Cite as: Mao XX, Gu LP, Li S, Shi L, Zhong LC. Deep learning visualization model for assisting junior sonographers in differentiating benign and malignant breast nodules [J]. Chin J Clin Res, 2026, 39(4):547-552.
DOI: 10.13429/j.cnki.cjcr.2026.04.012

Deep learning visualization model for assisting junior sonographers in differentiating benign and malignant breast nodules

MAO Xiaoxian*, GU Liping, LI Shuo, SHI Lin, ZHONG Lichang
**Shanghai Ocean University, Shanghai 201306, China*
Corresponding author: GU Liping, E-mail: guliping666@126.com

Abstract: **Objective** To evaluate the diagnostic value of a deep learning visualization model based on breast ultrasound images in assisting junior sonographers to differentiate benign and malignant breast nodules. **Methods** The clinical data and ultrasound images from 420 pathologically confirmed breast nodule cases identified through screening at Shanghai Sixth People s Hospital from September 2022 to September 2024 were collected. Patients were randomly divided into a training set (*n*=294) and an independent test set (*n*=126) in a 7 to 3 ratio. Six end-to-end deep learning models (VGG16, GoogleNet, AlexNet, ResNet50, Vision Transformer, DenseNet121) were constructed and compared using the training set images. Performance was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC) , accuracy, sensitivity, specificity, precision, and F1 value (harmonic average of accuracy and recall). The optimal model was selected and used to generate visualizations. Employing a self- controlled before- after design, a junior sonographer independently interpreted the test set images blinded to pathological results, assigning Breast Imaging Reporting and Data System (BI-RADS) categories. After a two-week washout period, the same junior sonographer re-interpreted the same images with the assistance of the ResNet50 visualization model. Diagnostic performance metrics were calculated and compared for the junior sonographer under independent and model- assisted reading conditions. **Results** The ResNet50 model demonstrated optimal performance on the training set, achieving an AUC of 0.937 (95%*CI*: 0.911-0.962) , with accuracy, sensitivity, specificity, precision, and F1 values of 0.864, 0.863, 0.865, 0.863 and 0.863, respectively. The core diagnostic performance metrics of the model surpassed those of the participating sonographer. With the assistance of ResNet50 visualization heatmaps, the AUC, accuracy, sensitivity, and specificity of junior sonographer in the second reading of breast nodule diagnosis increased from the initial 0.597, 59.5%, 80.8%, 38.5% to 0.904, 90.5%, 84.9%, and 95.9%, respectively. **Conclusion** Deep learning visualization models based on breast ultrasound images, such as ResNet50, exhibit strong capability in differentiating benign and malignant breast nodules. These models effectively assist junior sonographers by improving diagnostic sensitivity and overall performance, serving as valuable auxiliary tools to enhance standardized diagnostic skills and potentially shorten training cycles.

Keywords: Breast nodules; Ultrasonography; Differential diagnosis; Deep learning; Human - machine comparison; Visualization

Breast cancer is one of the most common malignant tumors threatening women's health [1]. According to global cancer statistics in 2022, breast cancer accounted for 11.6% of all newly diagnosed malignant tumors worldwide, ranking second only to lung cancer in overall incidence [2]. It remains the leading cause of cancer-related mortality among women aged 20 to 59 years [3]. Early screening and diagnosis are critical for optimizing long-term prognosis.

Given its non-invasiveness, wide accessibility and cost-effectiveness, ultrasonography is strongly recommended as a core imaging modality for population-based breast cancer screening [4-5]. Standardized diagnosis and treatment for early breast cancer can significantly elevate the 5-year survival rate (>90%) and improve patients' quality of life [6].

The Breast Imaging Reporting and Data System (BI-RADS), established by the American College of Radiology (ACR), provides a standardized framework for the clinical management of breast lesions [7]. Nevertheless, substantial inter-observer and intra-observer variability widely exists in practical BI-RADS application [8]. Breast ultrasonography diagnosis is highly operator-dependent and limited by the subjectivity of image interpretation [9].

Multiple clinical studies have confirmed that junior radiologists exhibit significantly lower diagnostic accuracy, sensitivity and specificity compared with senior physicians [10-12]. Such discrepancies are particularly prominent in the differential diagnosis of BI-RADS category 3–4 lesions. These lesions carry an intermediate malignant risk (category 3: 0–2%; category 4A: 2%–10%).

Diagnostic uncertainty regarding their benign or malignant nature frequently leads to overtreatment or delayed intervention [13-14], highlighting an urgent demand for objective and precise auxiliary tools to refine clinical decision-making.

Deep learning, as a core branch of artificial intelligence, simulates human cognitive mechanisms via multi-layer neural networks and autonomously extracts hierarchical high-dimensional features from massive datasets, demonstrating powerful capabilities in pattern recognition and classification within medical image analysis [15]. To address the "black-box" limitation of conventional deep learning models, visual interpretability

techniques, such as gradient-weighted class activation mapping (Grad-CAM) and saliency mapping, have been developed. These visualized approaches intuitively highlight the key imaging regions contributing to model decision-making [10], thereby greatly enhancing model interpretability and clinical acceptability. In recent years, deep learning techniques have achieved breakthrough advances in breast ultrasound research, with validated feasibility in lesion detection, benign and malignant differentiation, and molecular subtype prediction [16-18]. Although existing studies have verified the diagnostic efficacy of deep learning in breast lesion classification, the clinical value of visualized deep learning models for assisting junior radiologists remains systematically unevaluated.

This study aimed to explore the optimizing effects of a breast ultrasound-based visualized deep learning model on the diagnostic performance of junior physicians through rigorous controlled human-machine comparative trials.

1 Materials and Methods

1.1 General Data

Female patients with breast nodules categorized as BI-RADS 3-4 on routine ultrasonography were retrospectively enrolled from September 2022 to September 2024 at Shanghai Sixth People's Hospital. Inclusion criteria were as follows: (1) Definitive pathological results obtained via surgical resection or core needle biopsy; (2) Breast ultrasonography performed within 2 weeks before biopsy or surgery, with clear and complete two-dimensional grayscale images of the target nodule, complete imaging reports and integrated clinical-pathological data; (3) Target lesions classified as BI-RADS 3, 4A, 4B or 4C in accordance with the 2nd edition of the ACR BI-RADS atlas [7]. Exclusion criteria included:(1) Ambiguous or undetermined pathological findings; (2) Excessively large tumors with incomplete image display; (3) Preoperative anti-tumor treatment including chemotherapy, radiotherapy or endocrine therapy; (4) Incomplete imaging or clinical information.

A total of 420 eligible patients were finally enrolled, with an age range of 18-88 years and a mean age of (48.67±14.90) years. Clinical baseline characteristics including age, tumor size and postoperative pathological results were collected. To ensure statistical independence of each evaluation, for patients with multiple synchronous BI-RADS 3-4 breast lesions, only the lesion with the highest BI-RADS category was defined as the target nodule. All enrolled patients were randomly divided into a training set (n=294) and a test set (n=126) at a 7:3 ratio. This retrospective study was approved by the Ethics Committee of Shanghai Sixth People's Hospital (No: 2019-027), and the requirement for written informed consent was waived.

1.2 Research Methods

1.2.1 Ultrasonographic Examination

Ultrasonographic examinations were performed using S2000 ultrasound systems (Siemens, Germany) and GE ultrasound scanners, with a consistent probe frequency of 4-9 MHz. All patients underwent routine breast ultrasonography in the supine position within 2 weeks before surgery. Transverse and longitudinal scanning was performed for each target breast nodule. After acquiring high-quality two-dimensional images, the maximum cross-sectional view of each lesion was captured and stored in DICOM format. Basic information including lesion size, location and ultrasonic characteristics was simultaneously documented.

1.2.2 Image Preprocessing and Deep Learning Model Construction

Qualified breast ultrasound images were converted from DICOM to JPEG format.Given variations in image resolution from different ultrasound devices, all images were uniformly resized to 1024×1024 pixels. The open-source annotation tool LabelMe (http://labelme.csail.mit.edu/Release3.0) was used for manual labeling based on pathological golden standards, with malignant lesions labeled as "1" and benign lesions labeled as "0", and corresponding JSON files were generated for model training. Multiple deep learning backbones were applied for feature extraction and lesion classification. Each network was equipped with 512 fully connected layers, 2 output neurons and a softmax activation function at the top convolutional layer to output artificial intelligence (AI) scores for benign and malignant differentiation. The softmax function guaranteed that the sum of AI scores for benign and malignant classification of a single lesion equaled 1. The receiver operating characteristic (ROC) curve was adopted to evaluate the diagnostic performance of the ultrasound-based deep learning model, and external validation was conducted in the test cohort. The area under the curve (AUC) and other quantitative indicators were compared to screen the optimal model for visual interpretation. Six mainstream deep learning backbones were compared, including GoogleNet, VGG16, AlexNet, ResNet50, Vision Transformer and DenseNet121. ResNet50 yielded the highest AUC value in validation cohorts and was ultimately selected as the feature extraction backbone for model construction.

1.2.3 Diagnostic Evaluation by Junior Radiologists

All breast lesions were independently assessed and graded according to standardized ACR BI-RADS criteria [7] by a single junior radiologist for initial benign and malignant differentiation. Binary diagnostic classification was defined as follows: BI-RADS 3/4A lesions were classified as benign, while BI-RADS 4B/4C lesions were classified as malignant. After a 2-week washout period, the same junior physician re-evaluated all images with the assistance of the AI visualized auxiliary diagnosis system. Lesion localization, morphological analysis, total lesion

number, anatomical location and maximum diameter were recorded in the second evaluation. Diagnostic results from the two independent reading sessions were summarized for subsequent comparative analysis.

1.3 Statistical Analysis

All statistical analyses were performed using R software and SPSS 26.0, with pathological results regarded as the gold standard. Normality and homogeneity of variance tests were conducted for quantitative data. Normally distributed continuous variables were expressed as mean±standard deviation, and intergroup differences were compared using the independent sample *t*-test. Categorical variables were presented as case numbers and percentages, with the chi-square test for intergroup comparison. The DeLong test (MedCalc 20.01) was used for pairwise comparison of AUC values between different models. The Z-test was applied to compare ROC-derived AUC indicators. Intraclass correlation coefficient (ICC), Spearman rank correlation analysis and Z-score normalization were completed using Python 3.10. A two-tailed $P<0.05$ was considered statistically significant.

2 Results

2.1 Baseline Characteristic Comparison

A total of 420 patients with breast nodules were enrolled in this study, including 281 cases of BI-RADS 3/4A lesions and 139 cases of BI-RADS 4B/4C lesions. Postoperative pathological confirmation identified 209 benign lesions and 211 malignant lesions. No significant differences were observed in age, lesion location, nodule diameter, BI-RADS classification or pathological distribution between the training set and test set ($P>0.05$) **(Table 1).**

2.2 Diagnostic Performance Improvement Assisted by the Visualized Deep Learning Model

To select the optimal model for breast lesion differentiation, the diagnostic performance of six deep learning architectures was comprehensively compared (**Table 2**). ResNet50 achieved superior performance in terms of AUC, accuracy (except for the DenseNet121 training set) and sensitivity. The AUC values of the ResNet50 model reached 0.937 (95%CI: 0.911–0.962) in the training set and 0.891 (95%CI: 0.833–0.948) in the test set (**Figure 1**). Therefore, ResNet50 was determined as the final backbone architecture, and multi-form visual heatmaps were generated for auxiliary radiological reading (**Figure 2**).

Tab.1 Comparison of general data between two sets [case (%)]

Item	Training set (*n*=294)	Test set (*n*=126)	t/χ^2 value	*P* value
Age (years)[a]	49.18±14.65	47.49±15.48	1.065	0.287
Nodule diameter (mm)[a]	18.52±8.68	20.17±12.84	1.319	0.189
Nodule location			2.13	0.144
Left	122(41.50)	62(49.21)		
Right	172(58.50)	64(50.79)		
BI-RADS classification			0.005	0.946
Category 3, 4A	197(67.01)	84(66.67)		
Category 4B, 4C	97(32.99)	42(33.33)		
Pathological results			0.131	0.717
Benign	148(50.34)	61(48.41)		
Malignant	146(49.66)	65(51.59)		

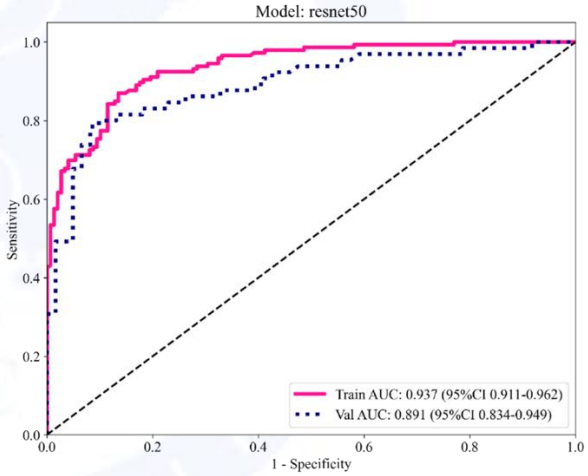


Fig.1 ROC curves of training set and test set

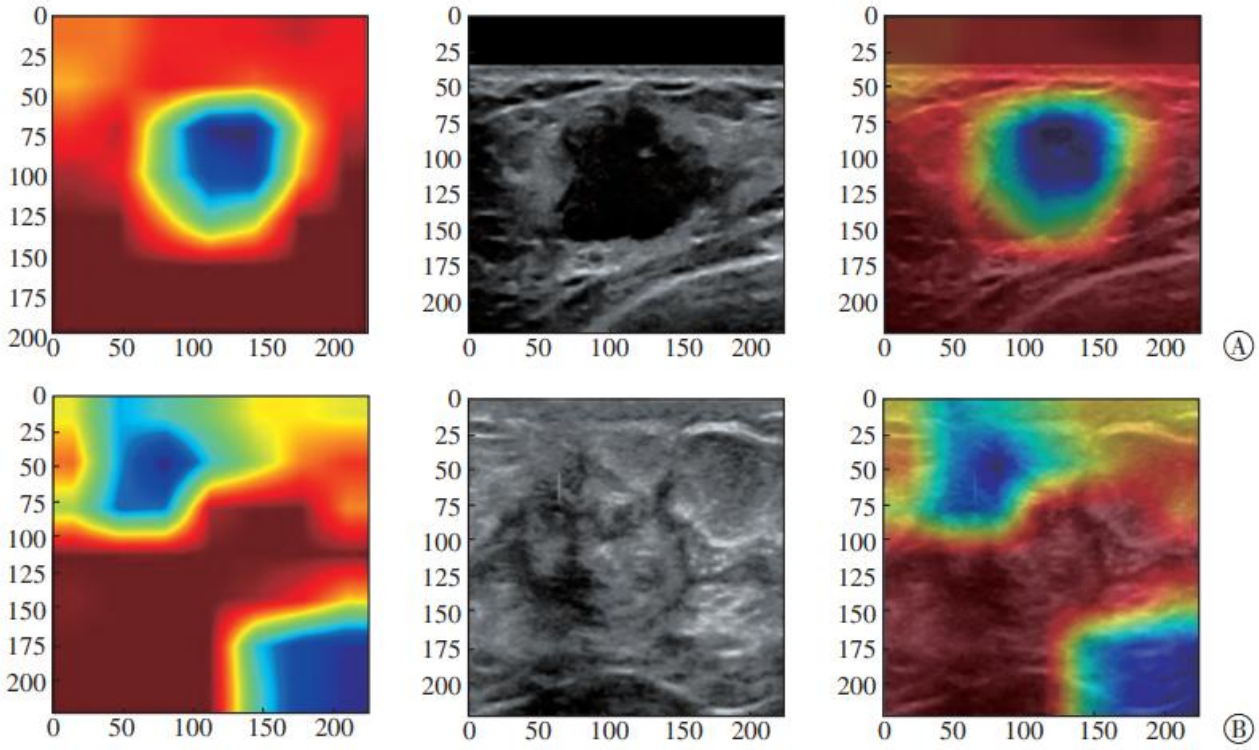
Tab.2 Diagnostic efficiency of different deep learning models

Deep learning model	Group	AUC (95%CI)	Accuracy	Sensitivity	Specificity	Precision	F1-score
VGG16	Training set	0.839(0.794~0.884)	0.786	0.733	0.838	0.817	0.773
	Test set	0.812(0.735~0.888)	0.77	0.754	0.787	0.79	0.772
Google Net	Training set	0.805(0.754~0.855)	0.752	0.692	0.811	0.783	0.735
	Test set	0.825(0.750~0.899)	0.77	0.662	0.885	0.86	0.748
AlexNet	Training set	0.760(0.705~0.814)	0.704	0.562	0.845	0.781	0.653
	Test set	0.789(0.709~0.868)	0.754	0.646	0.869	0.84	0.73
ResNet50	Training set	0.937(0.911~0.962)	0.864	0.863	0.865	0.863	0.863
	Test set	0.891(0.833~0.948)	0.841	0.769	0.918	0.909	0.833
Vision Transformer	Training set	0.562(0.496~0.628)	0.582	0.253	0.905	0.725	0.376
	Test set	0.631(0.533~0.727)	0.603	0.338	0.885	0.759	0.468
Dense Net121	Training set	0.933(0.905~0.960)	0.867	0.788	0.946	0.935	0.855
	Test set	0.879(0.816~0.941)	0.833	0.815	0.852	0.855	0.835

Compared with initial independent reading, the visualized deep learning model significantly improved comprehensive diagnostic efficiency. In the training set, the overall accuracy, sensitivity and specificity reached 0.864, 0.863 and 0.865, respectively. Corresponding indicators in the test set were 0.841, 0.769 and 0.918. Further comparative analysis demonstrated that with visualized AI assistance, the junior physician achieved significantly higher accuracy, sensitivity, specificity, precision and F1-score in breast lesion differentiation than initial unaided reading, with statistically significant differences (P<0.05) (**Table 3**).

Tab.3 Human-machine comparison before and after AI assistance

Category	Accuracy	AUC (95%CI)	Sensitivity	Specificity	Precision	F1-score
Deep learning visualization model	0.864	0.937(0.911~0.962)	0.863	0.865	0.863	0.863
Junior physicians	0.595	0.597(0.545~0.647)	0.808	0.385	0.565	0.665
Junior physicians assisted by deep learning visualization model	0.905	0.904(0.871~0.937)	0.849	0.959	0.965	0.899



Note: A is the visualization heatmap of malignant nodules; B is the visualization heatmap of benign nodules.

Fig.2 Visualization heatmap of ResNet50 model

3 Discussion

Based on ultrasound images from 420 patients with breast nodules, this study compared the differential performance of multiple deep learning models and validated the clinical value of a ResNet50-based visualized interpretable system in improving the diagnostic capacity of junior radiologists. The ResNet50 model exhibited robust and stable diagnostic efficacy in both training and test cohorts. With heatmap-based visual assistance, the overall diagnostic accuracy of the junior physician increased remarkably from 59.5% to 90.5%, accompanied by synchronous improvements in sensitivity, specificity, precision and F1-score. These findings not only validate the tremendous potential of deep learning in breast imaging diagnosis but also provide clinical evidence for human-machine collaborative diagnostic strategies, which is of great public health significance for optimizing grassroots breast cancer screening workflows [19].

Common clinical imaging modalities for breast cancer screening and diagnosis include magnetic resonance imaging (MRI), ultrasonography and mammography [20]. Mammography is recognized as a standardized global screening tool with high sensitivity for early microcalcification. However, increased breast density can reduce its sensitivity from 85% to 47.8%–64.4% [21]. Breast ultrasonography features non-radiation exposure and low cost, making it particularly suitable for women with dense breasts and young populations [22]. It presents high detection rates for solid masses, facilitates the differentiation of cystic and solid lesions, and enables real-time interventional puncture guidance. Nevertheless, its limited imaging resolution and operator-dependent characteristics remain major limitations [23]. Breast MRI offers excellent sensitivity for multifocal, multicentric and high-risk

breast lesions and can detect occult tumors invisible on mammography and ultrasonography [6], whereas its widespread application is restricted by high costs, prolonged examination duration and relatively high false-positive rates. The inherent limitations of individual imaging modalities have driven the rapid development of artificial intelligence and deep learning in breast radiology.

Compared with manual image interpretation, deep learning models can analyze imaging details objectively and consistently, and capture subtle abnormal morphological patterns imperceptible to the naked eye. AI-assisted diagnostic systems are expected to serve as an independent "second reader" [24] for lesion review and screening sensitivity enhancement.

In recent years, convolutional neural networks (CNNs) have been widely applied in medical image classification, target detection and semantic segmentation [25]. Accumulating evidence has demonstrated that CNN-based models achieve diagnostic performance comparable to experienced senior radiologists in lesion detection and segmentation based on mammography, ultrasonography and breast MRI [26-27]. Wu *et al.* [28] constructed a logistic regression model combining multi-modal radiomic features from MRI, ultrasound and mammography with a 5-mm peritumoral region, yielding an AUC of 0.905 in the test cohort. The ResNet50 visualized model established in the present study using two-dimensional ultrasound images achieved comparable or superior predictive performance. Artificial intelligence not only acts as an independent diagnostic tool but also effectively elevates the diagnostic level of less experienced physicians. A prospective clinical study focusing on the S-Detect system enrolled 216 patients and reported that the AI system reached a specificity of 92.7% and an accuracy of 90.4% for breast nodule evaluation. With AI assistance, the diagnostic sensitivity of junior radiologists increased from 85.0% to 93.3%, accuracy improved from 70.4% to 83.9%, and AUC elevated from 0.698 to 0.835, markedly narrowing the diagnostic gap between physicians with different clinical experience [29]. Consistent with previous research, our study revealed that visualized ResNet50 assistance increased the junior physician's accuracy from 59.5% to 90.5%, sensitivity from 80.8% to 84.9%, and specificity from 38.5% to 95.9%. The visualized deep learning system can automatically identify key ultrasonic features of breast lesions and provide intuitive heatmap prompts, enabling junior physicians to recognize pathological imaging characteristics, improve the identification of subtle lesions and ambiguous images, and enhance diagnostic sensitivity and specificity [30], thereby shortening the diagnostic disparity between junior and senior radiologists [31]. The "black-box" drawback of conventional deep learning models hinders their clinical translation. Various interpretable visualization strategies have been proposed to address this issue. Grad-CAM visualization technology has been proven to highlight critical perilesion structural features in RcdNet and EfficientNet-B7, ensuring that model-focused regions are consistent with human diagnostic basis [32-33]. Domain knowledge-based interpretable frameworks such as MUP-Net integrate B-mode and color Doppler ultrasound images, achieving an AUC of 0.902 and significantly improving diagnostic confidence among beginners [10].However, systematic reviews indicate that only a minority of studies have quantified the impact of interpretable AI models on clinical trust, and most existing evidence lacks comprehensive evaluation of data quality, model stability and clinical practicability [34]. Our study confirmed the auxiliary value of interpretability by verifying improved diagnostic accuracy, while quantitative assessment of physician trust and acceptance requires further prospective exploration in future research.

Several limitations of this study should be acknowledged. First, the relatively limited sample size of ultrasound images in the training and test sets may affect model stability and generalizability. Large-scale multicenter datasets will be included in further research to validate model robustness. Second, this is a single-center retrospective study. Well-designed prospective multicenter trials are essential to comprehensively evaluate the clinical application value of the visualized AI model.

Third, only two-dimensional grayscale ultrasound images were enrolled for model training, without incorporating three-dimensional imaging features, which may lead to the omission of partial tumor morphological information.

In conclusion, the two-dimensional ultrasound-based visualized deep learning model exhibits favorable efficacy in differentiating benign and malignant breast nodules.

As a reliable auxiliary tool, it effectively improves the diagnostic sensitivity and overall performance of junior radiologists, standardizes breast ultrasound interpretation, and shortens the professional training cycle for young physicians.

Conflict of Interest: None

Reference

[1] Bray F, Laversanne M, Weiderpass E, et al. The ever-increasing importance of cancer as a leading cause of premature death worldwide[J]. Cancer, 2021, 127(16): 3029-3030.

[2] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2024, 74 (3): 229-263.

[3] Pal R, Srivastava N, Chopra R, et al. Investigation of DNA da mage response and apoptotic gene methylation pattern in sporadic breast tumors using high throughput quantitative DNA methylation analysis technology[J]. Mol Cancer, 2010, 9(1): 303.

[4] Song PJ, Wang YL, Sun C, et al. A comparative study of AI diagnosis of ultrasound based on the real-world data and X-ray mammography examination for breast in the screening of early breast cancer[J]. China Med Equip, 2023, 20(9): 20-24.

[5] Zhou YP, Wei J, Wu DM, et al. Generating full-field digital mammogram from digitized screen-film mammogram for breast cancer screening with high-resolution generative adversarial network[J]. Front Oncol, 2022, 12: 868257.

[6] Zhou YY, Xie J, Zhang B, et al. Diagnostic values of conventional ultrasound, contrast-enhanced ultrasound, mammography, magnetic resonance imaging and their combinations in diagnosing patients with early breast cancer[J]. J Guizhou Med Univ, 2023, 48(2): 172-176.

[7] Maani N, Westergard S, Yang J, et al. NF1 patients receiving breast cancer screening: insights from the Ontario high risk breast screening

program[J]. Cancers, 2019, 11(5): 707.
[8] Ezeana CF, He TC, Patel TA, et al. A deep learning decision support tool to improve risk stratification and reduce unnecessary biop- sies in BI-RADS 4 mammograms[J]. Radiol Artif Intell, 2023, 5 (6): e220259.
[9] He P, Chen W, Bai MY, et al. Deep learning-based computer-aided diagnosis for breast lesion classification on ultrasound: a pro- spective multicenter study of radiologists without breast ultrasound expertise[J]. AJR Am J Roentgenol, 2023, 221(4): 450-459.
[10] Yan L, Liang ZY, Zhang H, et al. A domain knowledge-based in- terpretable deep learning system for improving clinical breast ultra- sound diagnosis[J]. Commun Med, 2024, 4(1): 90.
[11] Yao Z, Luo T, Dong YJ, et al. Virtual elastography ultrasound via generative adversarial network for breast cancer diagnosis[J]. Nat Commun, 2023, 14(1): 788.
[12] Wang J, Tian HT, Yang X, et al. Artificial intelligence in breast US diagnosis and report generation[J]. Radiol Artif Intell, 2025, 7 (4): e240625.
[13] Yang YT, Liao TT, Lin XH, et al. Reducing unnecessary biopsies of BI-RADS 4 lesions based on a deep learning model for mammog- raphy[J]. Front Oncol, 2025, 15: 1543553.
[14] MILA·Yeerlan, Zhang HJ, Hu ZH, et al. Value of ultrasound-based radiomics in identifying benign and malignant BI-RADS category 4a irregular breast lesions and reducing unnecessary biopsies[J]. J Mol Imag, 2023, 46(1): 12-20.
[15] Carriero A, Groenhoff L, Vologina E, et al. Deep learning in breast cancer imaging: state of the art and recent advancements in early 2024[J]. Diagnostics, 2024, 14(8): 848.
[16] Dembrower K, Crippa A, Colón E, et al. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study [J]. Lancet Digit Health, 2023, 5(10): e703-e711.
[17] Ng AY, Oberije CJG, Ambrózay É, et al. Prospective implementa- tion of AI-assisted screen reading to improve early detection of breast cancer [J]. Nat Med, 2023, 29(12): 3044-3049.
[18] Eisemann N, Bunk S, Mukama T, et al. Nationwide real-world implementation of AI for cancer detection in population-based mam- mography screening[J]. Nat Med, 2025, 31(3): 917-924.
[19] Wu TH, Liu XY, Si ZY, et al. Combination of ultrasound-based ra- diomics and deep learning with clinical data to predict response in breast cancer patients treated with neoadjuvant chemotherapy[J]. Front Oncol, 2025, 15: 1525285.
[20] Committee of Breast Cancer Society, China Anti-Cancer Association; Breast Oncology Group, Chinese Society of Oncology, Chinese Medical Association, Shao ZM. Guidelines for breast cancer diagnosis and treatment by China Anti-cancer Association(2024 edition)[J]. China Oncol, 2023, 33(12): 1092-1186.
[21] Thigpen D, Kappler A, Brem R. The role of ultrasound in screen-ing dense breasts-a review of the literature and practical solutions for implementation[J]. Diagnostics, 2018, 8(1): 20.
[22] Wu CW, Zou DZ. Diagnostic value of ultrasonography and X-ray molybdenum target in breast cancer and its correlation with Mo-lecular subtypes[J]. J Med Theory Pract, 2023, 36(24): 4149-4153.
[23] Vimala BB, Srinivasan S, Mathivanan SK, et al. Image noise re-moval in ultrasound breast images based on hybrid deep learning technique[J]. Sensors, 2023, 23(3): 1167.
[24] Frazer HML, Peña-Solorzano CA, Kwok CF, et al. Comparison of AI-integrated pathways with human-AI interaction in population mammographic screening for breast cancer[J]. Nat Commun, 2024, 15(1): 7525.
[25] Takahashi S, Sakaguchi Y, Kouno N, et al. Comparison of vision transformers and convolutional neural networks in medical image analysis: a systematic review[J]. J Med Syst, 2024, 48(1): 84.
[26] Luo LY, Wu MX, Li M, et al. A large model for non-invasive and personalized management of breast cancer from multiparametric MRI [J]. Nat Commun, 2025, 16(1): 3647.
[27] Iima M, Mizuno R, Kataoka M, et al. Deep learning applied to dif-fusion-weighted imaging for differentiating malignant from benign breast tumors without lesion segmentation[J]. Radiol Artif Intell, 2025, 7(1): e240206.
[28] Wu J, Li YX, Gong WQ, et al. Multi-modality radiomics diagnosis of breast cancer based on MRI, ultrasound and mammography[J]. BMC Med Imaging, 2025, 25(1): 265.
[29] Singla V, Garg D, Negi S, et al. Deep learning powered breast ul-trasound to improve characterization of breast masses: a prospec-tive study[J]. Acta Radiol, 2026, 67(1): 13-23.
[30]Yang L, Zhang NW, Jia JY, et al. Deep learning radiomics on gray-scale ultrasound images assists in diagnosing benign and malignant of BI-RADS 4 lesions[J]. Sci Rep, 2024, 14(1): 31479.
[31] Xiang HL, Wang X, Xu M, et al. Deep learning-assisted diagnosis of breast lesions on US images: a multivendor, multicenter study [J]. Radiol Artif Intell, 2023, 5(5): e220185.
[32] Liu BQ, Liu SY, Cao ZJ, et al. Accurate classification of benign and malignant breast tumors in ultrasound imaging with an en-hanced deep learning model[J]. Front Bioeng Biotechnol, 2025, 13: 1526260.
[33] Latha M, Kumar PS, Chandrika RR, et al. Revolutionizing breast ultrasound diagnostics with EfficientNet-B7 and Explainable AI [J]. BMC Med Imaging, 2024, 24(1): 230.
[34] Gurmessa DK, Jimma W. Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: a sys-tematic review[J]. BMJ Health Care Inform, 2024, 31(1): e100954.

Submission Received: 2025-11-03 Revised: 2026-01-12

·乳腺癌专题·论著·

深度学习可视化模型辅助低年资超声医师鉴别良恶性乳腺结节

毛晓娴[1,2]， 谷丽萍[2]， 李硕[1,2]， 师琳[2]， 钟李长[2]

1. 上海海洋大学，上海 201306；2. 上海交通大学医学院附属第六人民医院超声医学科，上海 200233

摘要：目的 探讨基于乳腺超声图像的深度学习可视化模型对低年资医师鉴别乳腺结节良恶性的辅助诊断价值。**方法** 回顾性收集2022年9月至2024年9月于上海市第六人民医院经超声筛查发现并经病理证实的乳腺结节病例（共420例）的临床资料及超声图像。按7∶3比例随机划分为训练集（*n*=294）与独立测试集（*n*=126）。基于训练集图像，构建并比较了VGG16、GoogleNet、AlexNet、ResNet50、Vision Transformer、DenseNet121六种端到端深度学习模型的性能，通过受试者工作特征（ROC）曲线下面积（AUC）、准确率、灵敏度、特异度、精确率、F1值（精确率和召回率的调和平均值）等指标筛选出最优模型并生成可视化结果。采用自身前后对照设计，由低年资医师在不知晓病理结果的情况下，独立对测试集图像进行诊断[基于乳腺影像报告与数据系统（BI-RADS）分类]；间隔2周后，同一低年资医师借助筛选出的最优可视化模型对相同图像进行辅助诊断。计算并比较低年资医师在独立诊断与模型辅助诊断模式下的诊断效能指标。**结果** ResNet50模型在训练集上表现最优，其AUC为0.937（95%*CI*：0.911~0.962），准确率、灵敏度、特异度、精确率、F1值分别为0.864、0.863、0.865、0.863、0.863。该模型的核心诊断效能指标优于参与研究的超声医师。低年资超声医师在ResNet50可视化热图辅助下，二次阅片对乳腺结节诊断的AUC、准确率、灵敏度、特异度分别由首次的0.597、59.5%、80.8%、38.5%，提高至0.904、90.5%、84.9%和95.9%。**结论** 基于乳腺超声的深度学习可视化模型（如ResNet50）具有良好的乳腺结节良恶性鉴别能力。该模型能有效辅助低年资医师提高诊断敏感性及整体效能，可作为提升低年资医师乳腺超声诊断规范化水平、缩短培训周期的有价值的辅助工具。

关键词： 乳腺结节；超声；鉴别诊断；深度学习；人机对比；可视化

中图分类号： R737.9 R445.1 **文献标识码：** A **文章编号：** 1674-8182(2026)04-0547-06

Deep learning visualization model for assisting junior sonographers in differentiating benign and malignant breast nodules

MAO Xiaoxian*, GU Liping, LI Shuo, SHI Lin, ZHONG Lichang

Shanghai Ocean University, Shanghai 201306, China

Corresponding author: GU Liping, E-mail: guliping666@126.com

Abstract: Objective To evaluate the diagnostic value of a deep learning visualization model based on breast ultrasound images in assisting junior sonographers to differentiate benign and malignant breast nodules. **Methods** The clinical data and ultrasound images from 420 pathologically confirmed breast nodule cases identified through screening at Shanghai Sixth People's Hospital from September 2022 to September 2024 were collected. Patients were randomly divided into a training set (*n*=294) and an independent test set (*n*=126) in a 7 to 3 ratio. Six end-to-end deep learning models (VGG16, GoogleNet, AlexNet, ResNet50, Vision Transformer, DenseNet121) were constructed and compared using the training set images. Performance was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC), accuracy, sensitivity, specificity, precision, and F1 value (harmonic average of accuracy and recall). The optimal model was selected and used to generate visualizations. Employing a self-controlled before-after

DOI：10.13429/j.cnki.cjcr.2026.04.012

通信作者： 谷丽萍，E-mail：guliping666@126.com

出版日期： 2026-04-20

QR code for English version

design, a junior sonographer independently interpreted the test set images blinded to pathological results, assigning Breast Imaging Reporting and Data System (BI-RADS) categories. After a two-week washout period, the same junior sonographer re-interpreted the same images with the assistance of the ResNet50 visualization model. Diagnostic performance metrics were calculated and compared for the junior sonographer under independent and model-assisted reading conditions. **Results** The ResNet50 model demonstrated optimal performance on the training set, achieving an AUC of 0.937 (95% *CI*: 0.911-0.962), with accuracy, sensitivity, specificity, precision, and F1 values of 0.864, 0.863, 0.865, 0.863 and 0.863, respectively. The core diagnostic performance metrics of the model surpassed those of the participating sonographer. With the assistance of ResNet50 visualization heatmaps, the AUC, accuracy, sensitivity, and specificity of junior sonographer in the second reading of breast nodule diagnosis increased from the initial 0.597, 59.5%, 80.8%, 38.5% to 0.904, 90.5%, 84.9%, and 95.9%, respectively. **Conclusion** Deep learning visualization models based on breast ultrasound images, such as ResNet50, exhibit strong capability in differentiating benign and malignant breast nodules. These models effectively assist junior sonographers by improving diagnostic sensitivity and overall performance, serving as valuable auxiliary tools to enhance standardized diagnostic skills and potentially shorten training cycles.

Keywords: Breast nodules; Ultrasonography; Differential diagnosis; Deep learning; Human-machine comparison; Visualization

乳腺癌是威胁女性健康的恶性肿瘤之一[1]，据2022年的数据显示，乳腺癌占当年全球全部新发恶性肿瘤的11.6%，在发病数上仅次于肺癌[2]。在20~59岁女性群体中乳腺癌是恶性肿瘤死亡的首要原因[3]。早期筛查与诊断是改善预后的关键，超声检查因其无创性、可及性和高性价比，被广泛推荐为乳腺癌筛查的核心影像学手段[4-5]。研究表明，早期乳腺癌的规范化诊疗可显著提升患者5年生存率(>90%)并改善生活质量[6]。

美国放射学会(American College of Radiology, ACR)制定的乳腺影像报告与数据系统(Breast Imaging Reporting and Data System, BI-RADS)为乳腺结节的临床管理提供了标准化框架[7]，然而，该系统的应用存在显著的观察者间差异与观察者内差异[8]。乳腺超声诊断高度依赖操作者经验，受限于图像解读的主观性和操作者依赖性[9]。多项研究证实，低年资医师的诊断准确率、灵敏度及特异度均显著低于高年资医师[10-12]。这种差距在BI-RADS 3~4类结节的鉴别诊断中尤为突出。此类结节具有中度恶性风险(3类：0~2%；4A类：2%~10%)，其良恶性判断的不确定性常导致过度诊疗或延迟诊断[13-14]，亟需开发客观、精准的辅助工具以优化临床决策。

深度学习作为人工智能的核心分支，通过构建多层神经网络模拟人脑认知机制，能够从海量数据中自主提取多层次特征表征，在医学影像分析领域展现出强大的模式识别与分类能力[15]。为解决传统深度学习模型的“黑箱”问题，可视化解释技术，如梯度加权类激活热图(gradient-weighted class activation mapping, Grad-CAM)、显著性热图应运而生，可直观呈现模型决策所依赖的影像学特征区域[10]，显著提升诊断过程的可解释性和临床可接受度。近年来，深度学习模型在乳腺超声领域取得突破性进展，已在病灶检测、良恶性分类、分子分型预测等方向验证其可行性[16-18]。

尽管现有研究验证了深度学习在乳腺结节分类中的效能，但针对低年资医师的可视化辅助诊断价值尚未系统评估。本研究通过严格的人机对照试验，探讨基于乳腺超声的深度学习可视化模型对低年资医师诊断性能的优化作用。

1 资料与方法

1.1 一般资料 回顾性收集2022年9月至2024年9月在上海市第六人民医院进行过常规检查超声发现乳腺结节并诊断为BI-RADS 3~4类的女性患者。纳入标准：(1) 有最终的外科手术或粗针穿刺病理诊断结果；(2) 活检或手术前2周内进行乳腺超声检查，且目标结节最大直径的二维灰阶超声图像清楚、完整，且均有完整的描述性报告及临床病理资料；(3) 根据第二版ACR BI-RADS[7]，将目标病变分为BI-RADS 3、4A、4B、4C类的结节。采用以下排除标准：(1) 病理结果不明确；(2) 肿瘤过大导致图像显示不完全的患者；(3) 患者术前接受抗癌治疗(化疗、放疗或内分泌治疗)；(4) 图像及临床资料不完整。

筛选后共纳入420例患者，年龄18~88(48.67±14.90)岁，并收集BI-RADS 3~4类乳腺结节患者的临床特征资料包括年龄、肿瘤大小、手术病理结果。为保证每次观察结果的统计独立性，多发的BI-RADS 3~4类乳腺病变的患者仅将类别最高的结节作

为目标结节纳入，以确保每次观察结果的统计独立性。将纳入的患者按7∶3比例随机划分为训练集（n=294）与测试集（n=126）。本研究获得上海市第六人民医院伦理委员会批准（No：2019-027）。由于研究为回顾性研究，豁免患者知情同意的要求。

1.2 方法

1.2.1 检查方法 本研究采用德国Siemens公司的S2000超声诊断仪及美国GE公司的超声诊断仪，两者探头检测频率均为4~9 MHz。患者均在术前2周内取仰卧位接受乳腺常规超声检查。对纳入的患者的目标结节进行横、纵切面扫描。在获取较好的二维图像后，选取乳腺结节最大切面，将图像以DICOM格式存储于硬盘，并同步记录乳腺结节的大小、部位及超声特征。

1.2.2 超声图像处理及深度学习模型建立 筛选符合纳入标准的乳腺结节超声图像并将其转化为JPEG格式。由于图像源自不同公司不同型号的超声诊断仪，尺寸不一，需统一调整为1 024×1 024像素。随后使用开源标签工具LabelMe（http://labelme.csail.mit.edu/Release3.0）根据病理结果标记图像，恶性为“1”、良性为“0”，同时生成JSON文件用以训练模型。基于多种深度学习模型主干提取深度学习特征并进行分类，在卷积层顶端分别加入512个全连接层、2个输出神经元和softmax激活函数，输出良、恶性乳腺结节的人工智能（artificial intelligence，AI）评分，softmax激活函数可确保一个病变的良恶性类别的AI评分和为1。以受试者工作特征（receiver operating characteristic，ROC）曲线评估超声深度学习模型的诊断效能，并在测试集上进一步验证模型。随后绘制模型的ROC曲线，对比曲线下面积（area under the curve，AUC）等指标，最终遴选出预测性能最优的模型进行可视化。共比较了GoogleNet、VGG16、AlexNet、ResNet50、Vision Transformer、DenseNet121这6种深度学习模型主干，选取验证队列中AUC最高的模型作为主干。ResNet50在大多数验证队列中AUC最高，将其作为特征提取的主干建立深度学习模型。

1.2.3 超声医师评估乳腺结节良恶性方法 根据ACR BI-RADS标准[7]，由一名低年资医师分别对乳腺结节进行BI-RADS分级及良恶性判断。依据以下标准对乳腺结节进行良恶性二分类诊断：BI-RADS 3和4A类乳腺结节定义为良性，BI-RADS 4B和4C类病灶定义为恶性。两周的洗脱期后，由同一低年资医师在AI辅助诊断系统支持下再次阅片，对结节进行定位与分析，并记录结节的数量、部位及最大径，随后汇总低年资医师两轮评读结果。

1.3 统计学方法 采用R软件和SPSS 26.0对数据进行统计分析，以病理结果作为金标准。对计量资料进行正态性及方差齐性检验，符合正态分布的连续变量以$\bar{x}\pm s$表示，组间比较采用独立样本t检验。分类变量以例（%）表示，比较采用χ^2检验。模型间的差异通过Delong检验（MedCalc软件，版本20.01）比较，临床变量的评估筛选使用SPSS 26.0。绘制ROC曲线，分析不同模型对乳腺结节良恶性鉴别诊断的效能，并通过Z检验比较AUC。类内相关系数检验（intraclass correlation coefficient，ICC）、Spearman秩相关分析、Z-score归一化处理均在Python（版本3.10）环境下完成。P<0.05为差异有统计学意义。

2 结 果

2.1 一般资料比较 本研究共计收集420例乳腺结节患者，BI-RADS分级为3/4A共281例，4B/4C共139例。其中术后病理证实良性病变为209例，恶性病变为211例。训练集和测试集患者年龄、结节位置、结节直径、BI-RADS分级及病理结果等指标差异均无统计学意义（P>0.05）。见表1。

2.2 深度学习模型辅助低年资医师诊断效果比较 为筛选出最优的乳腺结节良恶性鉴别模型，本研究对VGG16、GoogleNet、AlexNet、ResNet50、Vision Transformer以及DenseNet121的性能（表2）进行了比较。结果显示，ResNet50在AUC、准确率（DenseNet121训练集除外）和灵敏度方面均优于其他模型；其训练集和测试集的AUC分别达到了0.937（95%CI：0.911~0.962）和0.891（95%CI：0.833~0.948）（图1）。基于此，本研究最终选择ResNet50作为模型主干并对乳腺结节进行可视化分析，生成多种热图形式（图2），以辅助医师完成第二轮阅片。

表1 两集一般资料比较 ［例（%）］

Tab.1 Comparison of general data between two sets ［case（%）］

项目	训练集（n=294）	测试集（n=126）	t/χ^2值	P值
年龄（岁）[a]	49.18±14.65	47.49±15.48	1.065	0.287
结节直径（mm）[a]	18.52±8.68	20.17±12.84	1.319	0.189
结节位置				
左侧	122（41.50）	62（49.21）	2.130	0.144
右侧	172（58.50）	64（50.79）		
BI-RADS分级				
3、4A类	197（67.01）	84（66.67）	0.005	0.946
4B、4C类	97（32.99）	42（33.33）		
病理结果				
良性	148（50.34）	61（48.41）	0.131	0.717
恶性	146（49.66）	65（51.59）		

注：[a]表示数据形式为$\bar{x}\pm s$。

与首次阅片结果相比，深度学习可视化模型显著提升了诊断效率：训练集的准确率、灵敏度与特异度分别达到0.864、0.863和0.865。在测试集中，这三项指标分别为0.841、0.769和0.918。进一步分析比较两轮阅片可得，在深度学习可视化模型辅助下，低年资医师对乳腺肿块良恶性的鉴别准确率、灵敏度、特异度、精准率及F1值（精确率和召回率的调和平均值）均显著优于初次阅片，差异有统计学意义（$P<0.05$）。见表3。

表2　不同深度学习模型的诊断效能

Tab.2　Diagnostic efficiency of different deep learning models

深度学习模型	分组	AUC(95%*CI*)	准确率	灵敏度	特异度	精准率	F1值
VGG16	训练集	0.839(0.794~0.884)	0.786	0.733	0.838	0.817	0.773
	测试集	0.812(0.735~0.888)	0.770	0.754	0.787	0.790	0.772
GoogleNet	训练集	0.805(0.754~0.855)	0.752	0.692	0.811	0.783	0.735
	测试集	0.825(0.750~0.899)	0.770	0.662	0.885	0.860	0.748
AlexNet	训练集	0.760(0.705~0.814)	0.704	0.562	0.845	0.781	0.653
	测试集	0.789(0.709~0.868)	0.754	0.646	0.869	0.840	0.730
ResNet50	训练集	0.937(0.911~0.962)	0.864	0.863	0.865	0.863	0.863
	测试集	0.891(0.833~0.948)	0.841	0.769	0.918	0.909	0.833
Vision Transformer	训练集	0.562(0.496~0.628)	0.582	0.253	0.905	0.725	0.376
	测试集	0.631(0.533~0.727)	0.603	0.338	0.885	0.759	0.468
DenseNet121	训练集	0.933(0.905~0.960)	0.867	0.788	0.946	0.935	0.855
	测试集	0.879(0.816~0.941)	0.833	0.815	0.852	0.855	0.835

表3　AI辅助前后的人机对比

Tab.3　Human-machine comparison before and after AI assistance

类别	准确率	AUC(95%*CI*)	灵敏度	特异度	精准率	F1值
深度学习可视化模型	0.864	0.937 (0.911~0.962)	0.863	0.865	0.863	0.863
低年资医师	0.595	0.597 (0.545~0.647)	0.808	0.385	0.565	0.665
深度学习可视化模型辅助下低年资医师	0.905	0.904 (0.871~0.937)	0.849	0.959	0.965	0.899

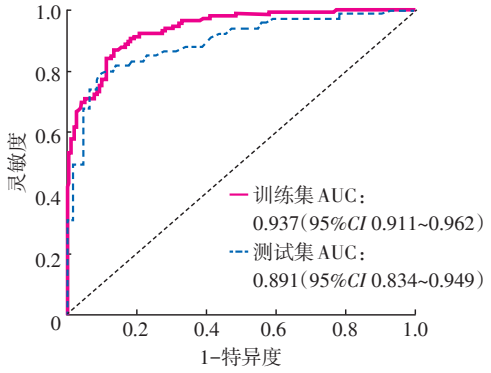
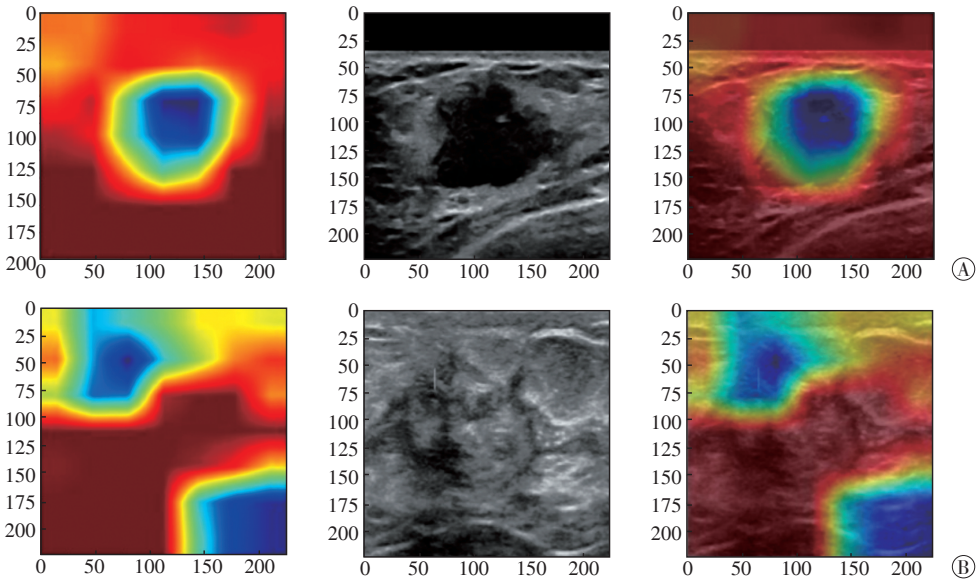


图1　训练集和测试集的ROC曲线

Fig.1　ROC curves of training set and test set



注：A为恶性结节可视化热图；B为良性结节可视化热图。

图2　ResNet50模型的可视化热图

Fig.2　Visualization heatmap of ResNet50 model

3 讨 论

本研究基于420例乳腺结节患者的超声图像，比较了多种深度学习模型在乳腺良恶性结节鉴别中的性能，并进一步探讨了以ResNet50为基础的可视化辅助系统对低年资医师诊断能力的提升效果。结果表明，ResNet50在训练集与测试集上均表现出优异的综合性能（AUC分别为0.937、0.891），且在引入热图可视化辅助后，低年资医师的诊断准确率从59.5%显著提升至90.5%，灵敏度、特异度、精准率及F1值亦同步改善。这一结果不仅验证了深度学习模型在乳腺影像诊断中的潜力，也为临床人机协同诊疗模式提供了实证支持，这对优化基层乳腺癌筛查流程具有重要公共卫生价值[19]。

目前乳腺癌诊断主要包括MRI、超声、乳腺钼靶等[20]。乳腺钼靶被视为全球标准化筛查手段，对微钙化等早期征象极为敏感，但乳腺密度增高时其灵敏度可由85%降至47.8%~64.4%[21]。乳腺超声无辐射、价格低，尤其适用于致密乳腺及年轻女性[22]；其对实质性肿块检出率高，便于区分囊实性病变并可实时引导穿刺，但成像分辨率有限且结果依赖操作者经验[23]。MRI在多灶、多中心病变及高危人群筛查中具有高灵敏度，可发现钼靶与超声难以识别的病灶[6]，但费用、检查时间及假阳性率均较高。由于上述成像方式各有短板，推动了深度学习的发展。与人工判读影像相比，深度学习模型可在一致、客观分析图像细节的同时捕捉人眼难辨的异常模式。深度学习辅助诊断系统有望作为“第二读片人”[24]，复核可疑病灶并显著提升筛查灵敏度。

近年来，卷积神经网络（convolutional neural network，CNN）被应用于图像分类、目标检测和语义分割等任务[25]。许多研究表明，基于CNN的模型在乳腺X线摄影、超声、磁共振成像等影像的病灶检测与分割中可媲美经验丰富超声科医师[26-27]。Wu等[28]采用MRI、超声及钼靶三模态影像的手工放射组学特征，并加入5 mm瘤周区域构建逻辑回归模型，测试集AUC为0.905。相比之下，本研究基于420例二维超声图像训练的ResNet50可视化模型，在训练与测试集中AUC分别为0.937（95% *CI*：0.911~0.962）、0.891（95% *CI*：0.833~0.948），与Wu等[28]的研究结果一致。人工智能不仅可以作为独立诊断工具，也能显著提升医师的诊断水平。在一项使用S-Detect系统的前瞻性研究中对216例患者进行分析发现，AI系统对乳腺结节诊断的特异度和准确率分别达92.7%和90.4%，在AI辅助下低年资放射科医生诊断的灵敏度从85.0%提高到93.3%，准确率从70.4%提升至83.9%，AUC由0.698升至0.835，显著缩小了不同经验医师间的差距[29]。本研究中，低年资医师在ResNet50可视化热图辅助下，准确率由59.5%提高至90.5%，灵敏度由80.8%升至84.9%，特异度由38.5%提升至95.9%，与上述报道一致，表明AI可以有效弥补经验不足带来的误判。这可能与深度学习可视化系统能够自动识别乳腺结节的关键特征并提供可视化分析有关，使低年资医师在深度学习可视化模型的帮助下识别乳腺结节的病理特征，提高低年资医师对微小病变和模糊图像的识别率，从而提高诊断的敏感度和特异度[30]，缩小其与高年资医师之间的诊断差距[31]。深度学习模型的“黑箱”特性限制了其临床应用，针对这一问题，各种可解释性方法被提出。Grad-CAM可视化技术在RcdNet、EfficientNet-B7等模型中已证明能够突出肿块周围的关键结构，使模型关注区域与医师的诊断依据一致[32-33]。MUP-Net等采用基于领域知识的可解释框架，通过联合利用B-mode和彩色多普勒图像，在提供AUC=0.902的同时，显著提高初学者的信心[10]。然而系统性评估显示目前仅有少数研究考察了可解释模型对临床信任的影响，大多数研究在数据质量、稳定性和可用性评估方面存在不足[34]。本研究通过热图反馈提升医师诊断准确率，间接体现了可解释性的作用，但缺乏对医师信任度的量化评价，未来需在用户研究和心理层面进一步探索。

尽管本研究取得了较好的结果，但仍有局限性。（1）本研究中训练集和测试集使用的超声图像数量相对有限，可能影响模型的稳定性，使其稳定性降低。在后续的工作中将扩大数据的规模，进一步验证模型的性能。（2）本研究为回顾性研究，需通过更大规模的前瞻性试验来全面评估模型的临床应用价值。（3）本研究的图像信息维度单一，仅利用二维超声图像，未纳入三维特征，可能遗漏肿瘤的部分特征信息。

综上所述，基于乳腺二维超声图像的深度学习可视化模型具有良好的乳腺结节良恶性鉴别能力。该模型能有效辅助低年资医师提高诊断的敏感度及整体效能，可作为提升低年资医师乳腺超声诊断规范化水平、缩短培训周期的有价值的辅助工具。

利益冲突 无

参考文献

[1] Bray F, Laversanne M, Weiderpass E, et al. The ever-increasing importance of cancer as a leading cause of premature death world-

wide[J]. Cancer, 2021, 127(16): 3029-3030.
[2] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2024, 74(3): 229-263.
[3] Pal R, Srivastava N, Chopra R, et al. Investigation of DNA damage response and apoptotic gene methylation pattern in sporadic breast tumors using high throughput quantitative DNA methylation analysis technology[J]. Mol Cancer, 2010, 9(1): 303.
[4] 宋鹏杰, 王艳蕾, 孙晨, 等. 基于真实世界数据的乳腺超声人工智能诊断和X射线钼靶摄影检查在早期乳腺癌筛查中的对比研究[J]. 中国医学装备, 2023, 20(9): 20-24.
[5] Zhou YP, Wei J, Wu DM, et al. Generating full-field digital mammogram from digitized screen-film mammogram for breast cancer screening with high-resolution generative adversarial network[J]. Front Oncol, 2022, 12: 868257.
[6] 周雅筠, 谢瑾, 张蓓, 等. 常规超声、超声造影、钼靶X线、磁共振成像及联合检测对早期乳腺癌患者的诊断价值[J]. 贵州医科大学学报, 2023, 48(2): 172-176.
[7] Maani N, Westergard S, Yang J, et al. NF1 patients receiving breast cancer screening: insights from the Ontario high risk breast screening program[J]. Cancers, 2019, 11(5): 707.
[8] Ezeana CF, He TC, Patel TA, et al. A deep learning decision support tool to improve risk stratification and reduce unnecessary biopsies in BI-RADS 4 mammograms[J]. Radiol Artif Intell, 2023, 5(6): e220259.
[9] He P, Chen W, Bai MY, et al. Deep learning-based computer-aided diagnosis for breast lesion classification on ultrasound: a prospective multicenter study of radiologists without breast ultrasound expertise[J]. AJR Am J Roentgenol, 2023, 221(4): 450-459.
[10] Yan L, Liang ZY, Zhang H, et al. A domain knowledge-based interpretable deep learning system for improving clinical breast ultrasound diagnosis[J]. Commun Med, 2024, 4(1): 90.
[11] Yao Z, Luo T, Dong YJ, et al. Virtual elastography ultrasound via generative adversarial network for breast cancer diagnosis[J]. Nat Commun, 2023, 14(1): 788.
[12] Wang J, Tian HT, Yang X, et al. Artificial intelligence in breast US diagnosis and report generation[J]. Radiol Artif Intell, 2025, 7(4): e240625.
[13] Yang YT, Liao TT, Lin XH, et al. Reducing unnecessary biopsies of BI-RADS 4 lesions based on a deep learning model for mammography[J]. Front Oncol, 2025, 15: 1543553.
[14] 米拉·也尔兰, 张海见, 胡峙珩, 等. 超声影像组学对BI-RADS 4a类不规则乳腺结节良恶性的鉴别价值[J]. 分子影像学杂志, 2023, 46(1): 12-20.
[15] Carriero A, Groenhoff L, Vologina E, et al. Deep learning in breast cancer imaging: state of the art and recent advancements in early 2024[J]. Diagnostics, 2024, 14(8): 848.
[16] Dembrower K, Crippa A, Colón E, et al. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study[J]. Lancet Digit Health, 2023, 5(10): e703-e711.
[17] Ng AY, Oberije CJG, Ambrózay É, et al. Prospective implementation of AI-assisted screen reading to improve early detection of breast cancer[J]. Nat Med, 2023, 29(12): 3044-3049.
[18] Eisemann N, Bunk S, Mukama T, et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography screening[J]. Nat Med, 2025, 31(3): 917-924.
[19] Wu TH, Liu XY, Si ZY, et al. Combination of ultrasound-based radiomics and deep learning with clinical data to predict response in breast cancer patients treated with neoadjuvant chemotherapy[J]. Front Oncol, 2025, 15: 1525285.
[20] 中国抗癌协会乳腺癌专业委员会, 中华医学会肿瘤学分会乳腺肿瘤学组. 中国抗癌协会乳腺癌诊治指南与规范(2024年版)[J]. 中国癌症杂志, 2023, 33(12): 1092-1186.
[21] Thigpen D, Kappler A, Brem R. The role of ultrasound in screening dense breasts-a review of the literature and practical solutions for implementation[J]. Diagnostics, 2018, 8(1): 20.
[22] 吴城炜, 邹大中. 超声联合X线钼靶用于乳腺癌的诊断价值及与分子亚型的相关性研究[J]. 医学理论与实践, 2023, 36(24): 4149-4153.
[23] Vimala BB, Srinivasan S, Mathivanan SK, et al. Image noise removal in ultrasound breast images based on hybrid deep learning technique[J]. Sensors, 2023, 23(3): 1167.
[24] Frazer HML, Peña-Solorzano CA, Kwok CF, et al. Comparison of AI-integrated pathways with human-AI interaction in population mammographic screening for breast cancer[J]. Nat Commun, 2024, 15(1): 7525.
[25] Takahashi S, Sakaguchi Y, Kouno N, et al. Comparison of vision transformers and convolutional neural networks in medical image analysis: a systematic review[J]. J Med Syst, 2024, 48(1): 84.
[26] Luo LY, Wu MX, Li M, et al. A large model for non-invasive and personalized management of breast cancer from multiparametric MRI[J]. Nat Commun, 2025, 16(1): 3647.
[27] Iima M, Mizuno R, Kataoka M, et al. Deep learning applied to diffusion-weighted imaging for differentiating malignant from benign breast tumors without lesion segmentation[J]. Radiol Artif Intell, 2025, 7(1): e240206.
[28] Wu J, Li YX, Gong WQ, et al. Multi-modality radiomics diagnosis of breast cancer based on MRI, ultrasound and mammography[J]. BMC Med Imaging, 2025, 25(1): 265.
[29] Singla V, Garg D, Negi S, et al. Deep learning powered breast ultrasound to improve characterization of breast masses: a prospective study[J]. Acta Radiol, 2026, 67(1): 13-23.
[30] Yang L, Zhang NW, Jia JY, et al. Deep learning radiomics on grayscale ultrasound images assists in diagnosing benign and malignant of BI-RADS 4 lesions[J]. Sci Rep, 2024, 14(1): 31479.
[31] Xiang HL, Wang X, Xu M, et al. Deep learning-assisted diagnosis of breast lesions on US images: a multivendor, multicenter study[J]. Radiol Artif Intell, 2023, 5(5): e220185.
[32] Liu BQ, Liu SY, Cao ZJ, et al. Accurate classification of benign and malignant breast tumors in ultrasound imaging with an enhanced deep learning model[J]. Front Bioeng Biotechnol, 2025, 13: 1526260.
[33] Latha M, Kumar PS, Chandrika RR, et al. Revolutionizing breast ultrasound diagnostics with EfficientNet-B7 and Explainable AI[J]. BMC Med Imaging, 2024, 24(1): 230.
[34] Gurmessa DK, Jimma W. Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: a systematic review[J]. BMJ Health Care Inform, 2024, 31(1): e100954.

收稿日期:2025-11-03 **修回日期**:2026-01-12 **编辑**:叶小舟