tumor origin, and preoperative staging [8, 9]. However, traditional imaging diagnosis remains highly dependent on physicians’ experience and subjective visual assessment, leading to intra- and inter-observer variability. For atypical or complex cases, even the combined use of US and MRI may not achieve

optimal diagnostic accuracy [10]. With the rapid development of artificial intelligence (AI), deep learning (DL), its core branch, has made revolutionary contributions to medical image analysis [11]. Convolutional Neural Networks (CNNs) can automatically learn and extract multi-level, high-dimensional features from large-scale imaging data that are often imperceptible to the human eye, enabling accurate disease detection, segmentation, and classification [12, 13]. Currently, the application of DL algorithms in the diagnosis and prognosis prediction of gynecological tumors is mainly focused on cervical cancer, followed by ovarian and endometrial cancers [14]. Meta-analyses have confirmed that AI- and DL-based imaging features possess high diagnostic value for ovarian tumors [15]. Previous studies have demonstrated that DL models analyzing either US or MRI images alone outperform traditional manual assessments, establishing a new paradigm for computer-aided diagnosis [16–19]. Integrating DL-derived features from both US and MRI modalities is expected to yield a more comprehensive and robust diagnostic framework, thereby achieving a “1 + 1 > 2” synergistic effect in diagnostic performance.

Therefore, this study aimed to investigate the diagnostic efficacy of a DL-based multimodal model combining US and MRI features for differentiating benign from malignant ovarian tumors, and to develop an intelligent auxiliary diagnostic tool for clinical practice.

2. Materials and methods

2.1 Study subjects

This study was a retrospective cohort analysis. Data were collected from patients with ovarian tumors treated at Shanxi Bethune Hospital from January 2022 to December 2024. Based on preliminary pilot study results (AUC of ultrasound alone: 0.82; expected AUC of the combined model: 0.95; $\alpha = 0.05$; $\beta = 0.2$), the required sample size was calculated using MedCalc 19.0 software (MedCalc Software Ltd, Ostend, Belgium). A total of 900 patients were initially considered; after applying the inclusion and exclusion criteria, 887 patients were finally enrolled, including 665 with benign tumors and 222 with malignant tumors. In the binary classification task, borderline tumors were classified as malignant tumors. The detailed patient selection flowchart is shown in Fig. 1. This study was approved by the Institutional Ethics Committee of Shanxi Bethune Hospital (approval number: 2025-356). All patients provided written informed consent, and the study complied with the Declaration of Helsinki.

2.2 Inclusion and exclusion criteria

Inclusion criteria: (a) Pathologically confirmed ovarian tumors (benign, borderline, and malignant); (b) Completed transvaginal ultrasound and MRI examinations within 1 week before surgery; (c) Image quality met diagnostic requirements (no severe motion artifacts and clear lesion visualization); (d) Complete clinical and pathological data (age, tumor markers, and pathological subtypes).

Exclusion criteria: (a) Previous history of radiotherapy, chemotherapy, targeted therapy, or surgery; (b) Concomitant

other pelvic malignancies (e.g., endometrial cancer metastasis); (c) Incomplete image data (e.g., missing MRI sequences, unrecorded blood flow signals on ultrasound).

2.3 Ultrasound and MRI feature data acquisition

2.3.1 Ultrasound examination

Ultrasound examinations were performed using Philips EPIQ 7 (Philips Healthcare, Amsterdam, Netherlands) and GE Voluson E10 (GE Healthcare, Chicago, IL, USA) color Doppler systems, equipped with a 5–9 MHz transvaginal probe. Scanning protocols included: (a) 2D grayscale ultrasound: routine acquisition of transverse, longitudinal, and oblique plane images, recording tumor size (maximum diameter), morphology (cystic/solid/mixed), margin, septa (thickness/smoothness), and papillary projections (number/size); (b) Color Doppler Flow Imaging (CDFI): Velocity scale set to 2–10 cm/s to assess intra- and peritumoral blood flow signals (Adler classification), and to measure the resistance index (RI) and pulsatility index (PI). The Adler classification was used to evaluate blood flow signals within and around tumors, grading them based on the richness and distribution pattern of blood flow to assist in determining the malignancy of tumors. All US images were stored in Digital Imaging and Communications in Medicine (DICOM) format, and 3–5 representative images were selected from each patient. US examinations were performed in real-time by experienced US specialists, who also performed image acquisition and preliminary interpretation.

2.3.2 MRI examination

MRI scans were performed using a Siemens Prisma 3.0T MRI scanner (Siemens Healthineers, Erlangen, BA, Germany) with a body phased-array coil. Scanning sequences and parameters are summarized in Table 1. For DCE-MRI, dynamic contrast-enhanced imaging was conducted immediately after contrast injection, acquiring five phases at approximately 60 seconds per phase. For analysis, early arterial and equilibrium phase images showing the most pronounced enhancement were selected to best capture tumor hemodynamics.

2.3.3 Image preprocessing

SimpleITK library (Python 3.8) was used to convert DICOM format images to Neuroimaging Informatics Technology Initiative (NIfTI) format, with coordinates standardized to the tumor center as the origin. Two radiologists with more than 5 years of experience in gynecological imaging diagnosis independently and manually delineated the tumor regions; discrepancies were resolved through consensus. The region of interest (ROI) masks were generated using 3D Slicer version 4.13, which served as the annotation tool for subsequent feature extraction. For Diffusion-Weighted Imaging (DWI), ROI were drawn within the solid components of the tumor, avoiding cystic, necrotic, or hemorrhagic areas to ensure accurate representation of viable tissue. Diffusion restriction was determined based on low-signal regions on the ADC maps corresponding to high signal on the DWI. Z-score normalization was applied to ultrasound images to reduce device-dependent grayscale

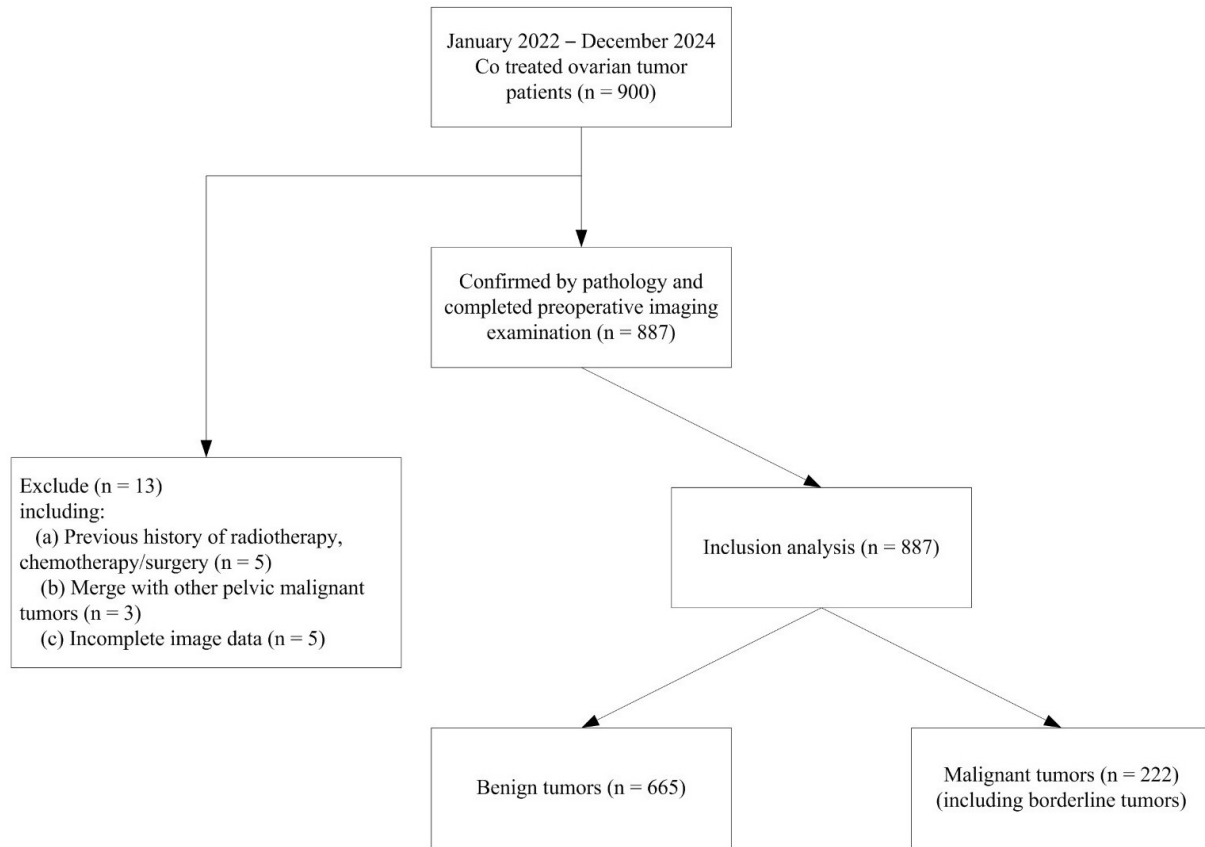


FIGURE 1. Flowchart of the study population selection process.

TABLE 1. MRI scanning sequences and parameters.

Sequence Type	TR (ms)	TE (ms)	Slice thickness (mm)	Vision (mm)	Matrix
T1WI axial position	400–600	8–12	3.0	280 × 280	320 × 320
T2WI axial/sagittal view	3000–4000	80–120	3.0	280 × 280	320 × 320
DWI (b-value = 0, 800 s/mm ²)	4000–5000	60–80	3.0	280 × 280	128 × 128
Dynamic Enhanced Scanning (DCE-MRI)	3.5–4.0	1.5–2.0	3.0	280 × 280	320 × 320

MRI: magnetic resonance imaging; TR: Repetition Time; TE: Echo Time; T1WI: T1-weighted imaging; T2WI: T2-weighted imaging; DWI: Diffusion-weighted imaging.

variation, while the MRI sequences were normalized to the range (0, 1) to preserve intra-sequence signal characteristics. All ROI images were resized to 256 × 256 pixels (US) or 128 × 128 × 32 voxels (MRI, 32 slices centered on the lesion).

Image quality control was performed by an independent physician (non-annotator) using a 4-point scale. Concisely, images scoring ≥ 3 were included in the analysis; substandard images (< 3 points) were re-acquired or excluded. For quality assurance, 20% of cases were randomly re-annotated under double-blind repeated conditions, and the intraclass correlation coefficient (ICC) was calculated to assess annotation consistency (*i.e.*, ICC > 0.8 indicating good consistency).

2.4 Deep learning model construction

A dual-channel multimodal fusion model was adopted (Fig. 2), comprising three main components:

1. Unimodal Feature Extraction Modules: The unimodal feature extraction module consists of an ultrasound branch

and an MRI branch. The ultrasound branch is based on the ResNet50 architecture, with inputs of 2D ultrasound images (grayscale + CDFI fused images), and extracts morphological features (edges, textures) and blood flow features (signal intensity distribution) through convolutional layers. The MRI branch is based on the 3D ResNeXt101 architecture, with inputs of MRI multi-sequence fused data (T2WI + DCE-MRI + ADC maps), and extracts spatial features (intratumoral structural heterogeneity) and functional features (degree of diffusion restriction, blood perfusion) through 3D convolutions.

2. Cross-modal Fusion Module: The cross-modal fusion module employs an attention mechanism fusion (Attention Fusion). It performs dimension alignment between the 2D feature maps (512 channels) output by the US branch and the 3D feature maps (1024 channels) output by the MRI branch (reducing 3D features to 2D via global average pooling); calculates modal attention weights (Eqn. 1), followed by weighted feature fusion (Eqn. 2):

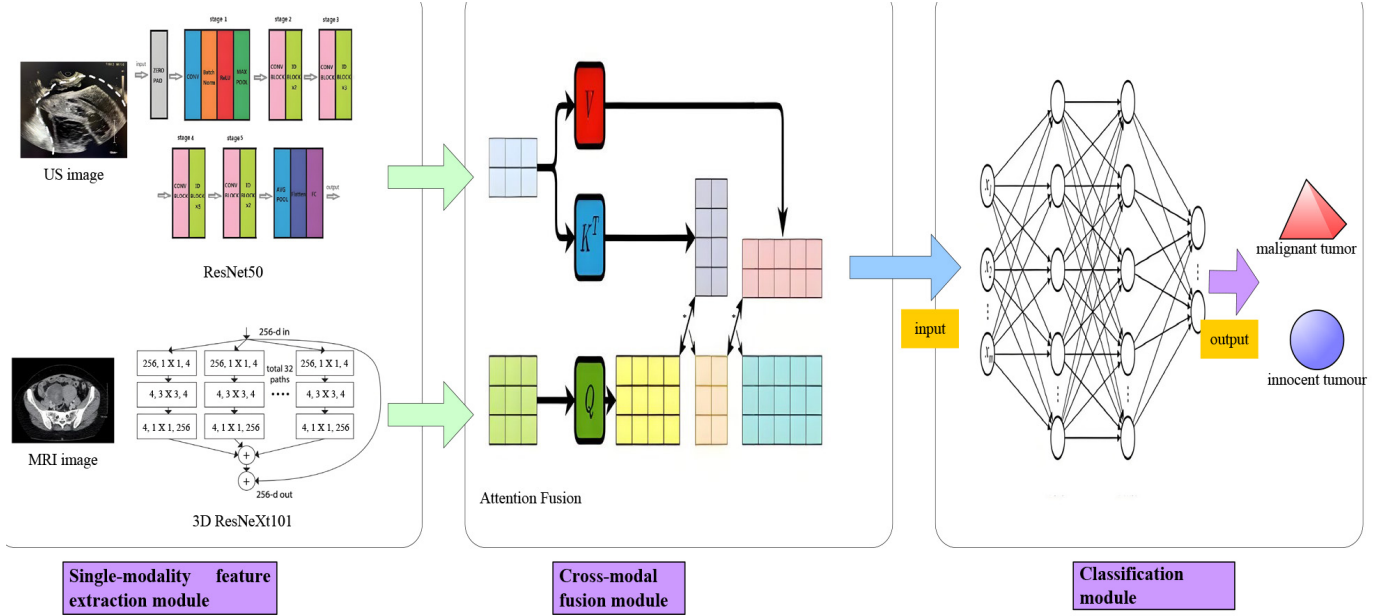


FIGURE 2. Schematic diagram of the deep learning multimodal fusion model architecture. US: ultrasound; MRI: magnetic resonance imaging; ZERO PAD: Zero Padding; CONV: Convolutional Layer; Batch Norm: Batch Normalization; ReLU: Rectified Linear Unit; MAX POOL: Max Pooling Layer; ID: Identity; AVG POOL: Global Average Pooling; FC: Fully Connected Layer; ResNet50: Residual Network with 50 layers. *: Sigmoid activation function.

$$\begin{aligned} W_u &= \text{Softmax}(\text{FC}(\text{GAP}(F_u))), \\ W_m &= \text{Softmax}(\text{FC}(\text{GAP}(F_m))) \end{aligned} \quad (1)$$

$$F_{\text{fusion}} = W_u \cdot F_u + W_m \cdot F_m \quad (2)$$

In Eqn. 1: W_u : ultrasound modal attention weight; F_u : ultrasound features; W_m : MRI modal attention weight; F_m : MRI features; GAP: global average pooling; FC: fully connected layer.

3. Classification module: After the fused features pass through two fully connected layers (with Dropout = 0.5), the predicted probabilities of benign and malignant tumors (benign = 1, malignant = 2) are output through the Softmax function.

2.5 Feature extraction and fusion

2.5.1 Unimodal features

Ultrasound features: High-dimensional features (512 dimensions) were automatically extracted, including: (a) Low-level features: edge gradients, texture entropy (reflecting cyst wall smoothness), and mean blood flow signal intensity (reflecting blood supply abundance); (b) High-level features: probability maps of papillary projections (confidence of papillary structures within the ROI output by the model), and septal thickness distribution histograms.

MRI features: High-dimensional features (1024 dimensions) were automatically extracted, including: (a) T2WI texture features: Gray-Level Co-occurrence Matrix (GLCM) parameters (contrast, correlation), and high-frequency components of wavelet transform (reflecting solid

component density); (b) DCE-MRI features: time-signal curve types (persistently rising/plateau/washout), and time to enhancement peak; (c) ADC features: ADC value histogram parameters (mean, median, skewness).

2.5.2 Fusion features

A 1536-dimensional joint feature vector was generated via the attention fusion module, where the first 512 dimensions were US features (after weighting) and the latter 1024 dimensions were MRI features (after weighting), for final classification.

2.6 Model training and evaluation

2.6.1 Dataset division

The data were divided into two parts using stratified sampling at a 7:3 ratio: (a) Training set (70%): including 1622 images of 621 ovarian tumors, used for model parameter learning; (b) Test set (30%): including 608 images of 266 ovarian tumors, to independently evaluate the model's generalization ability (the dataset was not accessed during model training).

2.6.2 Model training environment

Hardware: NVIDIA RTX A6000 GPU (48GB VRAM), Intel i9-12900K CPU, 64GB RAM.

Software: PyTorch 1.10 framework, CUDA 11.3 acceleration (NVIDIA Corporation, Santa Clara, CA, USA), Python 3.8, with open-source libraries including Torchvision, NiBabel, and Scikit-learn.

2.6.3 Training parameters

Optimizer: Adam (initial learning rate = 0.001, weight decay = 1×10^{-5} , $\beta_1 = 0.9$, $\beta_2 = 0.999$).

Loss function: Weighted cross-entropy loss (weights for benign: borderline: malignant = 1:2:3, to address class imbalance).

ance).

Batch size = 16 (US)/8 (MRI), number of iterations = 200 epochs.

2.7 Statistical methods

Statistical analyses were performed using SPSS 26.0 (IBM Corp., Armonk, NY, USA). Categorical variables were expressed as n (%), and comparisons between groups were performed using the chi-square test. Continuous variables with normal distribution were expressed as mean $\pm$ standard deviation (SD), and comparisons between groups were performed using the t -test; those with a non-normal distribution were expressed as M (Q25, Q75), and comparisons between groups were performed using the Mann-Whitney U test. The diagnostic performance of the DL model for differentiating benign from malignant ovarian tumors was analyzed, including accuracy, sensitivity, specificity, positive predictive value, and negative predictive value, with Receiver Operating Characteristic (ROC) curve analysis performed. The DL model was compared with the US-alone, MRI-alone, and the ADNEX (Assessment of Different NEoplasias in the adneXa) model; comparisons of AUC between groups were performed using the DeLong test. A p -value < 0.05 was considered statistically significant.

3. Results

3.1 Comparison of clinical characteristics of study subjects

The clinical and pathological characteristics of patients in the training and test groups are summarized in Table 2. There were no significant differences between the two groups in terms of age, maximum tumor diameter, Carbohydrate Antigen (CA)125 level, or pathological distribution ($p > 0.05$).

3.2 Diagnostic performance of different methods in differentiating benign and malignant ovarian tumors

In the test cohort, the DL-based multimodal model achieved excellent diagnostic performance, demonstrating the highest sensitivity, specificity, and overall accuracy, compared with US-alone, MRI-alone, and the ADNEX model. Details results are shown in Tables 3 and 4.

3.3 ROC curve analysis

The ROC curve analysis demonstrated that the DL multimodal fusion model achieved the highest area under the curve (AUC = 0.957) for differentiating benign and malignant ovarian tumors. The AUC was significantly higher than that of US alone ($z = 3.92$, $p < 0.001$), MRI alone ($z = 2.76$, $p = 0.006$), and the ADNEX model ($z = 2.07$, $p = 0.022$). Details are shown in Table 5 and Fig. 3.

4. Discussion

As the “king of gynecological cancers”, early diagnosis of ovarian cancer is crucial for improving patient prognosis. This

study developed a deep learning (DL)-based multimodal fusion model that combines ultrasound (US) and MRI to enhance the accuracy of differentiating between benign and malignant ovarian tumors. By automatically extracting and fusing high-dimensional imaging features from both modalities, the model achieved significantly higher diagnostic performance than US alone, MRI alone, or the ADNEX model, thereby demonstrating the substantial advantage of multimodal deep learning in ovarian tumor diagnosis.

The findings of this study indicate that combining the DL-extracted US and MRI features can significantly improve diagnostic accuracy for distinguishing benign from malignant ovarian tumors. Using Convolutional Neural Networks (CNN), the model automatically identifies latent high-dimensional features in US and MRI images, effectively compensating for the limitations of traditional visual assessment [20, 21]. Compared with the limitation of conventional imaging diagnosis relying on physicians’ subjective experience, this model effectively reduces inter-observer variability by automatically extracting deep features that are difficult for the human eye to recognize, such as ultrasound probability maps of papillary projections and MRI DCE-MRI time-signal curve types. It particularly exhibits better recognition ability for atypical cases, such as mixed tumors and small papillary projections [22, 23].

From a clinical practice perspective, the high sensitivity of the model can reduce the missed diagnosis rate of malignant tumors, while the high specificity can reduce over-treatment of benign tumors, which is of great significance for avoiding unnecessary surgical trauma and optimizing treatment decisions [24]. For patients with International Federation of Gynecology and Obstetrics (FIGO) stage I, early diagnosis can increase the 5-year survival rate to more than 90%, and the high diagnostic performance of this model is expected to provide support for timely intervention in more early-stage cases [25].

This study introduced an innovative dual-channel fusion architecture. The 2D residual architecture of ResNet50 was optimized for US 2D planar data, efficiently extracting morphological and hemodynamic features through residual connections and moderate depth [26]. The 3D ResNeXt101 branch, designed for volumetric MRI data, captured spatial-functional information using grouped convolutions and high cardinality. An attention-based fusion mechanism dynamically weighted the contributions of both modalities, ensuring optimal integration. The superior AUC of the combined model confirmed the complementarity of multimodal information [27]. By integrating the real-time vascular assessment capability of ultrasound with superior soft-tissue contrast and functional imaging power of MRI (DWI, DCE), this model provides a comprehensive reflection of the tumor biology, including blood supply, cell density, and perfusion patterns [28]. This fusion strategy offers a novel approach to addressing the problem of limited information from a single modality, and can serve as a framework for other multimodal medical image applications, such as breast and liver tumor diagnosis [29]. From a clinical translation perspective, the model could be integrated into hospital Picture Archiving and Communication Systems (PACS) workflows to provide real-time auxiliary diagnostic assistance during radiologists’ image review. This would reduce the inter-observer variability and shorten reporting time. For cases

TABLE 2. Comparison of clinical characteristics of study subjects.

Feature indicator	Training group (n = 621)	Test group (n = 266)	Statistical value	p value
Age (yr, $\bar{x} \pm s$)	55.57 $\pm$ 9.25	55.77 $\pm$ 10.64	$t = 0.260$	0.795
Maximum diameter of tumor (cm, $\bar{x} \pm s$)	8.39 $\pm$ 3.01	8.48 $\pm$ 2.78	$t = 0.448$	0.654
CA125 (U/mL, M (Q25, Q75))	30 (21, 76)	30 (23, 36)	$Z = 1.759$	0.079
Pathological examination results (n (%))				
Benign	454 (73.11)	199 (74.81)	$\chi^2 = 0.278$	0.598
Malignant	167 (26.89)	67 (25.19)		
Tumor nature (example) (n (%))				
Cystic	294 (47.34)	135 (50.75)	$\chi^2 = 1.481$	0.477
Reality	134 (21.58)	59 (22.18)		
Mixed	193 (31.08)	72 (27.07)		
FIGO Installment (n (%))				
I	117 (18.84)	55 (20.68)	$\chi^2 = 0.810$	0.847
II	83 (13.37)	32 (12.03)		
III	250 (40.26)	110 (41.35)		
IV	171 (27.54)	69 (25.94)		
Grading (n (%))				
Low-level	297 (47.83)	122 (45.86)	$\chi^2 = 0.325$	0.850
Medium level	210 (33.82)	92 (34.59)		
High-level	114 (18.36)	52 (19.55)		

CA125: Carbohydrate Antigen 125; FIGO: International Federation of Gynecology and Obstetrics.

TABLE 3. Diagnostic performance of different methods for benign and malignant ovarian tumors.

Detection method	Result	Pathological diagnosis		Total
		Malignant	Benign	
Ultrasound diagnosis alone				
	Malignant	35	11	46
	Benign	20	200	220
MRI diagnosis alone				
	Malignant	42	16	58
	Benign	13	195	208
ADNEX model diagnosis				
	Malignant	45	10	55
	Benign	10	201	211
DL model diagnosis				
	Malignant	51	3	54
	Benign	4	208	212
-	Total	55	211	-

DL: deep learning; ADNEX: Assessment of Different NEoplasias in the adnexa; MRI: magnetic resonance imaging.

TABLE 4. Sensitivity and specificity of different diagnostic methods for benign and malignant ovarian tumors (%).

Detection method	Sensitivity	Specificity	Positive predictive value	Negative predictive Value	Accuracy
Ultrasound diagnosis alone	63.64	94.79	76.09	90.91	88.35
MRI standalone diagnosis	76.36	92.42	72.41	93.75	89.10
ADNEX model diagnosis	81.82	95.26	81.82	95.26	92.48
DL model diagnosis	92.73	98.58	94.44	98.11	97.37

DL: deep learning; ADNEX: Assessment of Different NEoplasias in the adnexa; MRI: magnetic resonance imaging.

TABLE 5. Comparison of AUC values of different diagnostic methods.

Diagnostic method	AUC	Standard error	<i>p</i> value	95% confidence interval
Ultrasound diagnosis alone	0.792	0.041	<0.001	0.712–0.872
MRI standalone diagnosis	0.844	0.035	<0.001	0.774–0.913
ADNEX model diagnosis	0.885	0.032	<0.001	0.823–0.948
DL model diagnosis	0.957	0.021	<0.001	0.915–0.998

MRI: magnetic resonance imaging; ADNEX: Assessment of Different NEoplasias in the adnexa; DL: deep learning; AUC: Area Under the Curve.

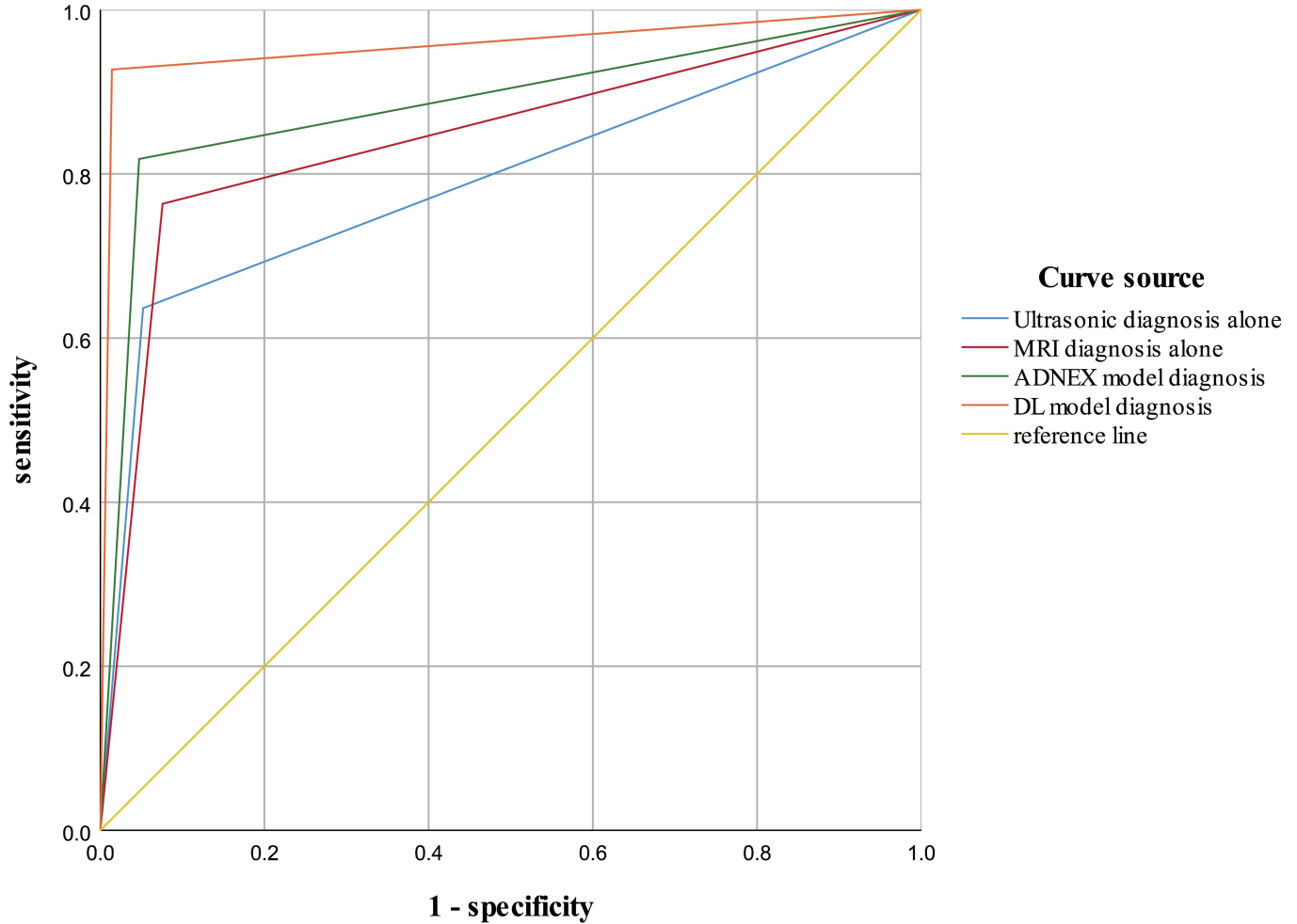


FIGURE 3. ROC Curves of Different Diagnostic Methods. MRI: magnetic resonance imaging; ADNEX: Assessment of Different NEoplasias in the adnexa; DL: deep learning.

with high Ovarian-Adnexal Reporting and Data System (O-RADS) scores (4–5) identified during the initial US screening, the system could combine the MRI-derived features to provide rapid malignancy risk stratification, supporting timely surgical or follow-up decisions and streamlining diagnostic workflows.

Recent studies have investigated the application of DL in ovarian tumor imaging, yet most have focused on single-modality approaches. For example, US-based CCN models have reported AUCs around 0.941 [30], while an MRI-based DL model achieved an AUC of 0.89 [31], both lower than the combined model in this study (0.957). The AUC of conventional combined diagnosis typically ranges from 0.85 to 0.90 and is limited by inter-physician experience variability [32–34]. In contrast, our automated feature fusion approach

eliminated subjective bias and revealed latent cross-modality relationships, achieving a notable improvement in diagnostic accuracy.

Compared with the ADNEX model, our DL model does not rely on serum markers, yet achieves higher accuracy using only imaging data, making it particularly suitable for early-stage patients with normal CA125 levels or atypical cases [35, 36]. This advantage expands the model's application scenarios and reduces reliance on laboratory tests. Furthermore, Adusumilli *et al.* [37] recently proposed a PyTorch-Medical Open Network for AI framework incorporating multiple architectures (ResNet, Densely Connected Convolutional Network, Transformer-based Unified Efficient Segmentation Transformer, and attention-based Multiple Instance Learning).

Although their design is more comprehensive, it remains conceptual, whereas our implemented ResNet-based model provides direct experimental validation and demonstrated clinical feasibility.

Despite the significant findings of this study, the following limitations should be acknowledged: (a) This was a single-center retrospective cohort study, and sample selection may be subject to bias. The generalizability of the model requires validation with multi-center data. (b) The US and MRI data were acquired from specific models of equipment. Variations in image quality and parameter settings across different brands of devices may affect the model's generalizability, and cross-device compatibility issues need to be addressed for practical application. (c) The study included benign, borderline, and malignant tumors; however, the diagnostic performance for borderline tumors was not analyzed separately. Borderline tumors exhibit biological behavior intermediate between benign and malignant, with more complex imaging features. Future studies should specifically optimize the model's recognition ability for this subgroup.

To address the above limitations, future research can be advanced in the following directions: (a) Conduct multi-center, prospective studies incorporating data from different devices and diverse populations (*e.g.*, young women, pregnant patients) to validate the model's stability and universality and explore its feasibility for deployment in primary hospitals. (b) Incorporating additional data modalities (*e.g.*, Computed Tomography, Positron Emission Tomography-Computed Tomography) and clinical biomarkers (*e.g.*, Human Epididymis Protein 4, miRNA profiles) to enhance diagnostic accuracy through multi-omics fusion. (c) Implementing advanced data harmonization techniques and exploring federated learning or domain adaptation to improve model generalization in real-world clinical environments. (d) Developing lightweight, mobile-compatible DL models that can integrate with PACS workflows, enabling real-time decision support for radiologists during patient evaluation.

5. Conclusions

This study demonstrates that a deep learning-based multimodal fusion model integrating US and MRI can significantly improve the diagnostic performance in differentiating benign and malignant ovarian tumors. The proposed deep learning model shows strong potential for clinical translation as an intelligent “second opinion” system to assist radiologists. By integrating it into hospital PACS workflows, this model could enhance diagnostic efficiency, reduce reporting time, and guide MRI utilization for suspicious cases, ultimately optimizing patient management and improving clinical outcomes.

AVAILABILITY OF DATA AND MATERIALS

The authors declare that all data supporting the findings of this study are available within the paper and any raw data can be obtained from the corresponding author upon request.

AUTHOR CONTRIBUTIONS

NS—designed the study and carried them out. NS, XLY, LLF—supervised the data collection; analyzed the data; interpreted the data. NS, NZ, YX—prepared the manuscript for publication and reviewed the draft of the manuscript. All authors have read and approved the manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical approval was obtained from the Ethics Committee of Shanxi Bethune Hospital (Approval No. 2025-356). Written informed consent was obtained from legally authorized representatives for anonymized patient information to be published in this article.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Tjokropawiro BA, Novitasari K, Ulhaq RA, Sulistya HA. Clinicopathological analysis of giant ovarian tumors. *European Journal of Obstetrics & Gynecology and Reproductive Biology*. X. 2024; 22: 100318.
- [2] Wang Y, Wang Z, Zhang Z, Wang H, Peng J, Hong L. Burden of ovarian cancer in China from 1990 to 2030: a systematic analysis and comparison with the global level. *Frontiers in Public Health*. 2023; 11: 1136596.
- [3] 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. *International Journal of Gynecological Cancer*. 2021; 31: 961–982.
- [4] Zhou Y, Duan Y, Zhu Q, Li S, Liu X, Cheng T, *et al.* Integrative deep learning and radiomics analysis for ovarian tumor classification and diagnosis: a multicenter large-sample comparative study. *Radiologia Medica*. 2025; 130: 889–904.
- [5] Zhao B, Wen L, Huang Y, Fu Y, Zhou S, Liu J, *et al.* A deep learning-based automatic recognition model for polycystic ovary ultrasound images. *Balkan Medical Journal*. 2025; 42: 419–428.
- [6] Zeng S, Jia H, Zhang H, Feng X, Dong M, Lin L, *et al.* Multimodal ultrasound-based radiomics and deep learning for differential diagnosis of O-RADS 4–5 adnexal masses. *Cancer Imaging*. 2025; 25: 64.
- [7] Maniaci A, La Via L, Lavalle S, Lentini M, Pavone P, Iannella G, *et al.* Presentation, radiologic features, and treatment options of congenital tongue tumors: a comprehensive review. *Annali Italiani di Chirurgia*. 2024; 95: 481–496.
- [8] Yang R, Zou Y, Li L, Liu WV, Liu C, Wen Z, *et al.* Enhancing repeatability of follicle counting with deep learning reconstruction high-resolution MRI in PCOS patients. *Scientific Reports*. 2025; 15: 1241.
- [9] Wang X, Quan T, Chu X, Gao M, Zhang Y, Chen Y, *et al.* Deep learning radiomics nomogram based on MRI for differentiating between borderline ovarian tumors and stage I ovarian cancer: a multicenter study. *Academic Radiology*. 2025; 32: 3485–3497.
- [10] Hsu WC, Wang Y, Wu YF, Chen R, Afyouni S, Liu J, *et al.* MRI-based ovarian lesion classification via a foundation segmentation model

- and multimodal analysis: a multicenter study. *Radiology*. 2025; 316: e243412.
- [11] Altinsoy E, Bakirarar B, Culcu S. Machine learning in predicting gastric cancer survival Presenting a novel decision support system model. *Annali Italiani di Chirurgia*. 2023; 94: 631–638.
 - [12] Yin R, Guo Y, Wang Y, Zhang Q, Dou Z, Wang Y, *et al*. Predicting neoadjuvant chemotherapy response and high-grade serous ovarian cancer from CT images in ovarian cancer with multitask deep learning: a multicenter study. *Academic Radiology*. 2023; 30: S192–S201.
 - [13] Yin R, Dou Z, Wang Y, Zhang Q, Guo Y, Wang Y, *et al*. Preoperative CECT-based multitask model predicts peritoneal recurrence and disease-free survival in advanced ovarian cancer: a multicenter study. *Academic Radiology*. 2024; 31: 4488–4498.
 - [14] Akazawa M, Hashimoto K. Artificial intelligence in gynecologic cancers: current status and future challenges—a systematic review. *Artificial Intelligence in Medicine*. 2021; 120: 102164.
 - [15] Xu HL, Gong TT, Liu FH, Chen HY, Xiao Q, Hou Y, *et al*. Artificial intelligence performance in image-based ovarian cancer identification: a systematic review and meta-analysis. *eClinicalMedicine*. 2022; 53: 101662.
 - [16] Du Y, Wang T, Qu L, Li H, Guo Q, Wang H, *et al*. Preoperative molecular subtype classification prediction of ovarian cancer based on multi-parametric magnetic resonance imaging multi-sequence feature fusion network. *Bioengineering*. 2024; 11: 472.
 - [17] Elguoshy A, Zedan H, Saito S. Machine learning-driven insights in cancer metabolomics: from subtyping to biomarker discovery and prognostic modeling. *Metabolites*. 2025; 15: 514.
 - [18] Geysels A, Garofalo G, Timmerman S, Barrenada L, De Moor B, Timmerman D, *et al*. Artificial intelligence applied to ultrasound diagnosis of pelvic gynecological tumors: a systematic review and meta-analysis. *Gynecologic and Obstetric Investigation*. 2025. PMID: 40340944; PMCID: PMC12180770.
 - [19] Du Y, Xiao Y, Guo W, Yao J, Lan T, Li S, *et al*. Development and validation of an ultrasound-based deep learning radiomics nomogram for predicting the malignant risk of ovarian tumours. *BioMedical Engineering OnLine*. 2024; 23: 41.
 - [20] Bogaerts JM, Steenbeek MP, Bokhorst JM, van Bommel MH, Abete L, Addante F, *et al*. Assessing the impact of deep-learning assistance on the histopathological diagnosis of serous tubal intraepithelial carcinoma (STIC) in fallopian tubes. *The Journal of Pathology: Clinical Research*. 2024; 10: e70006.
 - [21] Dai WL, Wu YN, Ling YT, Zhao J, Zhang S, Gu ZW, *et al*. Development and validation of a deep learning pipeline to diagnose ovarian masses using ultrasound screening: a retrospective multicenter study. *eClinicalMedicine*. 2024; 78: 102923.
 - [22] Du Y, Guo W, Xiao Y, Chen H, Yao J, Wu J. Ultrasound-based deep learning radiomics model for differentiating benign, borderline, and malignant ovarian tumours: a multi-class classification exploratory study. *BMC Medical Imaging*. 2024; 24: 89.
 - [23] El-Latif EIA, El-Dosuky M, Darwish A, Hassanien AE. A deep learning approach for ovarian cancer detection and classification based on fuzzy deep learning. *Scientific Reports*. 2024; 14: 26463.
 - [24] He D, Jin L, Geng H, Cao L. Deep learning-based analysis of gross features for ovarian epithelial tumors classification: a tool to assist pathologists for frozen section sampling. *Human Pathology*. 2025; 157: 105762.
 - [25] Hou B, Lee S, Lee JM, Koh C, Xiao J, Pickhardt PJ, *et al*. Deep learning segmentation of ascites on abdominal CT scans for automatic volume quantification. *Radiology: Artificial Intelligence*. 2024; 6: e230601.
 - [26] Bhuvaneshwari KV, Lahza H, Sreenivasa BR, Lahza HFM, Shawly T, Poornima B. Optimising ovarian tumor classification using a novel CT sequence selection algorithm. *Scientific Reports*. 2024; 14: 25010.
 - [27] Giourga M, Petropoulos I, Stavros S, Potiris A, Gerede A, Sapantzoglu I, *et al*. Enhancing ovarian tumor diagnosis: performance of convolutional neural networks in classifying ovarian masses using ultrasound images. *Journal of Clinical Medicine*. 2024; 13: 4123.
 - [28] Ma L, Gao W, Hu X, Zhou D, Wang C, Yu J, *et al*. An improved cancer diagnosis algorithm for protein mass spectrometry based on PCA and a one-dimensional neural network combining ResNet and SENet. *Analyst*. 2024; 149: 5675–5683.
 - [29] He X, Bai XH, Chen H, Feng WW. Machine learning models in evaluating the malignancy risk of ovarian tumors: a comparative study. *Journal of Ovarian Research*. 2024; 17: 219.
 - [30] Jung Y, Kim T, Han MR, Kim S, Kim G, Lee S, *et al*. Ovarian tumor diagnosis using deep convolutional neural networks and a denoising convolutional autoencoder. *Scientific Reports*. 2022; 12: 17024.
 - [31] Saida T, Mori K, Hoshiai S, Sakai M, Urushibara A, Ishiguro T, *et al*. Diagnosing ovarian cancer on MRI: a preliminary study comparing deep learning and radiologist assessments. *Cancers*. 2022; 14: 987.
 - [32] Yang Q, Zhang H, Ma PQ, Peng B, Yin GT, Zhang NN, *et al*. Value of ultrasound and magnetic resonance imaging combined with tumor markers in the diagnosis of ovarian tumors. *World Journal of Clinical Cases*. 2023; 11: 7553–7561.
 - [33] Mitchell S, Gleeson J, Tiwari M, Bailey F, Gaughran J, Mehra G, *et al*. Accuracy of ultrasound, magnetic resonance imaging and intraoperative frozen section in the diagnosis of ovarian tumours: data from a London tertiary centre. *BJC Reports*. 2024; 2: 50.
 - [34] Wang WH, Zheng CB, Gao JN, Ren SS, Nie GY, Li ZQ. Systematic review and meta-analysis of imaging differential diagnosis of benign and malignant ovarian tumors. *Gland Surgery*. 2022; 11: 330–340.
 - [35] Cui L, Xu H, Zhang Y. Diagnostic accuracies of the ultrasound and magnetic resonance imaging ADNEX scoring systems for ovarian adnexal mass: systematic review and meta-analysis. *Academic Radiology*. 2022; 29: 897–908.
 - [36] Hu Y, Chen B, Dong H, Sheng B, Xiao Z, Li J, *et al*. Comparison of ultrasound-based ADNEX model with magnetic resonance imaging for discriminating adnexal masses: a multi-center study. *Frontiers in Oncology*. 2023; 13: 1101297.
 - [37] Adusumilli P, Ravikumar N, Hall G, Scarsbrook AF. A methodological framework for AI-assisted diagnosis of ovarian masses using CT and MR imaging. *Journal of Personalized Medicine*. 2025; 15: 76.

How to cite this article: Ning Sun, Xiangli Yang, Lili Fan, Nan Zhang, Yan Xue. Research on the diagnostic value of ultrasound combined with MRI features based on deep learning in distinguishing benign and malignant ovarian tumors. *European Journal of Gynaecological Oncology*. 2025; 46(12): 57-65. doi: 10.22514/ejgo.2025.146.