Unveiling complex medical interdependencies through high-order correlation mining

Edited by
Donglin Di, Zhenzhong Deng, Lei Fan
and Juan Wang

Generative AI statement

Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Unveiling complex medical interdependencies through high-order correlation mining

Topic editors

Donglin Di — Li Auto, China

Zhenzhong Deng — University of Southern California, United States

Lei Fan — University of New South Wales, Australia

Juan Wang — The Second Affiliated Hospital of Xi'an Jiaotong University, China

Citation

Table of
contents

| frontiers | Frontiers in Medicine |

Check for updates

A hypergraph transformer method for brain disease diagnosis

Xiangmin Han[1], Jingxi Feng[2], Heming Xu[2], Shaoyi Du[2] and Junchang Li[3]*

[1]School of Software, Tsinghua University, Beijing, China, [2]The Institute of Artificial Intelligence and Robotics, College of Artificial Intelligence, Xi'an Jiaotong University, Xi'an, China, [3]Shenzhen Clinical Research Center for Mental Disorders, Shenzhen Kangning Hospital and Shenzhen Mental Health Center, Shenzhen, China

Objective: To address the high-order correlation modeling and fusion challenges between functional and structural brain networks.

Method: This paper proposes a hypergraph transformer method for modeling high-order correlations between functional and structural brain networks. By utilizing hypergraphs, we can effectively capture the high-order correlations within brain networks. The Transformer model provides robust feature extraction and integration capabilities that are capable of handling complex multimodal brain imaging.

Results: The proposed method is evaluated on the ABIDE and ADNI datasets. It outperforms all the comparison methods, including traditional and graph-based methods, in diagnosing different types of brain diseases. The experimental results demonstrate its potential and application prospects in clinical practice.

Conclusion: The proposed method provides new tools and insights for brain disease diagnosis, improving accuracy and aiding in understanding complex brain network relationships, thus laying a foundation for future brain science research.

1 Introduction

The structural and functional connections of brain networks reflect the interaction and collaboration between different brain regions (1, 2). Structural connectivity is typically represented by the distribution of neural fiber tracts (3), while functional connectivity describes the synchronous activity of different brain regions during specific tasks or at rest state (4). Understanding the structural and functional connectivity of brain networks is crucial for comprehending both normal brain function and pathological states (5). For instance, abnormalities in structural connectivity may be associated with brain tissue damage, while disruptions in functional connectivity could indicate communication issues between neurons. Therefore, studying functional and structural brain networks is essential for uncovering the mechanisms underlying brain disease diagnosis.

Current research on brain networks primarily relies on techniques such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI). fMRI captures brain activity during specific tasks or at rest, revealing functional connectivity between different brain regions. DTI, on the other hand, tracks the diffusion paths of water

molecules within neural fibers, providing information on the structural connectivity of the brain's white matter. Integrating data from fMRI and DTI offers a more comprehensive and enriched perspective for diagnosing brain diseases. For example, in Alzheimer's disease research, fMRI can reveal changes in functional connectivity, while DTI can demonstrate the degradation of white matter structure. In recent years, multimodal imaging techniques that combine fMRI and DTI have become mainstream in brain network research, further enhancing diagnostic accuracy and depth of understanding regarding brain diseases.

Artificial intelligence (AI) has achieved great success in various fields (6). For the brain network analysis task, graph and hypergraph methods (7–9) have shown great potential in brain network research. Graph methods represent brain networks as vertices and edges, allowing for the analysis of pairwise, low-order relationships. However, these methods have limitations, as they fail to effectively capture higher-order relationships within brain networks. For instance, traditional graph neural networks (GNNs) (10, 11) often underperform in handling complex high-order interactions (12, 13), neglecting the interactions among multiple vertices. Hypergraph methods (14) introduce hyperedges, which better model higher-order relationships in brain networks, but challenges remain in integrating functional and structural brain networks (15). Although hypergraphs can represent high-order relationships among multiple vertices, existing methods lack effective strategies for integrating information from different modalities, making it difficult to fully leverage the advantages of multimodal data. Thus, new methods are needed to address these issues and improve the accuracy and reliability of brain disease diagnosis.

This paper proposes a hypergraph Transformer (HGTrans) method for calculating high-order correlations between functional and structural brain networks. By utilizing hypergraphs, we can effectively model the high-order interactions within brain networks. The Transformer model provides robust feature extraction and integration capabilities, capable of handling complex multimodal data. Specifically, we use hypergraphs to represent high-order correlations in brain networks, including both functional and structural connectivity. Then, we propose the cross-attention Transformer module to extract features and integrate information from the hypergraphs, constructing a joint representation of the functional-structural brain network. This approach not only captures high-order functional and structural correlations but also effectively integrates information from different modalities, enhancing brain disease diagnosis performance. The main contributions of this paper are as follows:

- We propose a hypergraph-based method for modeling and computing the integration of functional and structural brain networks, effectively capturing high-order correlations. By using hypergraph modeling, we can accurately represent high-order interactions among multiple regions within brain networks, thereby enhancing our understanding and diagnosis performance of brain diseases.
- We introduce the Transformer model into hypergraph-based multimodal brain disease diagnosis, integrating diverse information from fMRI and DTI to improve diagnostic accuracy. The Transformer is conducted to refine

the structural embeddings by incorporating high-order relationships derived from the functional network, thereby enhancing the diagnosis of brain diseases.
- We validated our method on the ABIDE and ADNI datasets, showing that our approach outperforms all the traditional and graph-based methods for different types of brain diseases, demonstrating its potential and application prospects in brain disease diagnosis.

2 Materials and methods

2.1 Datasets and preprocessing

The proposed method is evaluated on the ABIDE (16) and ADNI (17) datasets. We utilized the NYU1 and TCD sites of the ABIDE database in this work. Specifically, the NYU1 dataset contains 55 subjects, of which 33 subjects are autism spectrum disorder (ASD) patients and 22 subjects are normal controls (NCs). The TCD site contains 40 subjects, of which 20 subjects are ASD subjects and 20 subjects are NCs. The ADNI dataset is collected from multiple sites that study for improving the clinical trials for the prevention and treatment of Alzheimer's disease (AD). We used a subset of ADNI in this work, consisting of 39 AD patients, 62 MCI patients, and 61 NCs. Each subject has both rs-fMRI and DTI data in this work. The AAL (18) brain atlas was used to segment the regions of interest (ROIs) of the brain network. We preprocessed the original rs-fMRI via DPARSF,[1] and the original DTI via PANDAS.

2.2 Method

2.2.1 Preliminaries of hypergraph computation

The hypergraph computation framework models high-order correlations by using hyperedges, which represent complex relationships beyond pairwise connections, and performs collaborative computation on these high-order interactions. Each hyperedge can connect multiple vertices, allowing it to capture both low-order (pairwise) correlations and high-order correlations across larger vertex sets. This approach leverages these high-order interactions to optimize data usage and improve overall task performance.

Given a hypergraph $\mathcal{H} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, where $\mathcal{V}$ and $\mathcal{E}$ represent the vertex set and the hyperedge set, respectively, and $\mathbf{W}$ denotes the weight matrix of the hyperedges. The incidence matrix of the hypergraph is defined as a $|\mathcal{V}| \times |\mathcal{E}|$ matrix, with each entry defined as

$$\mathbf{H}(v, e) = \begin{cases} w_e(v), & \text{if} \quad v \in e \\ 0, & \text{if} \quad v \notin e \end{cases}, \tag{1}$$

where $w_e(v) \in \mathbf{W}$ represents the weight of vertex v within the hyperedge e.

---

1 http://rfmri.org/DPARSF

2.2.2 HGTrans framework

As shown in Figure 1, the proposed HGTrans Framework consists of two main modules: the hypergraph computation module based on brain imaging and the structure-function Transformer module. The former constructs high-order relational structures from the information embedded in fMRI and DTI, exploring the complex relationships between different brain regions under fMRI and DTI, and generating high-order feature representations for fMRI and DTI. Then, semantic computations are performed using a hypergraph neural network to generate high-order feature representations. The latter uses the high-order features of the functional brain network as keys (K) and values (V), and the high-order features of the structural brain network as queries (Q) to achieve information interaction and fusion within the Transformer module. Finally, the fused features are fed into a classifier to enable brain disease diagnosis.

2.2.3 Hypergraph computation for fMRI and DTI

2.2.3.1 Higher-order functional brain network representation

To model the complex interactions within functional brain networks, we utilize hypergraphs, which allow for the connection of multiple ROIs in the brain, rather than just pairs of regions. This structure facilitates the representation of high-order associations that arise in functional brain activity. The time series data of 116 ROIs from each subject's resting-state fMRI (rs-fMRI) data is extracted, followed by calculating the Pearson correlation coefficient between each pair of ROIs. This correlation coefficient quantifies the degree of linear relationship, ranging from –1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no linear association. Using this approach, a functional connectivity (FC) matrix of size 116×116 is then generated, where each element represents the pairwise linear correlation between two ROIs.

In the hypergraph model, each of the 116 ROIs is treated as a vertex in the set $\mathcal{V} = \{v_1, v_2, \ldots, v_{116}\}$, with v_i representing the i-th vertex. In this work, we fix the K value as 3. The vertex feature set $X_f = \{x_{f1}, x_{f2}, \ldots, x_{f116}\}$ describes the Pearson correlation values between the i-th ROI and all other ROIs. To capture the structural relationships between ROIs, we apply a K-Nearest Neighbors (KNN) algorithm to identify the $k_1 - 1$ nearest neighbors for each vertex v_i. A hyperedge is then formed for each vertex, connecting it with its nearest neighbors. Each hyperedge e_j, constructed using KNN with a specified k-value, can be expressed as $e_j = \{v_1, v_2, \ldots, v_k\}$, where k represents the number of vertices in the hyperedge. The similarity between vertices is measured using Euclidean distance, calculated as follows:

$$E_{\text{dist}}(v_i, v_j) = \sqrt{\sum_{p=1}^{d^{(l)}} (z_{i,p} - z_{j,p})^2} \qquad (2)$$

where $E_{\text{dist}}(v_i, v_j)$ denotes the Euclidean distance between vertices v_i and v_j, and $d^{(l)}$ represents the number of feature dimensions in layer l.

By incorporating KNN with multiple values of k, representing local and global scales, the resulting hyperedges reflect complex high-order interactions between the vertices. The functional brain network hypergraph is then used for hypergraph convolution, allowing the learning of vertex representations. The HGNN+ convolution operation (19) consists of a two-step message-passing scheme. The process is formalized as follows:

$$Z^t = WH^T D_e^{-1} X^t$$
$$X^{t+1} = \sigma(D_v^{-1} HZ\theta^{t+1}) \qquad (3)$$

where $X^t \in \mathbb{R}^{|V| \times M_t}$ is the vertex feature matrix at layer t, and $Z^t \in \mathbb{R}^{|E| \times M_t}$ is the corresponding hyperedge feature matrix. The learnable parameter matrix $\theta^{t+1} \in \mathbb{R}^{M_t \times M_{t+1}}$ defines the transformation for the subsequent layer. Initially, the incidence matrix H guides the aggregation of vertex features to generate the hyperedge feature matrix Z^t. These features are then combined with vertex-specific hyperedge features using the learnable parameters θ^t, updating the vertex feature matrix X^{t+1}. A nonlinear activation function $\sigma(\cdot)$ is applied to facilitate the transformation of features.

The vertex embeddings derived from multiple layers of hypergraph convolution effectively capture high-order relationships between ROIs within the functional brain network. This modeling approach provides a superior representation of complex brain activity patterns.

2.2.3.2 High-order structural brain networks

DTI data is utilized to derive the structural connectivity (SC) matrix, which quantifies the fiber tract connections between various ROIs in the brain. This method facilitates a comprehensive evaluation of potential alterations in the structural brain network that may be associated with ASD, offering a holistic perspective on how the disease may impact brain function.

The structural brain network is characterized by features such as small-world architecture and rich-club organization, both of which are critical for understanding network efficiency and communication. High-order structural characteristics are captured by computing the clustering coefficient c_i and degree centrality d_i for each ROI, based on the SC matrix. The clustering coefficient assesses the extent of local interconnectivity, while degree centrality indicates the relative importance of each region within the broader network. These metrics provide valuable insights into the efficiency of information processing and communication within and between local brain regions.

The feature representation for each vertex in the network is defined as $X_s = \{x_{s1}, x_{s2}, \ldots, x_{s116}\}$, where x_{si} represents the feature vector for the i-th ROI, with $x_{si} = \{c_{si}, d_{si}\}$. These initial features serve as input for subsequent analysis and modeling. To capture the higher-order relationships between ROIs, a K-Nearest Neighbors (KNN) algorithm is employed to construct a hypergraph representation of the structural brain network. This hypergraph captures multi-dimensional interactions that extend beyond simple pairwise connections, allowing for a more detailed representation of the complex inter-regional relationships in the brain.

Following the construction of the hypergraph, HGNN+ (19) is applied for feature learning and information integration. The hypergraph convolution process mirrors the procedure used for the functional brain network, as indicated in Equation 3. After two layers of hypergraph convolution, the resulting vertex embeddings encode high-order structural features, which are used as the final representations of each brain region. These embeddings enable a

FIGURE 1
The pipeline of the proposed HGTrans framework.
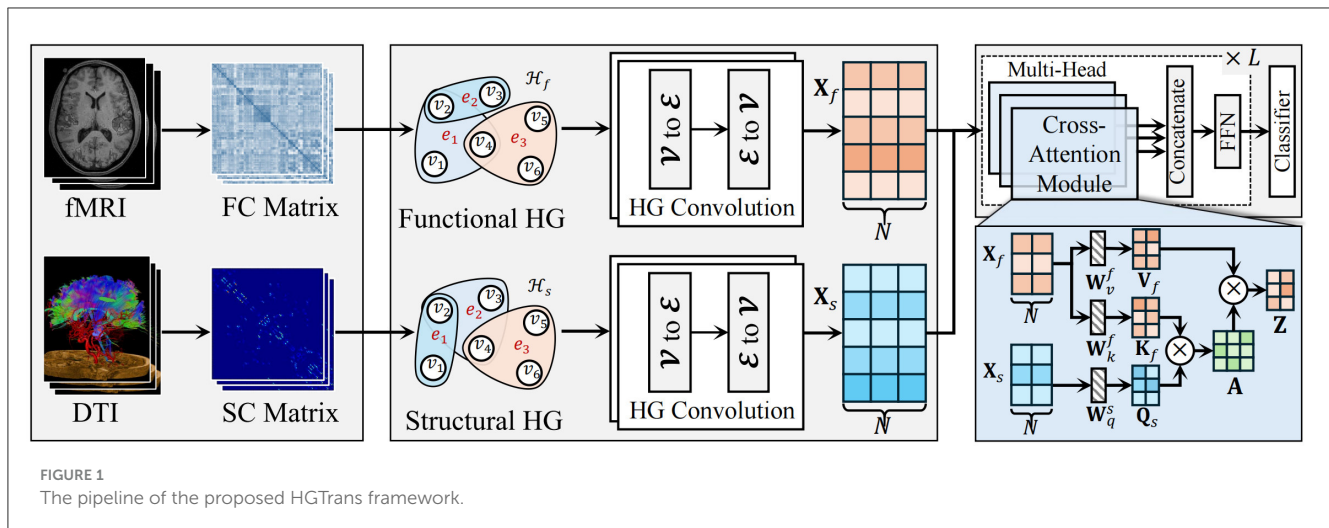
more nuanced analysis of the structural brain network, particularly in understanding the structural alterations associated with ASD. This approach provides a rigorous framework for examining both local and global connectivity patterns within the brain, offering valuable insights into the structural mechanisms underlying ASD.

2.2.3.3 Cross-attention transformer for multimodal integration

To effectively fuse functional and structural brain network features, a cross-attention Transformer module is introduced. This module leverages the Transformer architecture to model long-range dependencies between multimodal features, using the structural embeddings after hypergraph convolution as the query (Q) and the functional embeddings as the key (K) and value (V), enabling the integration of both modalities.

First, the structural embedding matrix X_s and the functional embedding matrix X_f obtained from hypergraph convolution are projected into Q, K, and V representations as follows:

$$Q_s = W_q^s X_s, \quad K_f = W_k^f X_f, \quad V_f = W_v^f X_f, \qquad (4)$$

where W_q^s, W_k^f, and W_v^f are learnable weight matrices that linearly project the structural and functional embeddings. This step maps both sets of features into a shared feature space, preparing them for cross-attention.

Next, through the cross-attention mechanism, the query matrix Q_s from the structural features attends to the key matrix K_f from the functional features, generating the attention weight matrix:

$$A = \text{softmax}\left(\frac{Q_s K_f^T}{\sqrt{d_k}}\right), \qquad (5)$$

where A represents the attention weight matrix, and d_k is the dimensionality of the key, used for scaling. These attention weights are then applied to the value matrix V_f from the functional features to generate updated structural feature embeddings:

$$Z = A V_f. \qquad (6)$$

This process refines the structural embeddings by incorporating high-order relationships derived from the functional network, enabling a more comprehensive representation of brain activity.

2.2.3.4 Brain disease diagnosis

The learned feature representations from the cross-attention Transformer module are then fed into the output layer for classification. The output layer consists of a fully connected layer and a log_softmax activation function to facilitate the final classification prediction.

Let x_i and b_i represent the input and bias for the i-th hidden layer, respectively, while W_i denotes the weight matrix facilitating connections from the i-th to the $i+1$-th hidden layer. Subsequently, the activation of the $i + 1$-th hidden layer is computed using the equation below:

$$Z_{i+1} = f(W_i x_i + b_i), \qquad (7)$$

Z_{i+1} is the activated output of layer $i+1$. f denotes the activation function.

$$f(x) = \max(0, x), \qquad (8)$$

The ReLU activation function constrains its output to the range $[0, \infty)$. For the final layer, a fully connected dense layer is paired with a log_softmax activation functionactivation function to execute the terminal classification predictions.

$$\text{log-softmax}(z_i) = z_i - \log\left(\sum_j e^{z_j}\right), \qquad (9)$$

The log-softmax function, which uses Euler's number e as the base for the natural logarithm. In a binary classification setting, it provides log probabilities as the output. We utilize the Adam optimizer for the optimization process, setting a relatively low learning rate of 1×10^{-5}. The negative log-likelihood loss function is utilized for the binary classification task, the loss function is defined as follows:

$$\mathcal{L} = -\log(p_y) \qquad (10)$$

where p_y represents the probability of the correct class y.

3 Results and discussion

The proposed method is compared against four categories of methods:

- Single-modality-based baseline: SVM (20), MLP (21)
- Single-modality-based graph methods: GCN (10), GAT (11), and GraphSage (22).
- Single-modality-based hypergraph method: HGNN+ (19).
- Multi-modality-based methods: BrainNN (23) and MVGCN (24).

A three-fold cross-validation approach was utilized to evaluate each method, quantifying the accuracy of ASD disease classification predictions using metrics such as accuracy, sensitivity, specificity, and F1 score. The final results is given by mean $\pm$ standard error. Tables 1, 2 show the experimental results of ABIDE and ADNI, respectively.

3.1 Comparison with single-modality baseline methods

In Tables 1, 2, single-modality baseline methods include traditional machine learning approaches such as Support Vector Machines (SVM) and Multilayer Perceptron (MLP). While these methods are widely used for classification tasks, they are limited to features from a single modality and cannot capture the complex interactions within brain networks. Specifically, on the ABIDE-NYU dataset, SVM achieved an accuracy of 0.746, and MLP achieved 0.655, both lower than the accuracy of 0.799 achieved by our proposed HGTrans. Similarly, on the ABIDE-TCD dataset, SVM and MLP achieved 0.625 and 0.646, respectively, which are significantly lower than HGTrans's 0.749. These results indicate that single-modality baseline methods are insufficient for effectively addressing the complexity of ASD data. By integrating both fMRI and DTI data, HGTrans can capture more informative features from different perspectives of the brain network, leading to superior classification performance. This highlights the necessity and effectiveness of multimodal data fusion.

3.2 Comparison with graph-based methods

Graph-based methods, including GCN, GAT, and GraphSAGE, utilize the graph structure of brain networks to model relationships between regions of interest (ROIs). These methods can capture more complex spatial topological features than traditional single-modality methods. However, as shown in Table 1, on the ABIDE-NYU dataset, GCN achieved an accuracy of 0.618, GAT 0.617, and GraphSAGE 0.582, all significantly lower than HGTrans's 0.799. Similarly, on the ABIDE-TCD dataset, GCN, GAT, and GraphSAGE achieved accuracies of 0.625, 0.623, and 0.623, respectively, which are lower than HGTrans's 0.749. Although graph-based methods can capture the topological information within brain networks, they are limited to modeling pairwise relationships and cannot fully represent higher-order interactions

between brain regions. In contrast, HGTrans leverages hypergraph modeling to capture more complex higher-order relationships in multimodal settings, which significantly improves classification performance over traditional graph methods.

3.3 Comparison with hypergraph-based methods

Hypergraph-based methods, such as HGNN+, extend the capabilities of graph models by capturing higher-order relationships between multiple brain regions through hypergraph structures. On the ABIDE-NYU dataset, HGNN+ achieved an accuracy of 0.707, and on the ABIDE-TCD dataset, it achieved 0.676, both close to but lower than HGTrans's 0.799 and 0.749, respectively. These results show that while hypergraph methods can capture more complex brain region interactions, performance remains limited when using single-modality data. HGTrans outperforms HGNN+ primarily due to its ability to not only capture higher-order spatial topological structures through hypergraphs but also effectively integrate functional and structural brain network features using cross-attention mechanisms. By jointly modeling multimodal data, HGTrans generates more robust embeddings, leading to superior performance compared to single-modality hypergraph methods. On the other hand, When we engage in cognitive activities such as reading, writing, and listening, multiple brain regions cooperate to complete the tasks (25, 26), rather than a single brain region or pairs of brain regions working independently. Traditional methods find it difficult to model such group high-order correlations. However, high-order correlation modeling and semantic computation based on hypergraphs can achieve high-order correlation-driven local brain region cooperative message passing, which is more efficient than traditional graph neural networks and contains richer information. Therefore, for brain disease diagnosis tasks, the hypergraph computation model can provide more abundant semantic information, thereby improving diagnostic performance.

3.4 Comparison with multimodal methods

Multimodal methods, such as MVGNN and BrainNN, integrate both fMRI and DTI data to capture complementary information from different brain modalities. As shown in Tables 1, 2, while these multimodal methods outperform single-modality and graph-based methods, HGTrans still achieves the highest accuracy across both datasets. On the ABIDE-NYU dataset, MVGNN achieved an accuracy of 0.748, and BrainNN 0.688, both lower than HGTrans's 0.799. On the ABIDE-TCD dataset, MVGNN, and BrainNN achieved accuracies of 0.698 and 0.672, respectively, also lower than HGTrans's 0.749. HGTrans's advantage lies in its ability to not only fuse multimodal data but also effectively capture the complex interactions between functional and structural brain networks through hypergraph structures and cross-attention mechanisms. This mechanism allows the model to fully leverage the relationships between functional and structural brain networks, resulting in more expressive features and higher

TABLE 1 The comparison results on the two ABIDE datasets.

Modality	Method		Accuracy	Sensitivity	Specificity	F1-score
ABIDE-NYU						
fMRI	SVM	(mean ± std)	0.746 ± 0.024	0.879 ± 0.043	0.547 ± 0.034	0.633 ± 0.024
	MLP	(mean ± std)	0.655 ± 0.048	0.879 ± 0.086	0.327 ± 0.185	0.400 ± 0.163
	GCN	(mean ± std)	0.618 ± 0.046	0.909 ± 0.074	0.185 ± 0.072	0.274 ± 0.090
	GAT	(mean ± std)	0.617 ± 0.053	0.667 ± 0.086	0.542 ± 0.083	0.590 ± 0.084
	GraphSage	(mean ± std)	0.582 ± 0.068	0.606 ± 0.086	0.542 ± 0.083	0.509 ± 0.077
	HGNN+	(mean ± std)	0.707 ± 0.098	0.849 ± 0.113	0.494 ± 0.149	0.672 ± 0.117
	BrainGB	(mean ± std)	0.691 ± 0.068	0.849 ± 0.043	0.452 ± 0.121	0.651 ± 0.085
	BrainGNN	(mean ± std)	0.727 ± 0.046	0.818 ± 0.074	0.583 ± 0.131	0.703 ± 0.056
DTI	SVM	(mean ± std)	0.582 ± 0.068	0.818 ± 0.074	0.185 ± 0.072	0.509 ± 0.077
	MLP	(mean ± std)	0.563 ± 0.011	0.818 ± 0.074	0.232 ± 0.139	0.244 ± 0.063
	GCN	(mean ± std)	0.527 ± 0.068	0.788 ± 0.086	0.143 ± 0.202	0.154 ± 0.218
	GAT	(mean ± std)	0.511 ± 0.078	0.697 ± 0.043	0.232 ± 0.139	0.266 ± 0.139
	GraphSage	(mean ± std)	0.545 ± 0.033	0.758 ± 0.113	0.232 ± 0.139	0.270 ± 0.112
	HGNN+	(mean ± std)	0.637 ± 0.0 61	0.758 ± 0.113	0.548 ± 0.034	0.622 ± 0.056
fMRI&DTI	BrainNN	(mean ± std)	0.688 ± 0.142	0.788 ± 0.086	0.536 ± 0.278	0.669 ± 0.049
	MVGCN	(mean ± std)	0.748 ± 0.104	0.909 ± 0.091	0.512 ± 0.238	0.602 ± 0.206
	HGTrans	(mean ± std)	0.799 ± 0.0957	0.909 ± 0.074	0.631 ± 0.144	0.778 ± 0.110
ABIDE-TCD						
fMRI	SVM	(mean ± std)	0.625 ± 0.112	0.650 ± 0.200	0.600 ± 0.255	0.628 ± 0.101
	MLP	(mean ± std)	0.646 ± 0.136	0.500 ± 0.175	0.794 ± 0.900	0.636 ± 0.140
	GCN	(mean ± std)	0.625 ± 0.126	0.794 ± 0.147	0.468 ± 0.258	0.526 ± 0.018
	GAT	(mean ± std)	0.623 ± 0.072	0.667 ± 0.294	0.611 ± 0.235	0.595 ± 0.112
	GraphSage	(mean ± std)	0.623 ± 0.072	0.556 ± 0.192	0.706 ± 0.107	0.652 ± 0.045
	HGNN+	(mean ± std)	0.676 ± 0.067	0.659 ± 0.124	0.363 ± 0.059	0.500 ± 0.082
	BrainGB	(mean ± std)	0.691 ± 0.068	0.849 ± 0.043	0.452 ± 0.121	0.651 ± 0.085
	BrainGNN	(mean ± std)	0.698 ± 0.070	0.746 ± 0.081	0.651 ± 0.059	0.697 ± 0.069
DTI	SVM	(mean ± std)	0.582 ± 0.068	0.818 ± 0.074	0.185 ± 0.072	0.509 ± 0.077
	MLP	(mean ± std)	0.526 ± 0.065	0.508 ± 0.259	0.548 ± 0.236	0.514 ± 0.112
	GCN	(mean ± std)	0.527 ± 0.068	0.788 ± 0.086	0.143 ± 0.202	0.154 ± 0.218
	GAT	(mean ± std)	0.601 ± 0.127	0.651 ± 0.059	0.564 ± 0.224	0.568 ± 0.166
	GraphSage	(mean ± std)	0.500 ± 0.031	0.540 ± 0.157	0.437 ± 0.224	0.436 ± 0.165
	HGNN+	(mean ± std)	0.498 ± 0.136	0.349 ± 0.059	0.667 ± 0.294	0.482 ± 0.123
fMRI&DTI	BrainNN	(mean ± std)	0.672 ± 0.080	0.444 ± 0.098	0.897 ± 0.074	0.732 ± 0.067
	MVGCN	(mean ± std)	0.698 ± 0.113	0.659 ± 0.170	0.746 ± 0.081	0.714 ± 0.101
	HGTrans	(mean ± std)	0.749 ± 0.098	0.970 ± 0.043	0.698 ± 0.022	0.748 ± 0.097

classification accuracy. There are also some domain adaption methods (27, 28) that can be used to transfer knowledge between structural and functional brain imaging. Although these cross-modal information transfer methods can achieve inference with only one modality in the testing phase, the performance is greatly limited by the lack of shared labels to guide the cross modality fusion.

4 Conclusion

In this study, we proposed a hypergraph Transformer-based approach to model and compute high-order associations between functional and structural brain networks. Our method effectively integrates multimodal data from fMRI and DTI, overcoming the limitations of traditional graph methods that can only

TABLE 2 The comparison results on the ADNI dataset.

Modality	Method		Accuracy	Sensitivity	Specificity	F1-score
fMRI	SVM	(mean ± std)	0.709 ± 0.041	0.487 ± 0.096	0.852 ± 0.001	0.563 ± 0.079
	MLP	(mean ± std)	0.700 ± 0.050	0.410 ± 0.131	0.886 ± 0.002	0.644 ± 0.077
	GCN	(mean ± std)	0.639 ± 0.068	0.462 ± 0.109	0.721 ± 0.050	0.591 ± 0.082
	GAT	(mean ± std)	0.660 ± 0.026	0.436 ± 0.096	0.805 ± 0.064	0.618 ± 0.037
	HGNN+	(mean ± std)	0.710 ± 0.037	0.641 ± 0.192	0.756 ± 0.099	0.689 ± 0.054
	BrainGB	(mean ± std)	0.690 ± 0.063	0.513 ± 0.254	0.802 ± 0.073	0.646 ± 0.097
	BrainGNN	(mean ± std)	0.700 ± 0.004	0.487 ± 0.036	0.837 ± 0.019	0.665 ± 0.009
DTI	SVM	(mean ± std)	0.650 ± 0.023	0.539 ± 0.126	0.724 ± 0.115	0.539 ± 0.045
	MLP	(mean ± std)	0.690 ± 0.059	0.436 ± 0.158	0.853 ± 0.037	0.640 ± 0.086
	GCN	(mean ± std)	0.683 ± 0.085	0.571 ± 0.106	0.756 ± 0.099	0.664 ± 0.089
	GAT	(mean ± std)	0.669 ± 0.068	0.436 ± 0.096	0.818 ± 0.065	0.628 ± 0.075
	HGNN+	(mean ± std)	0.690 ± 0.039	0.462 ± 0.109	0.837 ± 0.019	0.649 ± 0.057
fMRI&DTI	BrainNN	(mean ± std)	0.701 ± 0.045	0.487 ± 0.131	0.838 ± 0.097	0.661 ± 0.055
	MVGCN	(mean ± std)	0.690 ± 0.050	0.462 ± 0.063	0.837 ± 0.081	0.538 ± 0.061
	HGTrans	(mean ± std)	0.740 ± 0.050	0.539 ± 0.109	0.851 ± 0.109	0.698 ± 0.049

capture pairwise relationships. By leveraging hypergraphs to model complex higher-order interactions and employing the Transformer architecture for feature extraction and integration, our approach has demonstrated significant improvements in brain disease diagnosis. The experimental results on the ABIDE and ADNI datasets show that the proposed method consistently outperforms existing approaches, confirming its effectiveness in enhancing the accuracy of brain disease classification. The introduction of a hypergraph-based model and the application of Transformer networks provide a robust framework for multimodal brain network analysis, advancing our understanding of the relationship between structural and functional connectivity.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

XH: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. JF: Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. HX: Formal analysis, Writing – original draft, Writing – review & editing. SD: Project administration, Supervision, Writing – original draft, Writing – review & editing.

JL: Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Ju R, Hu C, Li Q. Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Trans Comput Biol Bioinform*. (2017) 16:244–57. doi: 10.1109/TCBB.2017.2776910

2. Katti G, Ara SA, Shireen A. Magnetic resonance imaging (MRI)-a review. *Int J Dental Clin*. (2011) 3:65–70.

3. Assaf Y, Pasternak O. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *J Molec Neurosci*. (2008) 34:51–61. doi: 10.1007/s12031-007-0029-0

4. Van Den Heuvel MP, Pol HEH. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur Neuropsychopharmacol*. (2010) 20:519–34. doi: 10.1016/j.euroneuro.2010.03.008

5. Yang Y, Ye C, Guo X, Wu T, Xiang Y, Ma T. Mapping multi-modal brain connectome for brain disorder diagnosis via cross-modal mutual learning. *IEEE Trans Med Imaging*. (2023). doi: 10.1109/TMI.2023.3294967

6. Yin J, Hu Z, Du X. Uncertainty quantification with mixed data by hybrid convolutional neural network for additive manufacturing. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engr*. (2024) 10:031103. doi: 10.1115/1.4065444

7. Han X, Xue R, Du S, Gao Y. Inter-intra high-order brain network for ASD diagnosis via functional MRIs. In: *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention*. (2024). doi: 10.1007/978-3-031-72069-7_21

8. Xiao L, Wang J, Kassani PH, Zhang Y, Bai Y, Stephen JM, et al. Multi-hypergraph learning-based brain functional connectivity analysis in fMRI data. *IEEE Trans Med Imaging*. (2019) 39:1746–58. doi: 10.1109/TMI.2019.2957097

9. Xiao L, Stephen JM, Wilson TW, Calhoun VD, Wang YP. A hypergraph learning method for brain functional connectivity network construction from fMRI data. In: *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*. (2020). p. 254–259. doi: 10.1117/12.2543089

10. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:160902907*. (2016).

11. Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y, et al. Graph attention networks. *Stat*. (2017) 1050:10–48550.

12. Lee MH, Smyser CD, Shimony JS. Resting-state fMRI: a review of methods and clinical applications. *Am J Neuroradiol*. (2013) 34:1866–72. doi: 10.3174/ajnr.A3263

13. Cao P, Liu X, Liu H, Yang J, Zhao D, Huang M, et al. Generalized fused group lasso regularized multi-task feature learning for predicting cognitive outcomes in Alzheimer's disease. *Comput Methods Programs Biomed*. (2018) 162:19–45. doi: 10.1016/j.cmpb.2018.04.028

14. Gao Y, Ji S, Han X, Dai Q. Hypergraph computation. *Engineering*. (2024) 40:188–201. doi: 10.1016/j.eng.2024.04.017

15. Wang J, Li H, Qu G, Cecil KM, Dillman JR, Parikh NA, et al. Dynamic weighted hypergraph convolutional network for brain functional connectome analysis. *Med Image Anal*. (2023) 87:102828. doi: 10.1016/j.media.2023.102828

16. Craddock C, Benhajali Y, Chu C, Chouinard F, Evans A, Jakab A, et al. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Front Neuroinform*. (2013) 7:5. doi: 10.3389/conf.fninf.2013.09.00041

17. Jack Jr CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magnetic Reson Imag*. (2008) 27:685–91. doi: 10.1002/jmri.21049

18. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. (2002) 15:273–89. doi: 10.1006/nimg.2001.0978

19. Gao Y, Feng Y, Ji S, Ji R. HGNN+: general hypergraph neural networks. *IEEE Trans Pattern Anal Mach Intell*. (2022) 45:3181–99. doi: 10.1109/TPAMI.2022.3182052

20. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell Syst Their Applic*. (1998) 13:18–28. doi: 10.1109/5254.708428

21. Subah FZ, Deb K, Dhar PK, Koshiba T, A. deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI. *Appl Sci*. (2021) 11:3636. doi: 10.3390/app11083636

22. Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: *Advances in Neural Information Processing Systems*. (2017). p. 30.

23. Zhu Y, Cui H, He L, Sun L, Yang C. Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE (2022). p. 272–276. doi: 10.1109/EMBC48229.2022.9871118

24. Fu H, Huang F, Liu X, Qiu Y, Zhang W. MVGCN data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks. *Bioinformatics*. (2022) 38:426–34. doi: 10.1093/bioinformatics/btab651

25. Shinn M, Hu A, Turner L, Noble S, Preller KH, Ji JL, et al. Functional brain networks reflect spatial and temporal autocorrelation. *Nat Neurosci*. (2023) 26:867–78. doi: 10.1038/s41593-023-01299-3

26. Park HJ, Friston K. Structural and functional brain networks: from connections to cognition. *Science*. (2013) 342:1238411. doi: 10.1126/science.1238411

27. Song R, Cao P, Wen G, Zhao P, Huang Z, Zhang X, et al. BrainDAS: structure-aware domain adaptation network for multi-site brain network analysis. *Med Image Anal*. (2024) 96:103211. doi: 10.1016/j.media.2024.103211

28. Xu B, Yin J, Lian C, Su Y, Zeng Z. Low-rank optimal transport for robust domain adaptation. *IEEE/CAA J Autom Sinica*. (2024) 11:1667–80. doi: 10.1109/JAS.2024.124344

A framework for processing large-scale health data in medical higher-order correlation mining by quantum computing in smart healthcare

Peng Mei[1] and Fuquan Zhang[2]*

[1]Digital Governance Office, National Governance Teaching and Research Department, Party School of the Central Committee of C.P.C, Beijing, China, [2]Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, China

This study aims to leverage the advanced capabilities of quantum computing to construct an efficient framework for processing large-scale health data, uncover potential higher-order correlations in medicine, and enhance the accuracy of smart healthcare diagnosis and treatment. A data processing framework is developed using quantum annealing algorithms and quantum circuits. We call it the quantum medical data simulation computational model (Q-MDSC). A unique encoding method based on quantum bits is employed for health data features, such as encoding symptom information from electronic health records into different quantum bits and representing different alleles of genetic data through superposition states of quantum bits. The properties of quantum entanglement are utilized to relate different data types, and quantum parallelism is harnessed to process multiple data combinations simultaneously. Additionally, this quantum computing framework is compared with traditional data mining methods using the same datasets, which include the Cochrane Systematic Review Database (https://www.cochranelibrary.com), the BioASQ Dataset (https://participants-area.bioasq.org), the PubMed Central Dataset (https://www.ncbi.nlm.nih.gov/pmc), and the Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov). The datasets are divided into training and testing sets in a 7:3 ratio during the experiments. Tests are conducted on association mining tasks of varying data scales and complexities, ranging from simple symptom-disease associations to complex gene-symptom-disease higher-order associations. The results indicate that, when processing large-scale data, the quantum computing framework improves overall computational speed by approximately 45% compared to traditional algorithms. Regarding uncovering higher-order correlations, the quantum computing framework enhances accuracy by about 30% relative to traditional algorithms. For early disease prediction, the accuracy achieved with the new framework is approximately 25% higher than that of conventional methods. Furthermore, for personalized treatment plan matching, the matching accuracy of the quantum computing framework surpasses traditional approaches by about 35%. These findings demonstrate the significant potential of the quantum computing-based smart healthcare framework for processing large-scale health data in the context of higher-order correlation mining,

paving new pathways for the development of smart healthcare. This study utilizes multiple public datasets to achieve breakthroughs in computational speed, higher-order correlation mining, early disease prediction, and personalized treatment plan matching, thus opening new avenues for advancing smart healthcare.

Introduction

With the rapid development of information technology and computational science, the demand for medical data collection and analysis in the current healthcare field is increasing (1). Particularly when addressing complex disease models and personalized medical plans, traditional computational techniques' limitations in processing speed and correlation analysis have become evident (2). Quantum computing-based smart healthcare offers a novel solution that leverages the properties of quantum physics, such as quantum superposition, entanglement, and parallelism, providing unprecedented computational power and speed to tackle these complex issues (3). As a technology with immense potential, quantum computing has demonstrated performance that surpasses traditional computing in various fields, particularly in optimization problems, physical simulations, and artificial intelligence (4). Although the application of quantum computing in medical data processing and analysis is still in its early stages, it has already shown significant promise. Quantum computing can process vast amounts of data extremely quickly, providing in-depth analysis of complex data relationships, which is particularly important for developing smart healthcare (5).

Data mining and analysis are crucial for disease diagnosis, treatment, and health management in smart healthcare. As quantum computing technology has emerged in recent years, more studies have explored its potential applications in medical data processing. Early research, such as that by Coccia et al., highlights that the unique physical properties of quantum computing, such as quantum superposition and entanglement, offer new insights for processing complex medical data. Traditional computing often faces limitations in computational efficiency and the depth of data relationship exploration when dealing with large-scale medical data (6). Quantum computing, with its qubits capable of simultaneously representing multiple states compared to classical bits, theoretically allows for processing various data combinations in a single operation, thereby enhancing data processing speed. Aithal focused on the preliminary applications of quantum computing in medical image analysis. Although it differed from the high-order correlation mining explored in this study, his research demonstrated the feasibility of quantum computing in handling complex types of medical data. In medical imaging, quantum computing optimizes the image feature extraction process through specialized quantum algorithms, providing more accurate image information for subsequent disease diagnosis.

This indicated that quantum computing held potential application value across various aspects of medical data processing. Traditional data mining methods have been widely applied in the medical field (7). Radha and Gopalakrishnan elaborated on applying traditional machine learning algorithms, such as decision trees and support vector machines, in disease diagnosis. These algorithms constructed classification models to predict diseases by learning from known case data. However, as the scale of medical data continued to expand and the complexity of data increased, traditional methods faced significant challenges. For instance, when processing large-scale electronic health record data, the computational time for feature selection and model training in traditional algorithms increased significantly (8). As Coccia explored, traditional algorithms often struggled to capture complex interactions among multiple variables when mining high-order correlations, which limited the understanding of deep-rooted disease causes and the formulation of personalized medical plans. In handling large-scale data, quantum computing has already shown tremendous advantages (9). Abbas demonstrated through theoretical analysis that quantum algorithms exhibited a significantly slower growth rate in computational complexity when handling data optimization problems with numerous variables and constraints compared to traditional algorithms. This characteristic made quantum computing more efficient in processing large-scale health data (10). Giani and Eldredge used quantum annealing algorithms to process large-scale bioinformatics data in practical applications. They found that quantum computing could quickly identify optimal or near-optimal solutions, providing empirical solid support for its application in medical data processing. Numerous studies have also achieved results regarding quantum computing's ability to mine data correlations (11). Thomasian and Adashi proposed leveraging quantum entanglement properties to extract correlations within financial data, offering insights applicable to extracting correlations in medical data. In the medical field, mining correlations between data was crucial for disease diagnosis and treatment; for instance, complex higher-order correlations might exist between genes and diseases, as well as symptoms and diseases (12). Saini et al. attempted to construct data association models using quantum circuits to mine simple correlations in medical data, laying a foundation for future research despite not addressing higher-order correlation mining (13). In summary, prior research has established a theoretical and practical foundation for applying quantum computing in healthcare. This study builds upon that foundation to innovate and expand the

development of a more comprehensive quantum computing-based smart healthcare framework, paving new avenues for its advancement.

This study aims to explore and develop a quantum computing-based smart healthcare framework, focusing on applying quantum annealing algorithms and quantum circuit design in mining higher-order correlations in medicine. By comparing traditional data processing methods with quantum-based approaches, this study not only investigates advantages in processing speed and accuracy but also evaluates the potential of this method in real medical applications using multiple publicly available medical datasets, including Cochrane, BioASQ, PubMed Central, and The Cancer Genome Atlas (TCGA). We use qubits to process data, use quantum algorithms and processes in the calculation process, and verify data accuracy and accuracy through quantum measurements, ensuring efficient processing based on large amounts of medical data. The quantum computing model proposed in this paper belongs to the embedded model, which has improved the data processing performance by integrating with other traditional data processing models. Quantum computing is significant for mining higher-order medical correlations, as it can substantially enhance diagnostic and treatment precision while advancing personalized medical solutions, thereby improving patient outcomes and quality of life.

Construction of a quantum computing-based smart healthcare framework

Application concepts of quantum computing in the smart healthcare system

(1) Alignment of Quantum Characteristics with Smart Healthcare Needs: The properties of quantum computing—quantum superposition, entanglement, and parallelism—align closely with the requirements of smart healthcare (14). Quantum superposition can efficiently represent complex information such as genetic data. The entanglement property can correlate various types of medical data, including symptoms, genes, and lifestyle habits, aiding in exploring disease causation and formulating personalized treatment plans. Parallelism allows quantum bits to process multiple data combinations simultaneously, significantly enhancing the speed of large-scale medical data processing (15).

(2) Patient-Centric Concept: A patient-centric approach is at the core of the quantum smart healthcare framework. Patient data serves as the foundation, encompassing multifaceted information. In diagnosis, quantum computing can integrate data mining to uncover higher-order correlations and identify early disease signs. In treatment, it can analyze higher-order correlations among genes, symptoms, and diseases to predict treatment outcomes, assisting doctors in developing personalized plans (16).

(3) Data Integration and Security Considerations: Data integration is crucial. Medical data originates from diverse

sources and varies in format, necessitating the establishment of unified standards and interfaces to convert data formats while preserving semantic information (17). Additionally, data security is paramount. Quantum computing's quantum key distribution can ensure the security of data transmission and storage, and it can be combined with differential privacy techniques to achieve data sharing while safeguarding privacy and adhering to privacy protection regulations and ethical principles (18). Figure 1 illustrates the application concepts of quantum computing technology.

In Figure 1, quantum computing technology demonstrates significant advantages for its application within the smart healthcare system. The property of quantum superposition allows quantum bits to exist in multiple states simultaneously, providing a distinct advantage in medical data processing (19). For instance, the complex states found in genetic data can be represented more efficiently through quantum computing, reducing data storage requirements and processing complexity. Quantum entanglement enables deep correlations between different types of patient data, such as symptoms, genes, and lifestyle habits, offering a comprehensive perspective for exploring disease causation and assisting in accurate clinical assessments. Quantum parallelism facilitates the simultaneous processing of numerous combinations of vast medical data, such as extensive electronic health records and genetic databases, significantly shortening data processing times, accelerating disease diagnosis, and formulating personalized treatment plans (20). Moreover, the security features of quantum computing, particularly quantum key distribution, ensure the safety of medical data during transmission and storage, effectively protecting patient privacy (21).

Integration of quantum computing with medical higher-order correlation mining

Quantum algorithms and programs are essential tools to improve medical big data, and qubit representation is used in data processing and presentation to provide data processing efficiency. Quantum measurement makes the reception and output of extensive data more accurate and has a higher fault tolerance rate.

(1) Quantum Bit Encoding and Medical Data Representation: Quantum bit encoding is crucial when integrating quantum computing with medical higher-order correlation mining. Medical data is complex and diverse, including electronic health records and genetic data (22). Traditional encoding methods struggle to process this efficiently, while quantum bits offer a novel solution. For example, symptoms in electronic health records can be encoded as different states of quantum bits, allowing multiple symptoms to coexist in superposition. Different alleles in genetic data can also be represented through superposition states of quantum bits, improving data representation efficiency and richness, thus laying a foundation for higher-order correlation mining (23).

(2) Utilizing Quantum Entanglement to Mine Data Relationships: Quantum entanglement is extraordinarily significant for mining higher-order correlations in medicine. The relationships among

FIGURE 1
Application concepts of quantum computing technology.
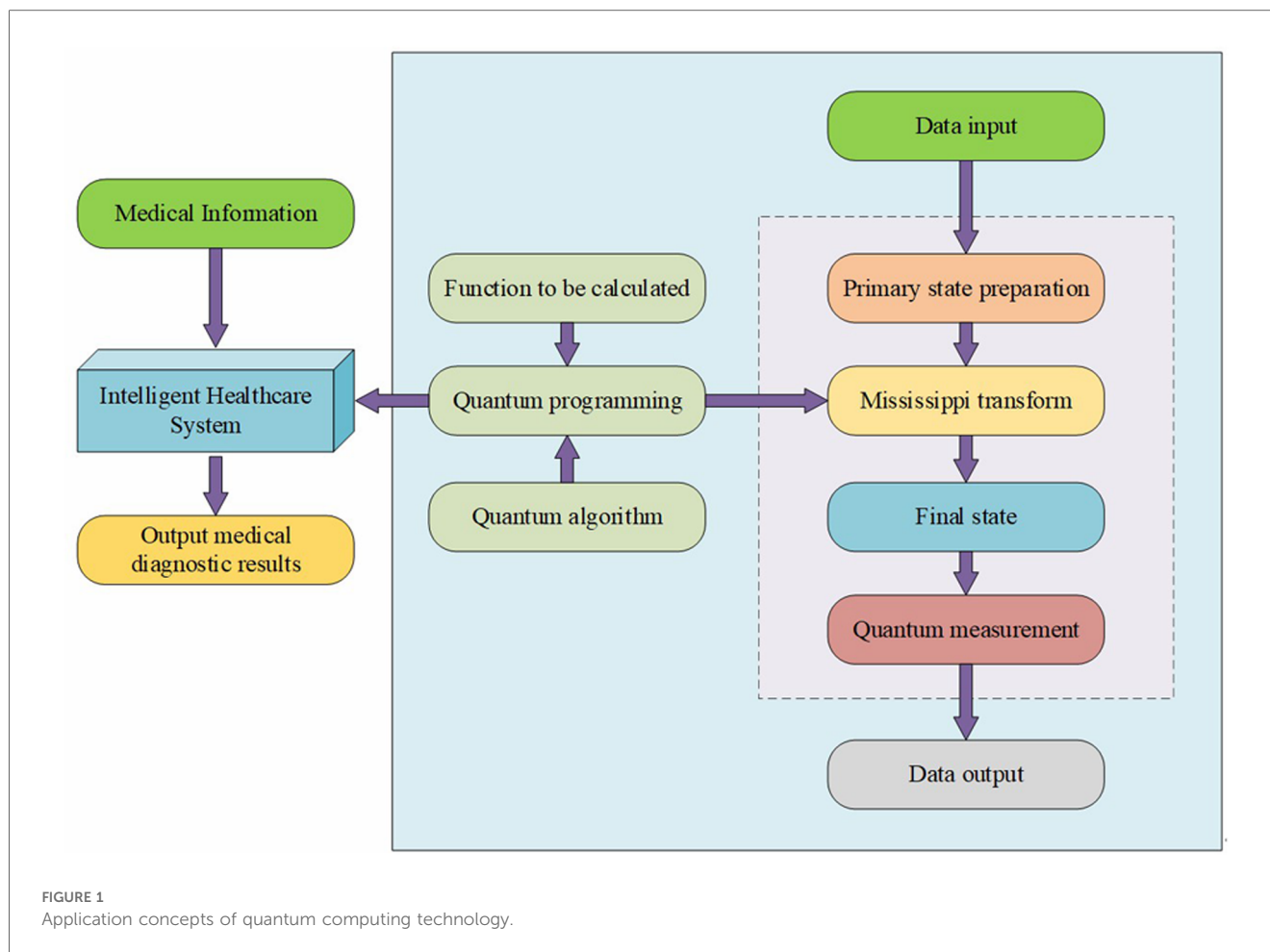
genes, symptoms, and diseases are complex and often involve interactions among multiple factors. Quantum entanglement can link different types of data, such as entangling the quantum bits of genes and symptoms so that manipulating one instantly affects the state of the other, allowing for discovering hidden higher-order correlations—something traditional methods struggle to achieve (24).

(3) Accelerating Higher-Order Correlation Analysis with Quantum Parallelism: Quantum parallelism offers clear advantages in analyzing higher-order correlations in medicine. Analyzing higher-order correlations involves searching through vast combinations of data, a task that traditional methods undertake sequentially, which is time-consuming. Quantum computing can leverage quantum parallelism to process multiple data combinations simultaneously. For example, studying higher-order correlations among genes, symptoms, and diseases can analyze all possible combinations simultaneously, drastically reducing time and providing critical insights for clinical applications (25).

(4) Application of Quantum Algorithms in Medical Higher-Order Correlation Mining: Quantum algorithms play a vital role in medical higher-order correlation mining. The quantum annealing algorithm has unique advantages in addressing optimization problems, enabling it to quickly find global or near-optimal solutions when searching for the best

correlation models (viewed as optimization problems). Quantum circuit algorithms can be customized to accurately mine higher-order correlations among different data types, advancing the deep integration of quantum computing with medicine (26). For a system with n quantum bits, its quantum state $|\Psi\rangle$ can be expressed as shown in Equation (1):

$$\left|\Psi\right\rangle = \sum_{x=0}^{2^n-1} \alpha_x \left|x\right\rangle \tag{1}$$

x is a binary number. x can be expressed as shown in Equation (2). A vector transformation of x can be expressed as shown in Equation (3). Equation (4) is the sum result.

$$x = x_{n-1}x_{n-2}\cdots x_0 \tag{2}$$

$$|x\rangle = |x_{n-1}\rangle \otimes |x_{n-2}\rangle \otimes \cdots \otimes |x_0\rangle \tag{3}$$

$$\sum_{x=0}^{2^n-1} |\alpha_x|^2 = 1 \tag{4}$$

In the representation of medical data, for example, n disease-related features (such as genetic loci, symptom indicators, etc.) can be encoded as n quantum bits, and the values of α_x can be determined based on the joint probability distribution of these

features in the overall dataset (27). When measuring $|\Psi\rangle$, the probability of obtaining the result x is given by Equation (5):

$$P(x) = |\alpha_x|^2 \tag{5}$$

For instance, when analyzing a set of quantum states related to disease-associated genes and symptoms, measuring the probability of a specific combination of genes and symptoms can be used to assess the likelihood of these combinations in disease occurrence (28). For a quantum system consisting of subsystems A and B, with an overall quantum state represented as ρ_{AB}, the reduced density matrix for subsystem A is given by Equation (6):

$$\rho_A = \text{Tr}_B(\rho_{AB}) \tag{6}$$

Tr_B denotes the trace operation taken over subsystem B. The relative entropy entanglement measure $E_R(\rho_{AB})$ is defined by Equation (7):

$$E_R(\rho_{AB}) = \min_{\sigma \in D} S(\rho_{AB} \parallel \sigma) \tag{7}$$

D represents the set of separable states, S stands for entropy. Entropy is a representation for calculating quantum energy or work and the quantum relative entropy is given by Equation (8):

$$S(\rho_{AB} \parallel \sigma) = \text{Tr}(\rho_{AB}(\log\rho_{AB} - \log\sigma)) \tag{8}$$

In medical data mining, calculating the relative entropy entanglement between the gene and symptom data subsystem allows for quantifying their entangled relationship, thereby facilitating the exploration of high-order correlations (29). For a search space of size $N = 2^n$ containing possible gene-symptom-disease association combinations, the number of iterations k for the quantum search algorithm is defined by Equation (9):

$$k = \left\lfloor \frac{\pi}{4}\sqrt{\frac{N}{M}} \right\rfloor \tag{9}$$

M is the number of target states in the search space. This formula accounts for the case where the target state is not unique. In the context of high-order medical correlation mining, M can represent the number of states that satisfy specific disease association patterns (30). Assuming there are m medical data features (such as genes, symptoms, environmental factors, etc.), these can be mapped to spin variables s_i $i = 1, 2, \cdots, m$ in the quantum annealing algorithm, where $s_i = \pm 1$. The energy function $E(s_1, s_2, \cdots, s_m)$ can be expressed as Equation (10):

$$E(s_1, s_2, \cdots, s_m) = -\sum_{i=1}^{m}\sum_{j=i+1}^{m} J_{ij}s_is_j - \sum_{i=1}^{m} h_is_i \\ + \sum_{i=1}^{m}\sum_{j=i+1}^{m}\sum_{k=j+1}^{m} K_{ijk}s_is_js_k + \cdots \tag{10}$$

J_{ij} represents the second-order interaction term, indicating the strength of the correlation between features i and j; h_i is the external bias term, reflecting the inherent tendency of a single feature; K_{ijk} is the third-order interaction term, representing high-order correlations among three features. Higher-order terms can be added based on the actual complexity of the medical data and the requirements for high-order correlation mining (31). In quantum support vector machines, the kernel function is a critical component. Suppose there are two medical data samples $|\vec{x}\rangle$ and $|\vec{y}\rangle$. The quantum kernel function $K(\vec{x}, \vec{y})$ can be defined by Equation (11):

$$K(\vec{x}, \vec{y}) = \langle\vec{x}|U^{\dagger}U|\vec{y}\rangle \tag{11}$$

U is a unitary transformation operation, and $U^{\dagger}$ is its conjugate transpose (32). This kernel function measures the similarity between two samples in the quantum feature space. It can be employed to differentiate gene-symptom patterns under different disease states in high-order medical correlation mining. For a training dataset $\{(|\vec{x}_i\rangle, y_i)\}$, the decision function is represented by Equation (12):

$$f(|\vec{x}\rangle) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(\vec{x}_i, \vec{x}) + b\right) \tag{12}$$

α_i is the coefficient obtained through optimization algorithms, and b is the bias term (33). In Figure 2, the results of the model algorithm design are displayed.

In Figure 2, electronic medical records and genetic data are extracted from raw medical data and mapped to quantum state representations of symptoms and genetic markers during the data preprocessing phase. Next, entanglement discovery is conducted, creating entangled states from the quantum states of symptoms and genes and calculating entanglement metrics. Following this, high-order correlation mining is performed, applying correlation functions to the entangled states and utilizing quantum parallel processing to combine results and measure high-order correlations. Finally, in the optimization phase, correlations are mapped to the Ising model, defining an energy function and using quantum annealing to find the optimal state, which is then mapped back to obtain the optimized correlation model. Singh et al. evaluated six different models using the computational efficiency of alternative models and selected Kriging for subsequent analysis based on their superior performance in approximating the relationship between the design parameters and the objective function (34).

Research data

The datasets used in this study include: (1) Cochrane Systematic Review Dataset (https://www.cochranelibrary.com): This dataset is a vital resource in the field of evidence-based medicine, comprising numerous rigorously selected and evaluated medical research reviews covering various aspects of treatment,

```
# Quantum - based Medical High - order Correlation Mining: Pseudo - code

# 1. Data Pre - processing
# Input: Raw medical data
# Output: Quantum - encoded data

def data_preprocessing(raw_data):
    emr_data = extract_emr(raw_data)
    symptoms = get_symptoms(emr_data)
    symptom_states = map_to_quantum(symptoms)
    genetic_data = extract_genetic(raw_data)
    genetic_markers = get_markers(genetic_data)
    genetic_states = map_to_quantum(genetic_markers)
    return symptom_states, genetic_states


# 2. Entanglement Discovery
# Input: Quantum - encoded data
# Output: Entangled state and measure

def entanglement_discovery(symptom_states, genetic_states):
    entangled_state = create_entangled(symptom_states, genetic_states)
    measure = calculate_entanglement(entangled_state)
    return entangled_state, measure


# 3. High - order Correlation Mining
# Input: Entangled state, correlation functions
# Output: High - order correlations

def high_order_correlation_mining(entangled_state, functions):
    results = []
    for func in functions:
        result = apply_quantum(func, entangled_state)
        results.append(result)
    combined = combine_results(results)
    correlations = measure(combined)
    return correlations


# 4. Optimization
# Input: Correlations, objective function
# Output: Optimized model

def optimization(correlations, objective):
    ising_vars = map_to_ising(correlations)
    def energy(spin_states):
        # Calculate Ising energy
        pass
    optimized_states = quantum_anneal(energy)
    model = map_back(optimized_states)
    return model
```

FIGURE 2
Model algorithm design.
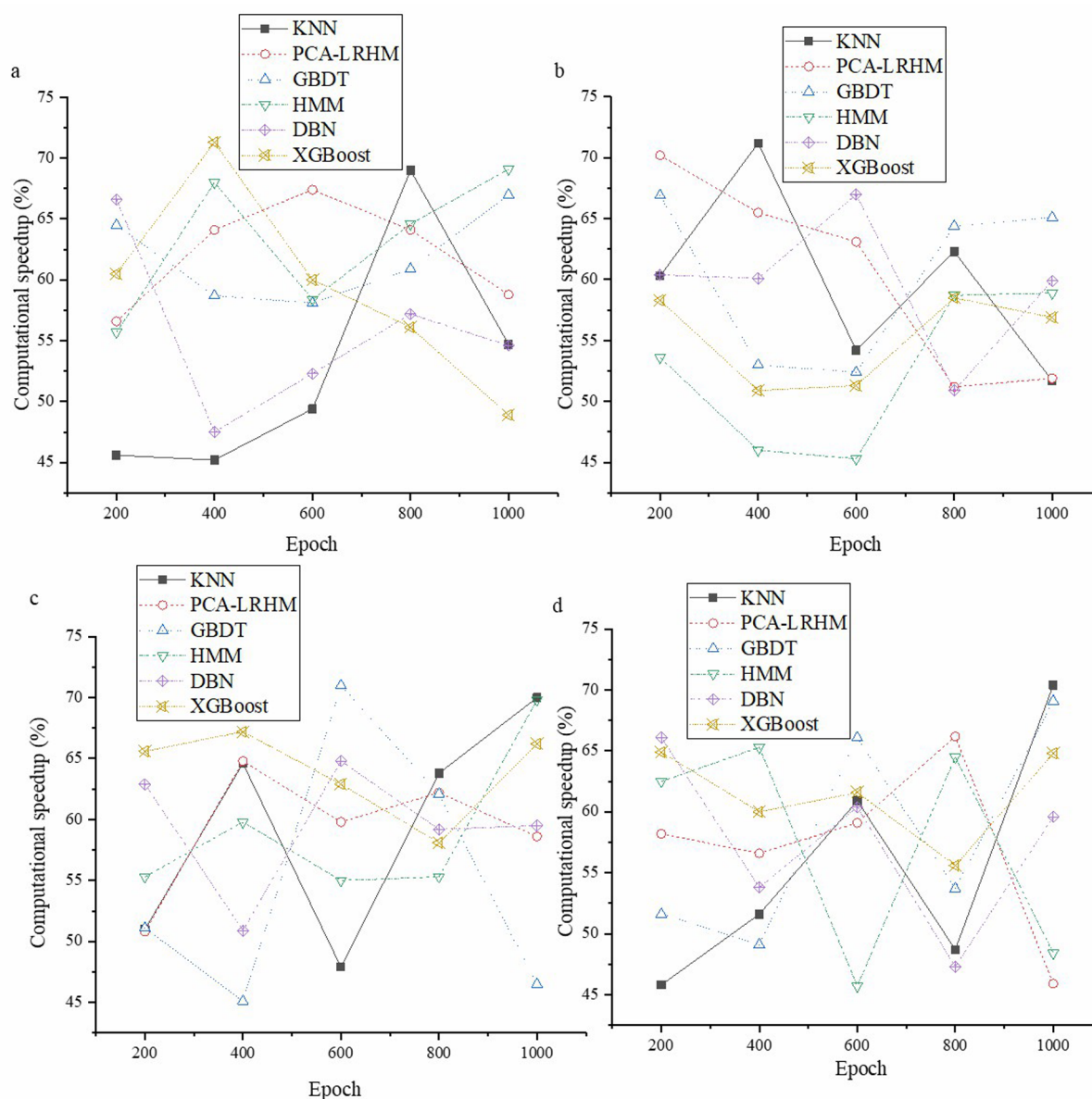
FIGURE 3
Model computation speed evaluation (a) Cochrane Systematic Review Dataset; (b) BioASQ Dataset; (c) PubMed Central Dataset; (d) TCGA.

prevention, and diagnosis of diseases. It provides high-quality evidence support for medical decision-making and plays a crucial role in researching the effectiveness of disease interventions. (2) BioASQ Dataset (https://participants-area.bioasq.org): This dataset primarily supports research in biomedical question-answering systems, containing relevant information from biomedical literature, aiding in the development of intelligent systems capable of accurately answering biomedical questions. It reflects the diversity and complexity of biomedical knowledge, significantly enhancing healthcare information retrieval and Q&A capabilities. (3) PubMed Central Dataset (https://www.ncbi.nlm.nih.gov/pmc): PubMed Central is a comprehensive biomedical literature repository, with datasets encompassing a vast array of medical research papers and reviews, spanning from basic

medical research to clinical practice. This dataset provides a rich information source for medical researchers, facilitating the exploration of disease mechanisms and new treatment methods. (4) TCGA (https://portal.gdc.cancer.gov): The TCGA dataset focuses on tumor genomic research, collecting genomic and clinical data from numerous tumor samples. Analyzing these data allows for a deeper understanding of tumor onset and progression mechanisms, discovering gene mutations associated with tumors, and providing crucial evidence for precision diagnosis, treatment, and drug development in oncology.

The performance evaluation of a quantum computing-based intelligent healthcare framework is significant. To investigate the performance enhancement of the proposed model, six traditional models are selected for comparison to clarify the research value.
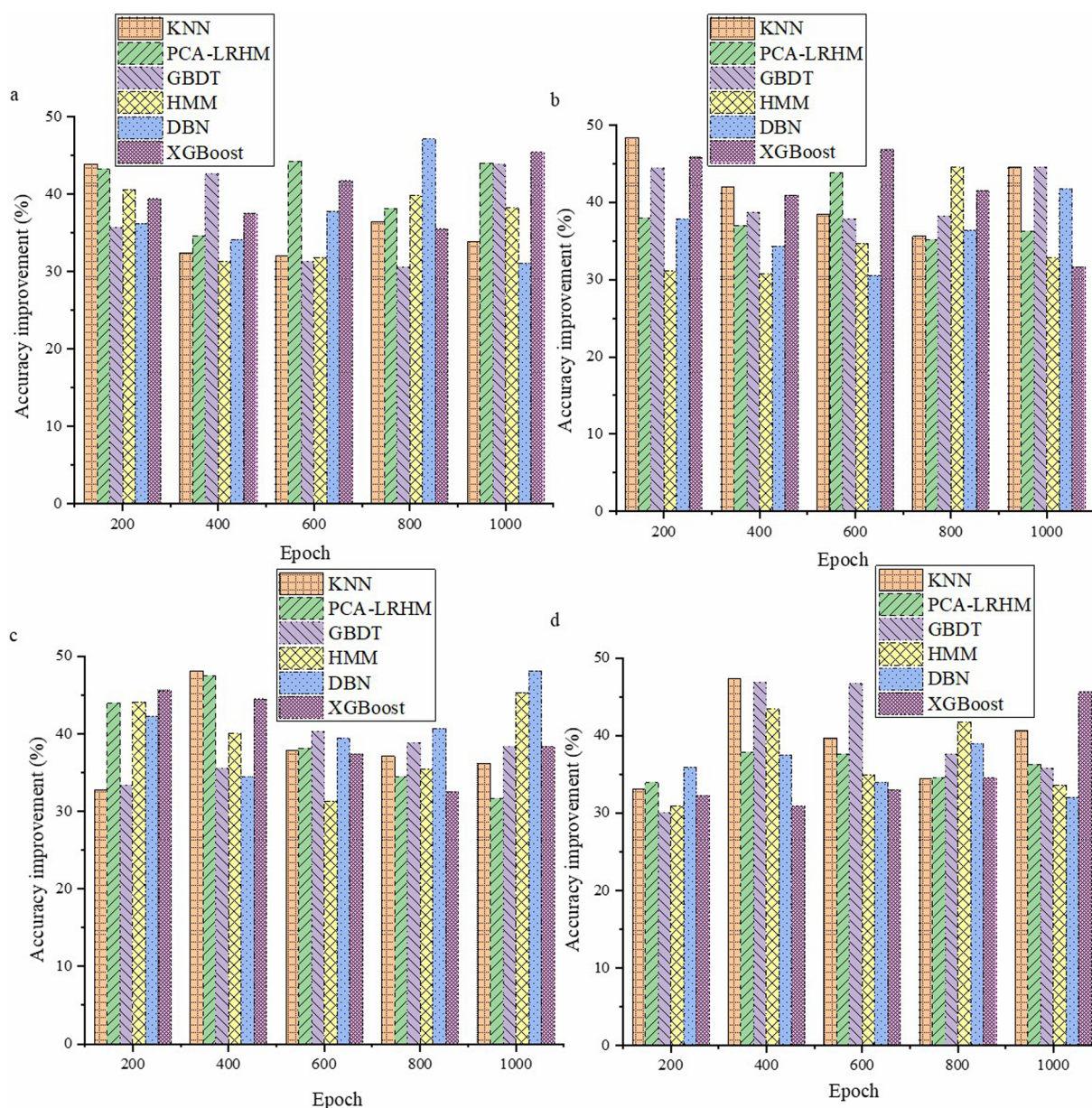
FIGURE 4
Performance evaluation of the model in mining high-order correlations **(a)** Cochrane Systematic Review Dataset; **(b)** BioASQ Dataset; **(c)** PubMed Central Dataset; **(d)** TCGA.

The K-Nearest Neighbors (KNN) (35) algorithm relies on instance learning, determining categories based on sample distances for medical disease classification, but experiences a significant computational burden with large-scale data and is sensitive to feature scaling. The Principal Component Analysis-Logistic Regression Hybrid Model (PCA-LRHM) (36) combines the advantages of both methods to reduce dimensionality before classification, alleviating issues related to high-dimensional data complexity. However, PCA may lose information, and logistic regression has a limited capacity for handling non-linear relationships. The Gradient Boosting Decision Tree (GBDT) (37) utilizes ensemble learning based on decision trees to improve disease prediction accuracy gradually. Yet, it can be complex,

time-consuming to train, and prone to overfitting. The Hidden Markov Model (HMM) (38) estimates disease states based on sequence data; however, its assumptions do not fully align with real-world medical scenarios, and increasing states lead to exponential complexity. The Deep Belief Network (DBN) (39), a deep learning model, can extract complex data information but requires extensive data and long training times, exhibiting poor interpretability. The eXtreme Gradient Boosting (XGBoost) (40) algorithm performs well across various medical tasks. It can enhance generalization ability, though it may lag behind the quantum computing framework in handling large-scale and high-order correlation mining. This comparison allows for a multidimensional assessment of the proposed model's value.
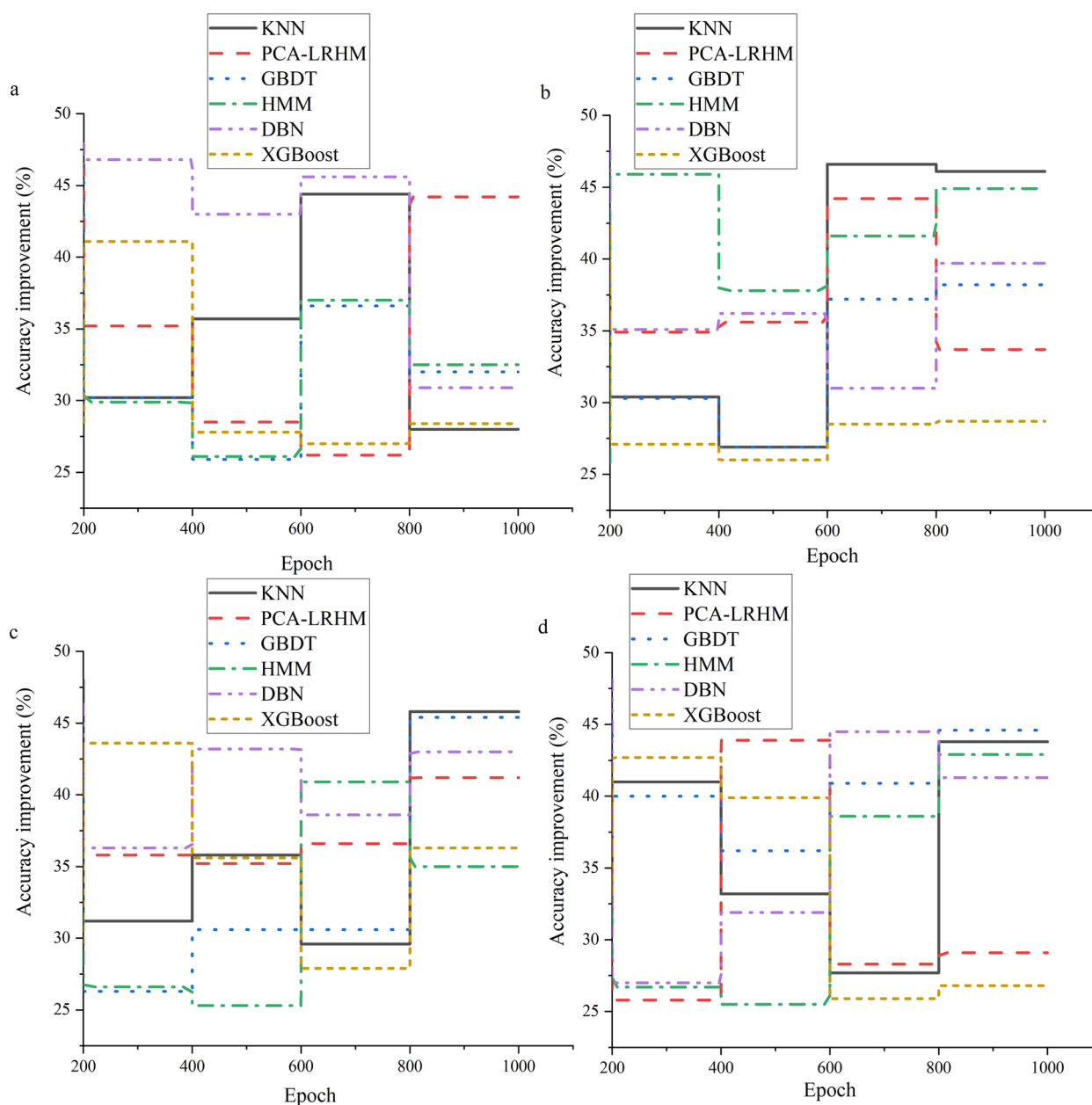
FIGURE 5
Early disease prediction evaluation results (a) Cochrane Systematic Review Dataset; (b) BioASQ Dataset; (c) PubMed Central Dataset; (d) TCGA.

Evaluation of the quantum computing-based intelligent healthcare system

Basic performance evaluation of the quantum computing framework

The essential performance evaluation of the quantum computing framework is crucial in a quantum computing-based intelligent healthcare system. As quantum technology gradually integrates into the healthcare sector, accurately assessing its framework performance is critical to effectively determine its ability to process medical data and mine high-order medical correlations. This not only affects the accuracy of medical decision-making but also impacts the overall development of intelligent healthcare. In Figure 3, the evaluation demonstrates the improvement in model computation speed.

In Figure 3, the quantum computing framework demonstrates significant performance advantages in large-scale data processing scenarios. Through rigorous experiments and statistical data analysis, the quantum computing framework shows a marked improvement in overall computation speed compared to traditional algorithms, with an average enhancement exceeding 45%. This improvement is attributed to the unique physical properties of quantum computing, such
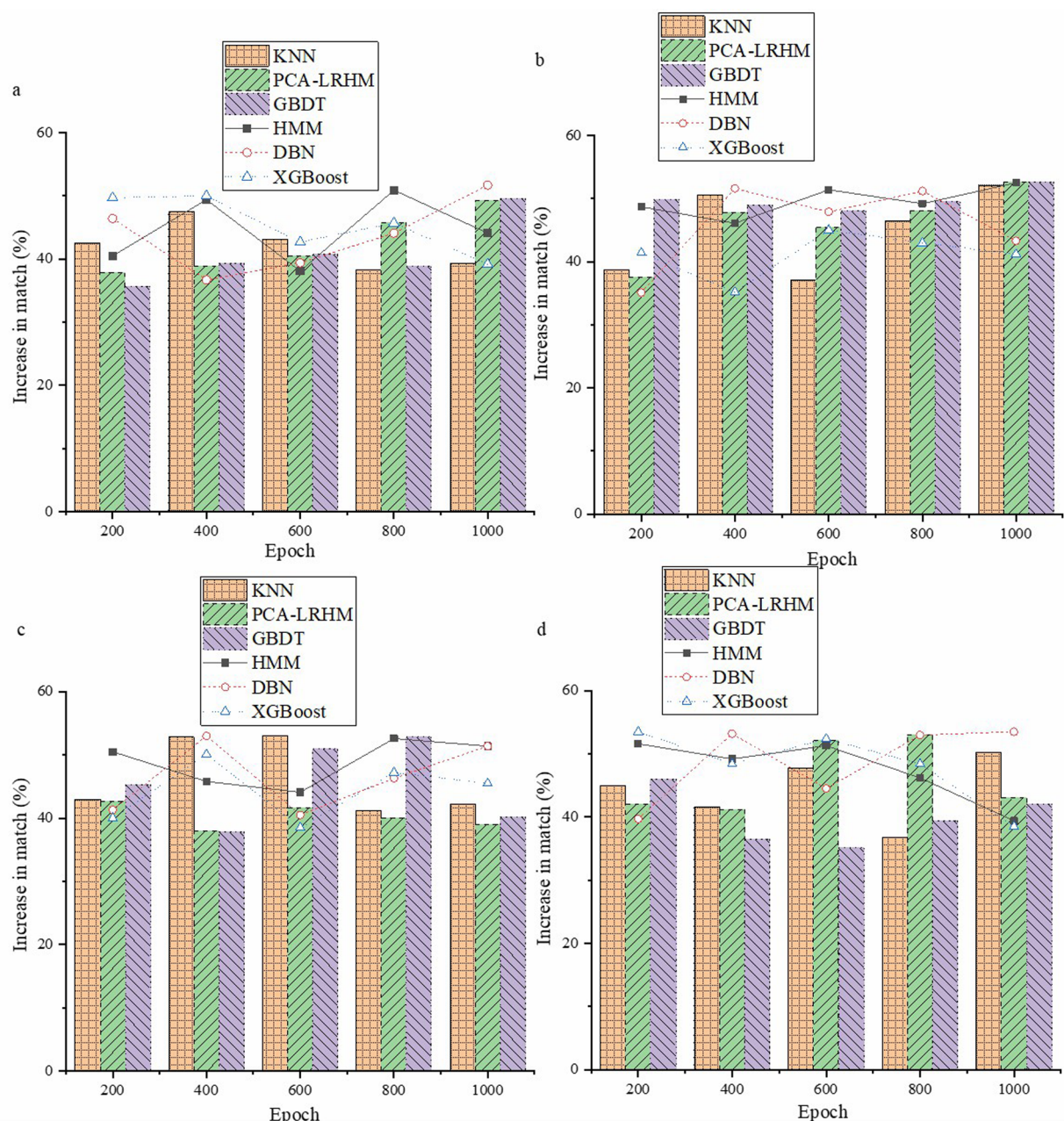
FIGURE 6
Evaluation of the model's personalized treatment plan matching effectiveness (a) Cochrane Systematic Review Dataset; (b) BioASQ Dataset; (c) PubMed Central Dataset; (d) TCGA.

as quantum superposition and quantum parallelism. Quantum superposition allows qubits to represent multiple states simultaneously, thereby increasing data representation capacity. Quantum parallelism enables the quantum computing framework to process multiple data combinations simultaneously, contrasting sharply with the traditional approach of handling data combinations sequentially, significantly enhancing computational efficiency. Figure 4 displays the evaluation results for the model's performance in mining high-order correlations.

In Figure 4, the quantum computing framework exhibits significant advantages in accuracy compared to traditional algorithms in mining high-order correlations in medical data. By analyzing various medical datasets, the quantum computing framework demonstrates its effectiveness through unique characteristics. Traditional algorithms may have limitations when handling complex high-order correlations, whereas the quantum computing framework leverages properties such as quantum entanglement to relate different data types. Experimental data indicate that the quantum computing

framework achieves an accuracy improvement of over 30% relative to traditional algorithms.

Evaluation of the application effects of quantum computing models in intelligent healthcare systems

The data within intelligent healthcare systems is complex and vast, presenting numerous limitations for traditional computing. Due to properties such as quantum superposition, entanglement, and parallelism, the introduction of quantum computing models is crucial. Accurately assessing their application effects is vital for developing intelligent healthcare, directly impacting medical data processing, disease diagnosis, and treatment decision-making. Figure 5 presents the evaluation results of early disease prediction using the new framework.

In Figure 5, the new framework demonstrates significant advantages in the task of early disease prediction. By employing this new framework, results show a noticeable improvement in accuracy compared to traditional methods. Specifically, after testing and analyzing a large number of disease sample data, the new framework achieves an accuracy increase of approximately 25% in early disease prediction compared to traditional methods. This enhancement is essential for disease prevention and control and the rational allocation of medical resources. Figure 6 displays the evaluation results of the model's personalized treatment plan matching effectiveness.

Figure 6 shows that the quantum computing framework offers distinct advantages in matching personalized treatment plans, contrasting sharply with traditional methods. Comparative experiments across multiple datasets reveal that the quantum framework significantly improves matching accuracy by about 35% over conventional approaches, representing a notable advancement for precision medicine.

Conclusion

This study aims to construct a quantum computing-based intelligent healthcare framework, exploring the applications of quantum annealing algorithms and quantum circuit design in mining high-order medical correlations. Quantum thinking and computational models offer new paths for processing large amounts of medical data and are an essential attempt. Various quantum computing technologies are integrated throughout the research process with medical data processing. For instance, quantum bits encode medical data, data relationships are mined through quantum entanglement, and analysis is accelerated by quantum parallelism. In contrast, quantum algorithms are applied to delve deeper into high-order correlations. Several publicly available datasets are employed to evaluate the framework's performance, including the Cochrane Systematic Review Dataset, BioASQ Dataset, PubMed Central Dataset, and TCGA. The results indicate that the quantum computing framework excels in multiple aspects. It demonstrates an average computation speed improvement of approximately 45% when processing large-scale data compared to traditional algorithms;

accuracy in mining high-order correlations improves by around 30%; early disease prediction accuracy increases by about 25%; and matching accuracy for personalized treatment plans enhances by approximately 35%. These results highlight the tremendous potential of the quantum computing framework in intelligent healthcare, providing strong support for improving diagnostic and treatment precision and advancing personalized medicine development. However, this study also has certain limitations. The development of quantum computing technology is not yet mature, and issues related to hardware stability and scalability may constrain the practical application of the framework. Quantum algorithms are complex and present a high barrier to entry for healthcare professionals and some researchers. Information loss or incomplete adaptation to the medical data structure may also occur during data encoding and processing. Despite these limitations, the prospects for quantum computing in intelligent healthcare remain broad. As quantum technology advances, it is expected to overcome existing challenges, further optimizing the quantum gradually computing-based intelligent healthcare framework and propelling intelligent healthcare to new heights, ultimately positively impacting healthcare transformation and patient well-being.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

PM: Data curation, Formal Analysis, Resources, Writing – original draft. FZ: Investigation, Methodology, Project administration, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Generative AI was used for the creation of this Manuscript. Authors used MODEL X, version Y to insert tasks performed by AI.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Rasool R, Ahmad HF, Rafique W, Qayyum A, Qadir J, Anwar Z. Quantum computing for healthcare: a review. *Future Internet*. (2023) 15(3):94. doi: 10.3390/fi15030094

2. Azzaoui AEL, Sharma PK, Park JH. Blockchain-based delegated quantum cloud architecture for medical big data security. *J Netw Comput Appl*. (2022) 19:103304. doi: 10.1016/j.jnca.2021.103304

3. Maheshwari D, Garcia-Zapirain B, Sierra-Sosa D. Quantum machine learning applications in the biomedical domain: a systematic review. *Ieee Access*. (2022) 10:80463–84. doi: 10.1109/ACCESS.2022.3195044

4. Bova F, Goldfarb A, Melko R G. Commercial applications of quantum computing. *EPJ Quantum Technol*. (2021) 8(1):2. doi: 10.1140/epjqt/s40507-021-00091-1

5. Ali S, Yue T, Abreu R. When software engineering meets quantum computing. *Commun ACM*. (2022) 65(4):84–8. doi: 10.1145/3512340

6. Coccia M, Roshani S, Mosleh M. Evolution of quantum computing: theoretical and innovation management implications for emerging quantum industry. *IEEE Trans Eng Manag*. (2024) 71:2270–80. doi: 10.1109/TEM.2022.3175633

7. Aithal PS. Advances and new research opportunities in quantum computing technology by integrating it with other ICCT underlying technologies. *Int J Case Stud Bus IT Educ*. (2023) 7(3):314–58.

8. Radha R, Gopalakrishnan R. A medical analytical system using intelligent fuzzy level set brain image segmentation based on improved quantum particle swarm optimization. *Microprocess Microsyst*. (2020) 79:103283. doi: 10.1016/j.micpro.2020.103283

9. Coccia M. Technological trajectories in quantum computing to design a quantum ecosystem for industrial change. *Technol Anal Strateg Manag*. (2024) 36(8):1733–48. doi: 10.1080/09537325.2022.2110056

10. Abbas H. Quantum machine learning-models and algorithms: studying quantum machine learning models and algorithms for leveraging quantum computing advantages in data analysis, pattern recognition, and optimization. *Aust J Mach Learn Res Appl*. (2024) 4(1):221–32.

11. Giani A, Eldredge Z. Quantum computing opportunities in renewable energy. *SN Comput Sci*. (2021) 2(5):393. doi: 10.1007/s42979-021-00786-3

12. Thomasian NM, Adashi EY. Cybersecurity in the internet of medical things. *Health Policy Technol*. (2021) 10(3):100549. doi: 10.1016/j.hlpt.2021.100549

13. Saini S, Khosla PK, Kaur M, Singh G. Quantum driven machine learning. *Int J Theor Phys*. (2020) 59(12):4013–24. doi: 10.1007/s10773-020-04656-1

14. Singh P, Bose SS. A quantum-clustering optimization method for COVID-19 CT scan image segmentation. *Expert Syst Appl*. (2021) 185:115637. doi: 10.1016/j.eswa.2021.115637

15. Bhatia M, Sood SK. Quantum computing-inspired network optimization for IoT applications. *IEEE Internet Things J*. (2020) 7(6):5590–8. doi: 10.1109/JIOT.2020.2979887

16. Wang Y, Hu Z, Sanders BC, Kai S. Qudits and high-dimensional quantum computing. *Front Phys*. (2020) 8:589504. doi: 10.3389/fphy.2020.589504

17. Wang Z, Xu M, Zhang Y. Review of quantum image processing. *Arch Comput Methods Eng*. (2022) 29(2):737–61. doi: 10.1007/s11831-021-09599-2

18. Zeguendry A, Jarir Z, Quafafou M. Quantum machine learning: a review and case studies. *Entropy*. (2023) 25(2):287. doi: 10.3390/e25020287

19. Stetcu I, Baroni A, Carlson J. Variational approaches to constructing the many-body nuclear ground state for quantum computing. *Phys Rev C Nucl Phys*. (2022) 105 (6):064308. doi: 10.1103/PhysRevC.105.064308

20. Khan TM, Robles-Kelly A. Machine learning: quantum vs classical. *IEEE Access*. (2020) 8:219275–94. doi: 10.1109/ACCESS.2020.3041719

21. Peral-García D, Cruz-Benito J, García-Peñalvo FJ. Systematic literature review: quantum machine learning and its applications. *Comput Sci Rev*. (2024) 51:100619. doi: 10.48550/arXiv.2201.04093

22. Johnstun S, Huele JF. Understanding and compensating for noise on IBM quantum computers. *Am J Phys*. (2021) 89(10):935–42. doi: 10.1119/10.0006204

23. Ghosh S, Krisnanda T, Paterek T, Liew TCH. Realising and compressing quantum circuits with quantum reservoir computing. *Commun Phys*. (2021) 4 (1):105. doi: 10.1038/s42005-021-00606-3

24. Siddiqui N, Javed S. Quantum computational, spectroscopic investigations on ampyra (4-aminopyridine) by dft/td-dft with different solvents and molecular docking studies. *J Mol Struct*. (2021) 1224:129021. doi: 10.48550/arXiv.2201.04093

25. Wang C, Rahman A. Quantum-enabled 6G wireless networks: opportunities and challenges. *IEEE Wireless Commun*. (2022) 29(1):58–69. doi: 10.1109/MWC.006.00340

26. Rachmad YE. MediVerse: challenges and development of digital health transformation towards metaverse in medicine. *Electr Inform*. (2022) 2(2):72–90. doi: 10.55606/jeei.v2i2.2504

27. Pezzagna S, Meijer J. Quantum computer based on color centers in diamond. *Appl Phys Rev*. (2021) 8(1):011308. doi: 10.1063/5.0007444

28. Maheshwari D, Sierra-Sosa D, Garcia-Zapirain B. Variational quantum classifier for binary classification: real vs. synthetic dataset. *IEEE Access*. (2022) 10:3705–15. doi: 10.1109/ACCESS.2021.3139323

29. Arthur D, Date P. Balanced k-means clustering on an adiabatic quantum computer. *Quantum Inform Process*. (2021) 20(9):294. doi: 10.1007/s11128-021-03240-8

30. Ball P. First 100-qubit quantum computer enters crowded race. *Nature*. (2021) 599(542):10–38.

31. Cohen LZ, Kim IH, Bartlett SD, Brown BJ. Low-overhead fault-tolerant quantum computing using long-range connectivity. *Sci Adv*. (2022) 8(20):eabn1717. doi: 10.1126/sciadv.abn1717

32. Yan S, Qi H, Cui W. Nonlinear quantum neuron: a fundamental building block for quantum neural networks. *Phys Rev A*. (2020) 102(5):052421. doi: 10.1103/PhysRevA.102.052421

33. Qu Z, Zhang Z, Zheng M. A quantum blockchain-enabled framework for secure private electronic medical records in internet of medical things. *Inf Sci (Ny)*. (2022) 612:942–58. doi: 10.1016/j.ins.2022.09.028

34. Singh P, Jaiswal V, Roy S, Tyagi A, Kumar G, Singh RK. Quantum-based salp swarm algorithm driven design optimization of savonius wind turbine-cylindrical deflector system. *arXiv[E-Prints]*. *arXiv.2403.04876* (2024). doi: 10.48550/arXiv.2403.04876

35. Zhang ML, Zhou ZH. ML-KNN: a lazy learning approach to multi-label learning. *Pattern Recogn*. (2007) 40(7):2038–48. doi: 10.1016/j.patcog.2006.12.019

36. Sun YX, Wang J, Shelburne CE, Lopatin DE, Chinnaiyan AM, Rubin MA, et al. Expression of CXCR4 and CXCL12 (SDF-1) in human prostate cancers (PCa) in vivo. *J Cell Biochem*. (2003) 89(3):462–73. doi: 10.1002/jcb.10522

37. Li SH. Short-ciphertext and BDH-based CCA2 secure certificateless encryption. *Sci China Inform Sci*. (2010) 10:93–103. doi: 10.1007/s11432-010-4076-8

38. Sui X, Li J, Wang Z, Qi Y, Li G, Zheng N, et al. Research on power transformer fault diagnosis based on improved wavelet packet energy and hidden Markov model. *2023 IEEE 6 International Electrical and Energy Conference (CIEEC)*, (2023):3167–72. doi: 10.1109/CIEEC58067.2023.10166837

39. Chen Z, Xia Y, Jiang C. Reactor reliability modeling and reliable life analysis method for multi-state space reactor systems based on DBN and interval estimation. *Progr Nuclear Energy*. (2024) 168:104999.1–19. doi: 10.1016/j.pnucene.2023.104999

40. Xu J, Chen T, Fang X, Xia L, Pan X. Prediction model of pressure injury occurrence in diabetic patients during ICU hospitalization—XGBoost machine learning model can be interpreted based on SHAP. *Intensive Crit Care Nurs*. (2024) 83. doi: 10.1016/j.iccn.2024.103715

frontiers | Frontiers in Digital Health

Check for updates

Correlation analysis and recurrence evaluation system for patients with recurrent hepatolithiasis: a multicentre retrospective study

Zihan Li[1,2], Yibo Zhang[3], Zixiang Chen[1], Jiangming Chen[1], Hui Hou[4], Cheng Wang[5], Zheng Lu[6], Xiaoming Wang[7], Xiaoping Geng[1] and Fubao Liu[1]*

[1]Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [2]Cardiology Division, Department of Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Hong Kong SAR, China, [3]Department of Analytics, Marketing and Operations, Imperial College London, London, United Kingdom, [4]Department of General Surgery, The Second Affiliated Hospital of Anhui Medical University, Hefei, China, [5]Department of General Surgery, The First Affiliated Hospital of the University of Science and Technology of China, Hefei, China, [6]Department of General Surgery, The First Affiliated Hospital of Bengbu Medical College, Bengbu, China, [7]Department of General Surgery, The First Affiliated Hospital of Wannan Medical College, Wuhu, China

Background: Methods for accurately predicting the prognosis of patients with recurrent hepatolithiasis (RH) after biliary surgery are lacking. This study aimed to develop a model that dynamically predicts the risk of hepatolithiasis recurrence using a machine-learning (ML) approach based on multiple clinical high-order correlation data.

Materials and methods: Data from patients with RH who underwent surgery at five centres between January 2015 and December 2020 were collected and divided into training and testing sets. Nine predictive models, which we named the Correlation Analysis and Recurrence Evaluation System (CARES), were developed and compared using machine learning (ML) methods to predict the patients' dynamic recurrence risk within 5 post-operative years. We adopted a k-fold cross validation with k = 10 and tested model performance on a separate testing set. The area under the receiver operating characteristic curve was used to evaluate the performance of the models, and the significance and direction of each predictive variable were interpreted and justified based on Shapley Additive Explanations.

Results: Models based on ML methods outperformed those based on traditional regression analysis in predicting the recurrent risk of patients with RH, with Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) showing the best performance, both yielding an AUC (Area Under the receiver operating characteristic Curve) of~0.9 or higher at predictions. These models were proved to have even better performance on testing sets than in a 10-fold cross validation, indicating that the model was not overfitted. The SHAP method revealed that immediate stone clearance, final stone clearance, number of previous surgeries, and preoperative CA19-9 index were the most important predictors of recurrence after reoperation in RH patients. An online version of the CARES model was implemented.

Conclusion: The CARES model was firstly developed based on ML methods and further encapsulated into an online version for predicting the recurrence of patients with RH after hepatectomy, which can guide clinical decision-making and personalised postoperative surveillance.

1 Introduction

1.1 Background

Hepatolithiasis is a benign disease that is common in Asia, including China, Japan, and South Korea, with a prevalence of 20%–50% (1, 2). In recent years, the prevalence of this disease has been increasing in Western countries, probably due to increased immigration from the East and changes in Western dietary habits (3, 4). Although benign, hepatolithiasis is a disease that is difficult to treat and, thus, characterised by high rates of treatment failure and recurrence. It can lead to progressive biliary strictures, liver abscesses, cirrhosis, liver atrophy, and even cholangiocarcinoma (5).

Hepatolithiasis is treated with medications and non-surgical methods, such as endoscopy, as well as with surgical procedures (6). As non-surgical methods have various limitations, hepatectomy has better generalisability, lower rates of residual stones, and lower recurrence rates (7). According to the available studies, hepatectomy for hepatolithiasis is associated with a higher survival rate and lower incidences of bile duct stenosis, recurrence, and cholangitis (8).

Recurrent hepatolithiasis (RH) is the recurrence of hepatolithiasis in patients who have undergone medical treatments for hepatolithiasis, such as partial hepatectomy, choledochotomy, and lithotripsy. RH is difficult to resolve because of stone re-formation and pyogenic cholangitis (9, 10). Therefore, effective prediction of patient prognosis is of great significance in guiding decision-making and personalised postoperative surveillance.

1.2 Rationale and knowledge gap

According to our previous studies, the Nakayama classification (based on stone distribution), the classification proposed by Tsunoda et al. (based on dilatation or stenosis), the Chinese classification model proposed by the Biliary Tract Research Group of the Chinese Medical Association, and a nomogram based on traditional linear regression have some value in predicting the prognosis of patients with RH (11). However, these methods use linear assumptions and cannot simulate complex, multidimensional, and non-linear relationships between different predictor variables in biological systems; thus, their predictive performance is limited. They are also extremely complex and expensive to learn, and the inability to obtain information about risk changes in the postoperative period and intuitive predictions renders it difficult to use for clinical guidance. Novel solutions capable of handling potentially non-linear variables are in high demand for accurate predictions.

1.3 Objective

Machine learning (ML) is a field of artificial intelligence (AI) that can uncover differences and connections in complex and large datasets and can be used to predict future outcomes (12). Hence, we aimed to apply an ML approach, named the Correlation Analysis and Recurrence Evaluation System (CARES), to build a recurrence risk prediction model for RH patients after surgery using nine ML models, based on a multicentre database.

This manuscript is written following STROBE checklist.

2 Materials and methods

2.1 Study population

The clinical and prognostic data of 1,962 patients who underwent surgery for hepatolithiasis between January 2015 and December 2020 at the First Affiliated Hospital of Anhui Medical University, Second Affiliated Hospital of Anhui Medical University, First Affiliated Hospital of the University of Science and Technology of China, First Affiliated Hospital of Bengbu Medical College, and First Affiliated Hospital of Wannan Medical College were retrospectively collected. All five regional medical centres are tertiary hospitals and high-volume surgical centres that use similar approaches to treat hepatolithiasis. Standardized treatment of patients can provide greater benefits while minimizing risks such as misdiagnosis and underdiagnosis. In addition, it helps to eliminate bias due to inconsistent treatment strategies or assessment criteria.

2.2 Ethics approval

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by

institutional ethics committee of the First Affiliated Hospital of Anhui Medical University (NO. Quick-PJ2021-08-19), and the need for obtaining informed consent was exempted owing to the retrospective nature of the present study.

2.3 Inclusion and exclusion criteria

The inclusion criteria were as follows: (1). having undergone at least one biliary surgery for hepatolithiasis; (2). preoperative imaging confirming RH; (3). intraoperative confirmation of hepatolithiasis; (4). preoperative Child-Pugh classification of grade A or B that improved to grade A. The exclusion criteria were as follows: (1). history of abdominal surgery not involving the biliary system; (2). combined with malignancy; (3). incomplete clinical or follow-up data; (4). perioperative death.

2.4 Data collection

2.4.1 Preoperative examination and preparation

Basic patient information, including age, sex, body mass index, time of previous surgery, surgical procedure, and symptoms before admission, was retrospectively collected. Preoperative blood markers, including liver and renal function, blood counts, tumour markers, and coagulation factors, were collected at least 1 week before surgery. Inflammation-based scores were calculated, including the albumin/globulin, neutrophil/lymphocyte, and platelet/lymphocyte ratios. Imaging tools, including ultrasound (US), computed tomography (CT), magnetic resonance imaging, and magnetic resonance cholangiopancreatography (MRCP), were selectedly used to document in detail the distribution of stone locations, biliary narrowing, and hepatic lobe atrophy. In some patients with complex bilateral stones, the future residual liver volume and total functional liver volume were calculated using three-dimensional visualisation techniques, and the indocyanine green 15 min retention rate was tested to ensure the safety of the procedure. This test will not be used in patients with a history of indocyanine green allergy and a history of iodine allergy (indocyanine green contains iodine and therefore may cause iodine allergy). If the patients did not reach Child-Pugh class A preoperatively, they received hepatoprotective therapy until their liver function improved to Child-Pugh class A.

2.4.2 Intraoperative strategy and findings

All the surgeries were performed by experienced hepatobiliary surgeons. As patients who had undergone one or more laparotomies tended to have more severe abdominal adhesions, a detailed surgical plan and biliary drainage strategy were formulated based on the location of the stone, sphincter of Oddi function, cirrhosis, and hepatic lobe atrophy, which were confirmed in the preoperative examination and reconfirmed intraoperatively after the surgery. Detailed intraoperative findings, operative approach and duration of surgery were recorded, and choledochoscopy was performed to assess whether the stones were immediately removed. Bile acid was collected intraoperatively for bacterial culture and drug sensitivity testing.

2.4.3 Postoperative examination examination and decision

Postoperative specimens were pathologically diagnosed and described by experienced pathologists from five medical centres. Postoperative complications, including bile leakage, pancreatic fistula, infection, and abdominal bleeding, as well as postoperative blood markers, bile culture, and blood culture results were recorded. Before discharge, abdominal CT and cholangiography or choledochoscopy was used in patients with external T-tube drainage to confirm whether the stone was immediately removed. For patients without instant clearance, choledochoscopy is usually performed through the T-tube sinus tract several times at 6–8 weeks postoperatively until the stone is removed or cannot be removed by any means. For patients with instant clearance, T-tube cholangiography was performed 2 weeks postoperatively. If residual stones were observed, choledochoscopy would be performed, as described above.

2.4.4 Follow-up and data collection

All patients were followed up every 3 months after discharge by the supervising physician in the hepatobiliary surgery clinic or by telephone. Follow-up evaluation included assessment of clinical signs and symptoms, routine blood tests, liver function assessment, and US, CT, or MRCP for residual or recurrent stones. Prognosis was evaluated according to the Terblanche criteria (13) and was considered poor if it was Terblanche classification grade III (serious bile duct-related symptoms requiring treatment) or IV (with anastomotic stricture or bile duct stone formation requiring surgical treatment, resulting in disease-related cancer or death), which was the endpoint of this study.

2.4.5 Missing data handling

Regarding data collection, missing data were dealt with differently in model training and deployment.

During Model Training, for the construction of our machine learning model, we believe in utilizing the most complete and accurate dataset possible. Thus, when an entry has one or more missing feature values, we decided to exclude it from the training process. This approach ensures that our model is trained only on complete cases, minimizing potential biases or inaccuracies that might arise from imputed data.

In our preprocessing steps, the dropna() function was employed to exclude such entries. We're confident that this method is appropriate given our dataset's size and the relative infrequency of missing values. Moreover, we ensured that the removal of these data points did not introduce any bias by examining the distribution of outcomes among the dropped and retained entries.

During Model Deployment, we deemed that in a real-world clinical setting, excluding a patient's data due to a single missing value might not be feasible or desirable. Thus, when our model is used on new patient data, if any feature values are missing, we replace them with the average (mean) value derived from our training dataset. It allows our model to generate predictions even when some data might be temporarily unavailable or missing, and using the mean value from

our training set serves as a neutral placeholder, minimizing the potential impact on the model's prediction.

2.5 Statistical analysis

2.5.1 Data splitting

In our study, the dataset was divided between training and testing sets. The patient data from the First Affiliated Hospital of Anhui Medical University, Second Affiliated Hospital of Anhui Medical University, and First Affiliated Hospital of the University of Science and Technology of China (82.7%) were used for the training set and those from the First Affiliated Hospital of Bengbu Medical College and First Affiliated Hospital of Wannan Medical College (17.3%) for the testing set. This testing set is entirely independent from the training set, thereby enabling out-of-sample evaluation.

Differences in the clinical characteristics of the included patients were compared using independent samples t-test, Mann–Whitney U-test, or χ^2 test, and the statistical significance level was set at 0.05.

2.5.2 Model training

Nine machine learning models were used to build a predictive model for recurrence after RH. These models were selected because they represent different types of machine learning algorithms, including linear models [Logistic Regression (LR)], tree-based models [Decision Tree (DT), Random Rorest (RF), Light Gradient-Boosting Machine (LightGBM), Extreme Gradient Boosting (XGBoost)], integrated methods [XGBoost and Adaptive Boosting (AdaBoost)], support vector machine (SVM), neural network (NNW), and instance-based methods [K-nearest neighbour (KNN)]. By comparing the performance of these different models, the model that performs the best for this particular prediction task can be identified.

All features underwent scaling using the StandardScaler(). This method ensured features were on a similar scale, centering them around zero with a standard deviation of one. To address dataset size limitation and potential class imbalance, ADASYN (Adaptive Synthetic Sampling) was chosen as our oversampling technique. This method was preferred over others like RandomOverSampler due to its ability to generate synthetic samples in regions where the data distribution is sparse. This adaptive approach minimized the risk of overfitting while effectively balancing the class distribution.

To improve the predictive efficacy of the model, five time nodes were set with a spacing of 1 in the range of 1–5 years. For patients who experience recurrence within the first year, we will still incorporate them into the model development in the second year. This was because our time nodes is measured in "k" years, rather than specifically in the "kth" year. This decision was based on the clinical significance of predicting a patient's recurrence in a few years, and providing an intuitive and dynamic recurrence curve, rather than solely predicting recurrence in a specific year. From our original dataset, two key variables were present: "recurrence" (a binary indicator) and "recurrence_time" (quantified in months). Utilizing these, we generated our target variables, "recurrence_in_k_years".

All 84 features were retained in the model to ensure comprehensive data capture and to avoid the premature exclusion of potentially relevant predictors. The reliance on advanced algorithms such as XGB and LightGBM, known for their proficiency in handling high-dimensional data, further justified this decision. The study of feature importance was not conducted for optimization purposes, but rather to provide clinically relevant insights. By understanding which features were deemed most influential by the models, valuable information can be provided to the clinical community about the factors crucial for predicting disease recurrence. Recognizing the distinct consequences of false negatives vs. false positives in medical scenarios, we additionally assigned a cost ratio for False Positives (FP) to False Negatives (FN) of 1:4. This emphasizes the criticality of not overlooking potential risks, as missing a true positive case can have significant ramifications. Beyond the cost matrix, all models were utilized with default configuration.

2.5.3 K-fold cross validation

Concerning our methodology of using only a training and a testing set, without a dedicated validation set, we had specific considerations. Given the limited size of our dataset, we believed that allocating a portion to a validation set could adversely impact the model's performance. Moreover, research indicated that with small datasets, the models often perform best with default hyperparameters, and that hyperparameter tuning might negatively influence performance (14, 15). These factors led us to the decision of not engaging in hyperparameter tuning and adpoting a k-fold Cross Validation with $k = 10$. Our testing set, being independent from the training set, serves to effectively evaluate the model's performance on unseen data.

In cross validation, training set was split randomly into 10 folds. For each iteration, 9 of the 10 folds were used as training set and 1 as validation set. An average AUC was calculated for each model to evaluate if the model was overfitted and used as a benchmark for the model's performance on the testing set. XGBoost and LightGBM consistently outperformed other models in every time node, with AUC of 83.97% and 83.02%, indicating a solid performance of our model and no sign of overfitting. Since the difference between XGBoost and LightGBM is trivial, we decided to conduct final model selection based on their performance on testing set.

2.5.4 Performance evaluation

For each time node, the performance of each model was compared, and the comprehensive evaluation indices were AUC, sensitivity, specificity, accuracy, and F2 score. Considering the ability of the AUC score to evaluate the performance of a model across all thresholds, it was used as a single metric to select the best model at each time node and the model with the highest performance. These metrics were also compared with those of k-fold Cross Validation, to see if the model was overfitted to the training set, in which condition, metrics of validation would be significantly higher than those of testing set.

Descriptive statistics and machine learning analyses were performed using SPSS version 23.0 (IBM Corp, Armonk, NY,

USA) and Python version 3.6.15 (Python Software Foundation, Wilmington, DE, USA).

3 Results

3.1 Patient basic characteristics and clinical outcomes

Based on these criteria, the data of 488 patients who underwent hepatolithiasis surgery in the five medical centres during the 5-year period were evaluated, with 294 patients admitted at the First Affiliated Hospital of Anhui Medical University, 51 patients admitted at the Second Affiliated Hospital of Anhui Medical University, 59 patients admitted at the First Affiliated Hospital of the University of Science and Technology of China, 32 patients admitted at the First Affiliated Hospital of Bengbu Medical College, and 52 patients admitted at the First Affiliated Hospital of Wannan Medical College (Figure 1).

Overall, 488 patients were included in the ML model [mean age, 57.9 ± 12.0 years; >60 years, $n = 235$ (48.2%); female, $n = 331$, 67.8%]. A total of 157 patients (32.1%) underwent more than one surgical treatment, and 89 patients (18.2%) underwent hepatectomy. The characteristics of the training and testing sets were not significantly different (Table 1). A total of 135 patients (27.7%) had a recurrence within 5 years (Table 2). All predictor variables were incorporated into the ML model to predict the risk of recurrence in patients with RH.

In Table 1, we have presented the preoperative clinical characteristics of the patients in a simplified categorical or hierarchical manner for clarity and ease of understanding for the readers. Please note that during the actual model-building process, the original continuous values of these variables were utilized. We believe using the continuous data during model-building aids in capturing subtle nuances and providing a more accurate representation, whereas the categorized data in the table helps in presenting an easier-to-read overview.

3.2 Model performance

The nine models were built and externally validated. The AUC values of the models are presented in Table 3. In terms of predicting RH recurrence at 3 years and more, XGBoost showed optimal performance, with AUCs of about 0.9 or greater, which fully demonstrates its strength. It can efficiently and flexibly handle multivariate data and assemble weak prediction models to build an accurate one (16, 17). In the prediction of recurrence within 1 year and 2 years, LightGBM was more advantageous, with AUCs of 0.981 and 0.924, respectively, whereas the

FIGURE 1
Flow chart of patient enrollment. RH, recurrent hepatolithiasis; ML, machine learning; XGBoost, extreme gradient boosting; LightGBM, light gradient-boosting machine; RF, random forest; SVM, support vector machine; AdaBoost, adaptive boosting; NNW, neural network; DT, decision tree; LR, logistic regression; KNN, K-nearest neighbour; CARES, correlation Analysis and Recurrence Evaluation System.

TABLE 1 Preoperative clinical characteristics of patients with recurrent hepatolithiasis after surgery.

Characteristic	Total (n = 488)	Training set (n = 404)	Testing set (n = 84)	Statistic	P value
Gender, n (%)				$\chi^2 = 0.257$	0.612
Male	157 (32.17)	128 (31.68)	29 (34.52)		
Female	331 (67.83)	276 (68.32)	55 (65.48)		
Age, n (%)				$\chi^2 = 1.520$	0.218
<60	239 (48.98)	203 (50.25)	36 (42.86)		
≥60	249 (51.02)	201 (49.75)	48 (57.14)		
BMI, mean ± SD	21.97 ± 2.85	21.93 ± 2.89	22.17 ± 2.67	$t = -0.700$	0.484
Abdominal pain, n (%)				$\chi^2 = 0.000$	0.987
No	70 (14.34)	58 (14.36)	12 (14.29)		
Yes	418 (85.66)	346 (85.64)	72 (85.71)		
Fever, n (%)				$\chi^2 = 0.284$	0.594
No	303 (62.09)	253 (62.62)	50 (59.52)		
Yes	185 (37.91)	151 (37.38)	34 (40.48)		
Emesis, n (%)				$\chi^2 = 2.067$	0.151
No	383 (78.48)	322 (79.70)	61 (72.62)		
Yes	105 (21.52)	82 (20.30)	23 (27.38)		
Icterus, n (%)				$\chi^2 = 0.722$	0.395
No	384 (78.69)	315 (77.97)	69 (82.14)		
Yes	104 (21.31)	89 (22.03)	15 (17.86)		
Pressing pain, n (%)				$\chi^2 = 0.776$	0.378
No	311 (63.73)	261 (64.60)	50 (59.52)		
Yes	177 (36.27)	143 (35.40)	34 (40.48)		
Smoking, n (%)				$\chi^2 = 0.449$	0.503
No	400 (81.97)	329 (81.44)	71 (84.52)		
Yes	88 (18.03)	75 (18.56)	13 (15.48)		
Drinking, n (%)				$\chi^2 = 0.491$	0.483
No	418 (85.66)	344 (85.15)	74 (88.10)		
Yes	70 (14.34)	60 (14.85)	10 (11.90)		
Number_of_operations, n (%)				Fisher	0.399
1	331 (67.83)	278 (68.81)	53 (63.10)		
2	100 (20.49)	77 (19.06)	23 (27.38)		
3	47 (9.63)	40 (9.90)	7 (8.33)		
≥4	10 (2.05)	9 (2.23)	1 (1.19)		
Previous hepatectomy, n (%)				$\chi^2 = 2.159$	0.142
No	175 (35.86)	139 (34.41)	36 (42.86)		
Yes	313 (64.14)	265 (65.59)	48 (57.14)		
Liver cirrhosis, n (%)				$\chi^2 = 3.785$	0.052
No	428 (87.7)	349 (86.39)	79 (94.05)		
Yes	60 (12.3)	55 (13.61)	5 (5.95)		
Surgical method, n (%)				$\chi^2 = 2.477$	0.116
Open surgery	436 (89.34)	365 (90.35)	71 (84.52)		
Laparoscopic surgery	52 (10.66)	39 (9.65)	13 (15.48)		
Intrahepatic narrow, n (%)				$\chi^2 = 1.130$	0.288
No	367 (75.2)	300 (74.26)	67 (79.76)		
Yes	121 (24.8)	104 (25.74)	17 (20.24)		
Hepatic lobe atrophy, n (%)				$\chi^2 = 0.231$	0.630
No	215 (44.06)	176 (43.56)	39 (46.43)		
Yes	273 (55.94)	228 (56.44)	45 (53.57)		
AGR, n (%)				$\chi^2 = 0.671$	0.413
>1.5	158 (32.38)	134 (33.17)	24 (28.57)		
≤1.5	330 (67.62)	270 (66.83)	60 (71.43)		
NLR, n (%)				$\chi^2 = 3.156$	0.076
<2.462	292 (59.84)	249 (61.63)	43 (51.19)		
≥2.462	196 (40.16)	155 (38.37)	41 (48.81)		
PLR, n (%)				$\chi^2 = 0.168$	0.682
<173.74	393 (80.53)	324 (80.20)	69 (82.14)		
≥173.74	95 (19.47)	80 (19.80)	15 (17.86)		
TBIL, n (%)				$\chi^2 = 0.071$	0.790
<34.2	400 (81.97)	332 (82.18)	68 (80.95)		

(Continued)

TABLE 1 Continued

Characteristic	Total (*n* = 488)	Training set (*n* = 404)	Testing set (*n* = 84)	Statistic	*P* value
≥34.2	88 (18.03)	72 (17.82)	16 (19.05)		
ALT, *n* (%)				$\chi^2 = 0.728$	0.393
<50	299 (61.27)	251 (62.13)	48 (57.14)		
≥50	189 (38.73)	153 (37.87)	36 (42.86)		
AST, *n* (%)				$\chi^2 = 0.008$	0.929
<40	300 (61.48)	248 (61.39)	52 (61.90)		
≥40	188 (38.52)	156 (38.61)	32 (38.10)		
ALP, *n* (%)				$\chi^2 = 0.899$	0.343
<200	309 (63.32)	252 (62.38)	57 (67.86)		
≥200	179 (36.68)	152 (37.62)	27 (32.14)		
GGT, *n* (%)				$\chi^2 = 0.192$	0.661
<150	243 (49.8)	203 (50.25)	40 (47.62)		
≥150	245 (50.2)	201 (49.75)	44 (52.38)		
CA19-9, *n* (%)				$\chi^2 = 2.288$	0.130
<34	338 (69.26)	274 (67.82)	64 (76.19)		
≥34	150 (30.74)	130 (32.18)	20 (23.81)		

This table summarizes patient data on key clinically significant variables only. BMI, body mass index; AGR, albumin-to-globulin ratio; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; TBIL, total bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; GGT, γ-glutamyl transpeptidase; CA19-9, carbohydrate antigen19-9;.

TABLE 2 The number of recurrent patients in k years.

In k years	1	2	3	4	5
Number	44	108	126	132	135

performance of the DT and KNN models was unsatisfactory, probably because the sample size was not sufficiently large (Figure 2) (18). It was worth noticing that model showed better performance on testing set than validation, indicating that it was not overfitted to the training set.

For the clinical results at each time point, Shapley Additive Explanations (SHAP) were generated to construct a comprehensive explainable framework showing the importance and direction of each predictor variable, increasing the interpretability of the model. The position of each predictor variable on the *y*-axis was ranked in order of relative importance, with the most important predictor variable at the top. For each predictor variable, the position of each point on the *x*-axis (red indicates higher values or the presence of binary factors) represents the contribution of the individual participant to the

overall SHAP value, with highly positive contributions on the far right (Figure 3).

3.3 Predictive analysis and clinical application

Instant and final clearance were of considerable importance in the prediction of almost every time point, whereas the number of previous surgeries and the neutrophil/lymphocyte ratio were also of great importance, which is in line with our previous findings (11). Moreover, advanced ML models can capture higher-order non-linear interactions among predictors; therefore, we also found many previously unappreciated or undetected factors that have great impact on recurrence, such as the function of the sphincter of Oddi (SO), carbohydrate antigen 19-9 (CA19-9), symptom score, and platelet count.

The system named CARES employs five specialized models, each optimized for predicting the risk of disease recurrence for years 1–5 post-surgery. Specifically, CARES has 5 system components and goes through the following steps.

TABLE 3 Area under the receiver operating characteristic curve (AUC) of each model at different time nodes.

Model	AUC within 1 year	AUC within 2 years	AUC within 3 years	AUC within 4 years	AUC within 5 years
XGBoost	0.941	0.906	0.922	0.917	0.887
LightGBM	0.981	0.924	0.889	0.907	0.885
RF	0.903	0.825	0.852	0.849	0.774
SVM	0.900	0.856	0.836	0.843	0.832
AdaBoost	0.659	0.779	0.732	0.661	0.781
NNW	0.747	0.852	0.823	0.845	0.813
DT	0.469	0.650	0.674	0.636	0.542
LR	0.819	0.839	0.810	0.833	0.795
KNN	0.600	0.592	0.585	0.576	0.568

This table summarizes area under the receiver operating characteristic curve (AUC) of each model at different time nodes only. Additional data on optimal parameters and performance of each model is summarized in Supplementary Appendix. XGBoost, extreme gradient boosting; LightGBM, light gradient-boosting machine; RF, random forest; SVM, support vector machine; AdaBoost, adaptive boosting; NNW, neural network; DT, decision tree; LR, logistic regression; KNN, K-nearest neighbour.
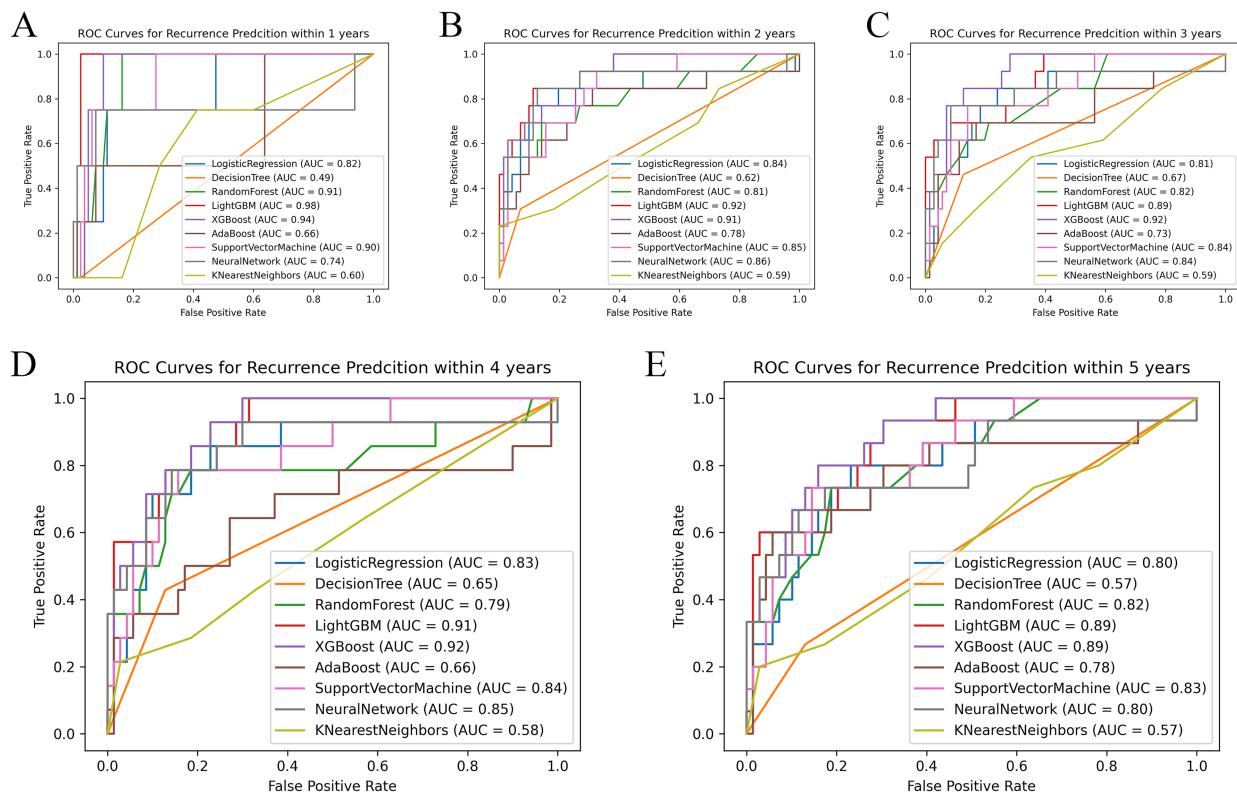
FIGURE 2
Comparison of ROC curves of each model at different time nodes. Panels A−E respectively show the ROC curves and AUC of each model at the time points set to 1, 2, 3, 4, and 5 years. AUC, area under the receiver operating characteristic curve.

Firstly, for each k (ranging from 1 to 5), a dedicated model is trained using the entire dataset to predict the probability of a patient experiencing disease recurrence k years after surgery. This results in 5 distinct models, each optimized for its specific prediction year. Secondly, for a new patient, measurements and relevant clinical information serve as the input. In instances where certain data points are missing, these are substituted with the sample average to ensure a comprehensive data input. Thirdly, each of the 5 models processes the input data, providing individual probability estimates of the patient's risk of disease recurrence for years 1 through 5. Fourthly, to ensure that the risk curve exhibits clinical coherence (i.e., the risk doesn't drop in subsequent years, which would be counterintuitive), an isotonic regression is applied to the predicted probabilities. Lastly, the output of the CARES system is a graphical representation or "risk curve". This curve offers a clear visualization of a patient's estimated risk of recurrence across the 5-year period post-surgery.

This system was encapsulated and deployed online. When the user inputs the patient's predictors, it outputs a curve of recurrence risk over time; when the patient's recurrence risk is higher at a certain time point or spikes at a certain period of time, we notify the user of the output on the output graph to draw attention to the patient's recurrence risk (Figure 4). This incorporation of individual and aggregated predictive models aids in offering a comprehensive and nuanced risk profile. Compared with previous scoring systems, our calculator is easier to use and the output is more intuitive, with

greater utility and a higher predictive value. The CARES is available for free online (19) and can also be accessed by scanning the QR code.

In terms of evaluation, the model's efficacy can be gauged by comparing its predictions against actual recurrence events in a real-world clinical setting. After deployment in real practice, continual validation and recalibration can further refine the model, ensuring its sustained relevance and accuracy.

4 Discussion

4.1 Principal findings

In this study, ML methods and multicentre clinical data were combined to build CARES, an accurate, efficient, and user-friendly prediction model that integrates clinical characteristics to predict the dynamic recurrence risk of RH after surgery, and then analysed the risk factors that may be associated with recurrence using the SHAP method. Based on SHAP at various time points, immediate stone clearance, final stone clearance, number of previous surgeries, and preoperative CA19-9 index were the most significant predictors of recurrence after reoperation in RH patients. We employed state-of-the-art algorithms, such as XGB and LightGBM. It's noteworthy that, to our knowledge, these algorithms have not been previously utilized in modeling recurrence of this specific disease. CARES is the first model that uses ML to assess the prognosis of patients
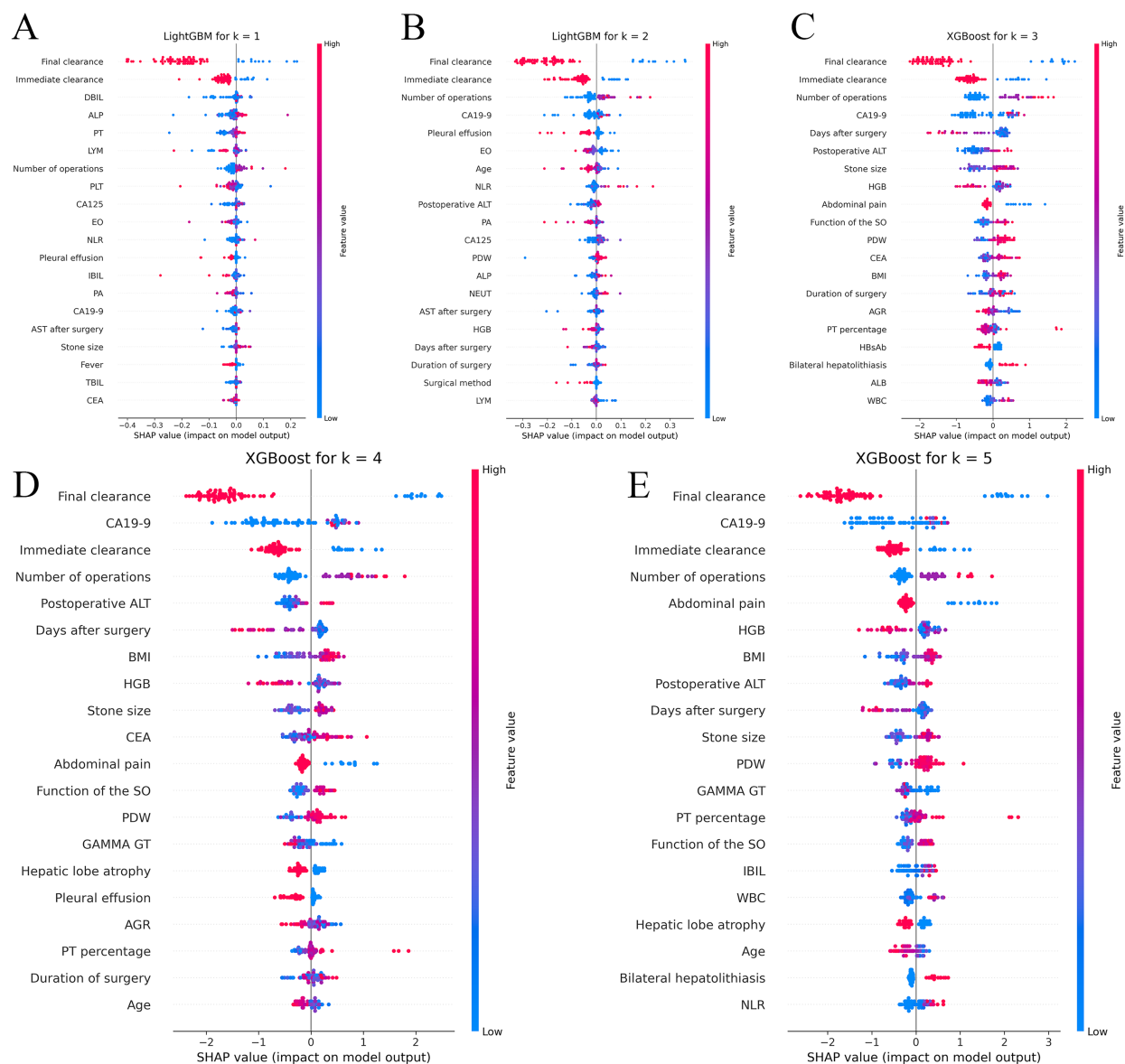
FIGURE 3

Shapley additive explanations (SHAP) analyses of the best-performing machine learning models for predicting recurrence of hepatolithiasis. Panels **A,B** respectively show the Shapley additive explanations (SHAP) for the LightGBM model, which performed the best at the 1-year and 2-year time points, while panels **C–E** respectively show the SHAP for the XGBoost model at the 3-year, 4-year, and 5-year time points. XGBoost, extreme gradient boosting; LightGBM, light gradient-boosting machine; DBIL, direct bilirubin; ALP, alkaline phosphatase; PT, prothrombin time; LYM, lymphocyte; PLT, platelet count; CA125, carbohydrate antigen 125; EO, eosinophil; NLR, neutrophil-to-lymphocyte ratio; IBIL, indirect bilirubin; PA, prealbumin; CA19-9, carbohydrate antigen19-9; AST, aspartate aminotransferase; TBIL, total bilirubin; CEA, carcinoembryonic antigen; PDW, platelet distribution width; NEUT, neutrophil count; AGR, albumin-to-globulin ratio; HBsAb, hepatitis B surface antibody; WBC, white blood cell; BMI, body mass index; HGB, hemoglobin; GGT, γ-glutamyl transpeptidase.

with RH after biliary surgery. We incorporated the latest dataset available, which, to the best of our knowledge, is unparalleled in its scale and comprehensiveness for this subject.

4.2 Interdisciplinary integration

Hepatolithiasis is a relatively common benign disease in East Asia; however, the management of patients with hepatolithiasis has been challenging owing to the high rates of treatment failure,

recurrence, and complications (20–22). Patients with RH are also more difficult to re-treat because they have already undergone one or multiple surgeries, and repeat surgery places a greater psychological and financial burden on patients. Therefore, a model that accurately predicts the individual dynamic recurrence risk of patients with RH after surgical treatment could provide great value in guiding the assessment of postoperative efficacy as well as the development of a follow-up strategy (23).

The application of AI in healthcare is growing rapidly with potential applications in various subspecialties and subfields (24–
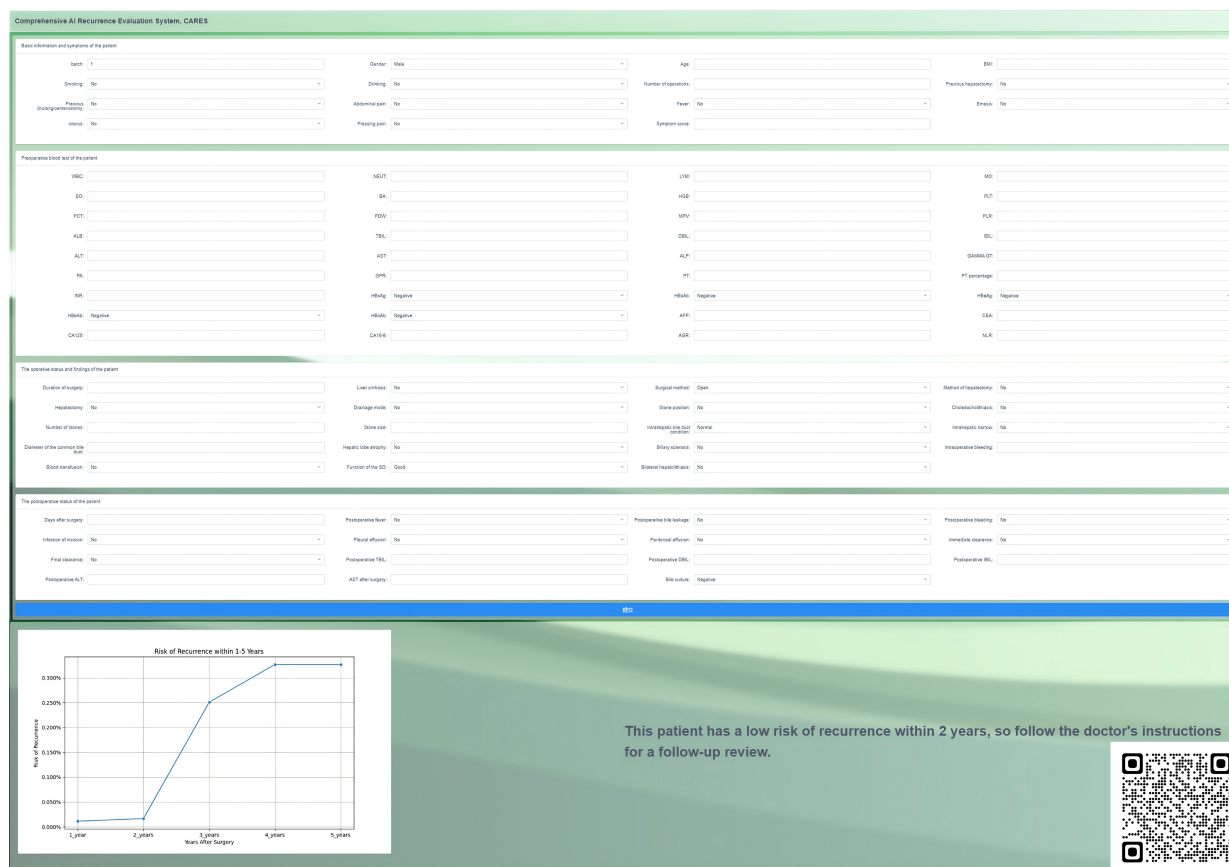
FIGURE 4
Page presentation of the online correlation analysis and recurrence evaluation system (CARES), which is available for free at http://www.ahmucares. tech:5000/ or by scanning the QR code.

26). As an important branch of AI, ML can be trained by inputting large amounts of labelled data (27) and analysing these data to identify relevant patterns that can then be used to predict future events or states (28). It has the ability to learn automatically from data and algorithms and uses past experience to improve performance (29). Unlike traditional regression-based methods, ML algorithms capture higher-order non-linear interactions between predictors (30) and thus focus on detecting hard-to-recognise patterns in complex data. CARES allows the comparison of multiple learning algorithms to identify the algorithm with the best performance.

When developing CARES, a different oversampling method was used, ADASYN, to prevent the imbalance in the amount of negative vs. positive data from distorting the model's performance. Unlike random oversampling, which simply replicates existing examples, ADASYN generates new synthetic examples in a small number of classes that are slightly different from existing examples, with a particular focus on samples that are more difficult to learn. These synthetic examples make the model more robust and reduce the risk of overfitting because they introduce more variability and help the model to better generalise the training data to new data.

Our study also demonstrated that a prediction model based on ML techniques was superior to the traditional regression analysis method in terms of predictive performance. Previous studies had

few predictive models for postoperative recurrence in patients with RH. We used traditional LR to build a recurrence prediction model for patients with RH after biliary surgery, which had an AUC of 0.754 and was not fully satisfactory (11). In contrast, with the help of ML techniques, the AUC of LightGBM reached 0.981 and 0.924 for patients with recurrence within 1 year and 2 years after surgery, respectively, whereas XGBoost performed exceptionally well for patients with recurrence at 3 years and beyond, with AUCs of 0.922, 0.917, and 0.887 at 3, 4, and 5 years, respectively.

As a widely used model in biological and medical analyses, XGBoost is a boosting algorithm with many advantages. First, several variables may have affected disease recurrence. By building an ensemble of decision trees, XGBoost can capture complex relationships between features and outcomes, which may be particularly important in medical scenarios where multiple factors interact to influence outcomes. Second, our dataset contains a large number of predictor variables, including binary, numerical, and categorical data. XGBoost can handle all these types of data, allowing us to incorporate all potentially relevant information into the prediction (31). Finally, our dataset was considered unbalanced, with a limited number of samples and fewer positive data. XGBoost addresses this issue. It also provides resilience against overfitting and supports parallel processing to maximise the use of resources

(32). Therefore, XGBoost tends to have excellent performance when the number of predictor variables is large and the dataset is not balanced. The present study also indicated that the prediction model based on XGBoost had the best performance.

As ML becomes more computationally powerful and the complexity of models increases, understanding the underlying logic and decision factors of the models becomes increasingly difficult. Therefore, enhancing the interpretability of black boxes so that people can understand the reasons for their predictions can considerably improve the applicability and credibility of models (33). Therefore, we combined the predictions of CARES with SHAP to construct a comprehensive explanatory framework for presenting the contribution of each predictor variable to the results and to increase the transparency of the model (34). SHAP has many advantages. It can calculate the contribution of various factors, determine the positivity or negativity of each contribution, quantify each factor's contribution to the stone recurrence/non-occurrence probability, and predict recurrence without decreasing the predictive model's accuracy (33, 35). These advantages are important for the prediction of potential recurrence risk, clinical focus of influencing factors, and interpretation of CARES prediction results.

4.3 Clinical findings and contributions

According to the results of the SHAP, instant and final clearance of stones were the most important predicting factors. Patients who fail to achieve instant clearance and final clearance appear to be at a much higher risk of recurrence, showing that perfect preoperative examination and fine intraoperative operation are quite beneficial in improving the patient's prognosis. Therefore, the surgical method should be carefully selected to remove all stones intraoperatively, based on preoperative examination. For patients in whom intraoperative stone extraction is difficult, such as those with stones in both the hepatic and biliary ducts, severe lateral hepatectomy combined with choledochoscopic lithotripsy can be attempted to obtain a high stone removal rate (36, 37). Stones that are difficult to remove intraoperatively should be removed postoperatively using trans-T-tube sinusoidal choledochoscopy.

The number of previous surgeries was also a major concern. According to the SHAP, a greater number of previous surgeries significantly increases a patient's risk of recurrence. According to previous studies, up to 95% of prior abdominal surgeries result in intra-abdominal adhesions (38), which may be related to intraoperative vascular and intestinal injuries (39). A complex abdominal environment can greatly increase the difficulty of surgery, making accurate resection of lesions and removal of stones difficult. Therefore, care should be taken when choosing a surgical procedure for patients who have undergone multiple laparotomies. Open approach may be a better option than laparoscopic approach because in patients with severe abdominal adhesions, improper placement of the trocar may prevent effective laparoscopic surgery and may damage the viscera or vascular around the adhesions. Loosening the abdominal adhesions to accurately identify the anatomical landmarks can be a challenge during surgery.

In our study, CA19-9 played an important role in recurrence at certain time points, higher CA19-9 levels in patients on preoperative examination suggested a higher risk of recurrence. Previous studies on the relationship between CA19-9 and hepatolithiasis have often been limited to whether it is associated with malignancy in biliary diseases; little research has been conducted on its relationship with recurrence. According to Ker et al. (40), the concentration of CA19-9 is not only affected by tumours but is also increased by severe infections in patients with hepatolithiasis. Cases of stone-induced acute bile duct inflammation leading to elevated CA19-9 levels were also reported by Sheen-Chen et al. (41). We hypothesised that patients with elevated CA19-9 levels may have more severe tract infections, which may disrupt the biliary environment and increase the risk of recurrence.

In addition to the aforementioned key risk factors, the function of SO also affected recurrence in our prediction model. The primary function of the SO is to regulate bile influx into the duodenum and to prevent duodenal reflux (42). Duodenal reflux of food debris can lead to Escherichia coli infections and a decrease in biliary pH. E. coli can generate β-glucuronidase, which hydrolyses water-soluble direct bilirubin into water-insoluble indirect bilirubin, thereby facilitating stone formation in the biliary tract (43). Consequently, patients with poorer SO function are more prone to recurrence. Therefore, maintaining the functional integrity of SO helps to reduce the recurrence rate in patients with RH. In patients with normal SO function, the best method of biliary drainage is T-tube drainage, which is relatively simple, has a high stone-clearance rate, and preserves the structural integrity and continuity of the extrahepatic bile ducts because it preserves SO function. T-tube drainage significantly reduces the incidence of post-operative reflux cholangitis in patients with normal SO function. However, in patients with complete loss of function or stenosis of the SO, Roux-en-Y hepatico-jejunostomy is one of the best methods available for biliary drainage. Roux-en-Y hepatico-jejunostomy has the advantage that it reduces reflux of duodenal fluid, but this procedure abandons the SO (44). Therefore, to reduce the recurrence rate in patients with RH, the surgeon should carefully choose the method for different states of SO function and preserve SO function as much as possible to prevent the occurrence of reflux cholangitis.

Naturally, other factors seem to influence the recurrence of hepatolithiasis, but the direct link between these factors, such as postoperative fever, and the recurrence of hepatolithiasis is difficult to understand. However, ML has the advantage of observing complex, multidimensional, and non-linear relationships between different predictor variables in biological systems. Perhaps in the future, we can aim to understand how these factors cause physiological and pathological "butterfly effects" in the human body and isolate them to demonstrate a complete "chain of evidence."

To improve the application value of the model, we encapsulated the CARES as a recurrence risk curve calculator and deployed it online. By inputting patient information, the calculator outputs a dynamic recurrence risk curve that increases with time after the operation, and the user can approximate the patient's possible risk of recurrence based on the output. An open interface is reserved in CARES for interfacing with the hospital information system.

CARES not only has a better performance but can also visually output the change in recurrence risk of patients in each period from 1 to 5 years after surgery, suggesting the period when doctors and patients need to be extra cautious, as well as the indicators and guidelines that they need to focus on.

4.4 Limitations

This study has some limitations. First, the retrospective nature of the methodology may lead to a selection bias, and prospective studies are needed to validate the accuracy of the results. Second, during model training, due to the imbalanced nature of our dataset, we adapted ADASYN as oversampler. We acknowledged that while ADASYN helped address class imbalance, it may not fully capture the complexities of real-world distributions in clinical settings. Third, the explainable internal working logic of the model remains one of the biggest barriers to implementing cutting-edge ML techniques in biomedical research. We must better understand the evolving and complex relationships between physicians and smart tools in clinical settings to provide better treatment strategies for patients.

5 Conclusions

Multiple ML algorithms were used to construct CARES, which integrates various clinical data to predict the dynamic recurrence risk of RH patients after surgery. The predictive power of our model was externally validated based on a multicentre database. We believe that CARES can provide critical prognostic predictions for patients after RH surgery and may facilitate more efficient clinical decision-making by surgeons and patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the committee on Medical Ethics, the first affiliated hospital of Anhui Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

ZL: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. YZ: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Validation, Writing – original draft. ZC: Investigation, Methodology, Validation, Visualization, Writing – review & editing. JC: Data curation, Investigation, Project administration, Resources, Supervision, Validation, Writing – original draft. HH: Data curation, Formal Analysis, Project administration, Writing – original draft. CW: Data curation, Formal Analysis, Project administration, Writing – original draft. ZL: Data curation, Formal Analysis, Project administration, Writing – original draft. XW: Data curation, Formal Analysis, Project administration, Writing – original draft. XG: Formal Analysis, Methodology, Resources, Supervision, Writing – review & editing. FL: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2024.1510674/full#supplementary-material

References

1. Kim HJ, Kim JS, Joo MK, Lee BJ, Kim JH, Yeon JE, et al. Hepatolithiasis and intrahepatic cholangiocarcinoma: a review. *World J Gastroenterol*. (2015) 21 (48):13418–31. doi: 10.3748/wjg.v21.i48.13418

2. Lei J, Huang J, Yang X, Zhang Y, Yao K. Minimally invasive surgery versus open hepatectomy for hepatolithiasis: a systematic review and meta analysis. *Int J Surg*. (2018) 51:191–8. doi: 10.1016/j.ijsu.2017.12.038

3. Shoda J, Tanaka N, Osuga T. Hepatolithiasis–epidemiology and pathogenesis update. *Front Biosci*. (2003) 8:e398–409. doi: 10.2741/1091

4. Tazuma S. Gallstone disease: epidemiology, pathogenesis, and classification of biliary stones (common bile duct and intrahepatic). *Best Pract Res Clin Gastroenterol*. (2006) 20(6):1075–83. doi: 10.1016/j.bpg.2006.05.009

5. Tan J, Tan Y, Chen F, Zhu Y, Leng J, Dong J. Endoscopic or laparoscopic approach for hepatolithiasis in the era of endoscopy in China. *Surg Endosc*. (2015) 29(1):154–62. doi: 10.1007/s00464-014-3669-5

6. Lorio E, Patel P, Rosenkranz L, Patel S, Sayana H. Management of hepatolithiasis: review of the literature. *Curr Gastroenterol Rep*. (2020) 22(6):30. doi: 10.1007/s11894-020-00765-3

7. Tazuma S, Unno M, Igarashi Y, Inui K, Uchiyama K, Kai M, et al. Evidence-based clinical practice guidelines for cholelithiasis 2016. *J Gastroenterol*. (2017) 52 (3):276–300. doi: 10.1007/s00535-016-1289-7

8. Cheon YK, Cho YD, Moon JH, Lee JS, Shim CS. Evaluation of long-term results and recurrent factors after operative and nonoperative treatment for hepatolithiasis. *Surgery*. (2009) 146(5):843–53. doi: 10.1016/j.surg.2009.04.009

9. Uchiyama K, Kawai M, Ueno M, Ozawa S, Tani M, Yamaue H. Reducing residual and recurrent stones by hepatectomy for hepatolithiasis. *J Gastrointest Surg*. (2007) 11 (5):626–30. doi: 10.1007/s11605-006-0024-8

10. Pu Q, Zhang C, Huang Z, Zeng Y. Reoperation for recurrent hepatolithiasis: laparotomy versus laparoscopy. *Surg Endosc*. (2017) 31(8):3098–105. doi: 10.1007/s00464-017-5631-9

11. Pu T, Chen JM, Li ZH, Jiang D, Guo Q, Li AQ, et al. Clinical online nomogram for predicting prognosis in recurrent hepatolithiasis after biliary surgery: a multicenter, retrospective study. *World J Gastroenterol*. (2022) 28(7):715–31. doi: 10.3748/wjg.v28.i7.715

12. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol*. (2019) 20(5):e262–e73. doi: 10.1016/S1470-2045(19)30149-4

13. Terblanche J, Worthley CS, Spence RA, Krige JE. High or low hepaticojejunostomy for bile duct strictures? *Surgery*. (1990) 108(5):828–34.

14. Probst P, Boulesteix A-L, Bischl B. Tunability: importance of hyperparameters of machine learning algorithms. *J Mach Learn Res*. (2019) 20(1):1934–65. doi: 10.48550/arXiv.1802.09596

15. Galindo JA, Dominguez AJ, White J, Benavides D. Large language models to generate meaningful feature model instances. *Proceedings of the 27th ACM International Systems and Software Product Line Conference - Volume A; Tokyo, Japan: Association for Computing Machinery* (2023). p. 15–26

16. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. *J Transl Med*. (2020) 18(1):462. doi: 10.1186/s12967-020-02620-5

17. Yuan KC, Tsai LW, Lee KH, Cheng YW, Hsu SC, Lo YS, et al. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *Int J Med Inform*. (2020) 141:104176. doi: 10.1016/j.ijmedinf.2020.104176

18. Li X, Yang L, Yuan Z, Lou J, Fan Y, Shi A, et al. Multi-institutional development and external validation of machine learning-based models to predict relapse risk of pancreatic ductal adenocarcinoma after radical resection. *J Transl Med*. (2021) 19 (1):281. doi: 10.1186/s12967-021-02955-7

19. Correlation Analysis and Recurrence Evaluation System, CARES. Available online at: http://www.ahmucares.tech:5000/

20. Tsui WM, Chan YK, Wong CT, Lo YF, Yeung YW, Lee YW. Hepatolithiasis and the syndrome of recurrent pyogenic cholangitis: clinical, radiologic, and pathologic features. *Semin Liver Dis*. (2011) 31(1):33–48. doi: 10.1055/s-0031-1272833

21. Park JS, Jeong S, Lee DH, Bang BW, Lee JI, Lee JW, et al. Risk factors for long-term outcomes after initial treatment in hepatolithiasis. *J Korean Med Sci*. (2013) 28 (11):1627–31. doi: 10.3346/jkms.2013.28.11.1627

22. de Andres Olabarria U, Garcia Bruna L, Maniega Alba R, Ibanez Aguirre FJ. Hepatolitiasis masiva secundaria a síndrome del sumidero. *Cir Esp (Engl Ed)*. (2019) 97(3):176. doi: 10.1016/j.ciresp.2018.08.006

23. Truong M, Slezak JA, Lin CP, Iremashvili V, Sado M, Razmaria AA, et al. Development and multi-institutional validation of an upgrading risk tool for

gleason 6 prostate cancer. *Cancer*. (2013) 119(22):3992–4002. doi: 10.1002/cncr.28303

24. Li Z, Wang L, Wu X, Jiang J, Qiang W, Xie H, et al. Artificial intelligence in ophthalmology: the path to the real-world clinic. *Cell Rep Med*. (2023) 4(7):101095. doi: 10.1016/j.xcrm.2023.101095

25. Stafie CS, Sufaru IG, Ghiciuc CM, Stafie II, Sufaru EC, Solomon SM, et al. Exploring the intersection of artificial intelligence and clinical healthcare: a multidisciplinary review. *Diagnostics (Basel)*. (2023) 13(12):1995. doi: 10.3390/diagnostics13121995

26. Hamid N, Portnoy JM, Pandya A. Computer-assisted clinical diagnosis and treatment. *Curr Allergy Asthma Rep*. (2023) 23(9):509–17. doi: 10.1007/s11882-023-01097-8

27. Manickam P, Mariappan SA, Murugesan SM, Hansda S, Kaushik A, Shinde R, et al. Artificial intelligence (AI) and internet of medical things (IoMT) assisted biomedical systems for intelligent healthcare. *Biosensors (Basel)*. (2022) 12(8):562. doi: 10.3390/bios12080562

28. Kulkarni S, Seneviratne N, Baig MS, Khan AHA. Artificial intelligence in medicine: where are we now? *Acad Radiol*. (2020) 27(1):62–70. doi: 10.1016/j.acra.2019.10.001

29. Kang J, Hanif M, Mirza E, Khan MA, Malik M. Machine learning in primary care: potential to improve public health. *J Med Eng Technol*. (2021) 45(1):75–80. doi: 10.1080/03091902.2020.1853839

30. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. *Cell*. (2015) 163 (5):1079–94. doi: 10.1016/j.cell.2015.11.001

31. Ahirwal J, Nath A, Brahma B, Deb S, Sahoo UK, Nath AJ. Patterns and driving factors of biomass carbon and soil organic carbon stock in the Indian Himalayan region. *Sci Total Environ*. (2021) 770:145292. doi: 10.1016/j.scitotenv.2021.145292

32. Bertini A, Salas R, Chabert S, Sobrevia L, Pardo F. Using machine learning to predict complications in pregnancy: a systematic review. *Front Bioeng Biotechnol*. (2022) 9:780389. doi: 10.3389/fbioe.2021.780389

33. Zhang J, Ma X, Zhang J, Sun D, Zhou X, Mi C, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J Environ Manage*. (2023) 332:117357. doi: 10.1016/j.jenvman.2023.117357

34. Goodwin NL, Nilsson SRO, Choong JJ, Golden SA. Toward the explainability, transparency, and universality of machine learning for behavioral classification in neuroscience. *Curr Opin Neurobiol*. (2022) 73:102544. doi: 10.1016/j.conb.2022.102544

35. Lapuschkin S, Waldchen S, Binder A, Montavon G, Samek W, Muller KR. Unmasking clever hans predictors and assessing what machines really learn. *Nat Commun*. (2019) 10(1):1096. doi: 10.1038/s41467-019-08987-4

36. Li SQ, Liang LJ, Peng BG, Hua YP, Lv MD, Fu SJ, et al. Outcomes of liver resection for intrahepatic stones. *Ann Surg*. (2012) 255(5):946–53. doi: 10.1097/SLA.0b013e31824dedc2

37. Connell M, Sun WYL, Mocanu V, Dang JT, Kung JY, Switzer NJ, et al. Management of choledocholithiasis after roux-en-Y gastric bypass: a systematic review and pooled proportion meta-analysis. *Surg Endosc*. (2022) 36(9):6868–77. doi: 10.1007/s00464-022-09018-y

38. Wei X, Lu J, Siddiqui KM, Li F, Zhuang Q, Yang W, et al. Does previous abdominal surgery adversely affect perioperative and oncologic outcomes of laparoscopic radical cystectomy? *World J Surg Oncol*. (2018) 16(1):10. doi: 10.1186/s12957-018-1317-6

39. Parsons JK, Jarrett TJ, Chow GK, Kavoussi LR. The effect of previous abdominal surgery on urological laparoscopy. *J Urol*. (2002) 168(6):2387–90. doi: 10.1016/S0022-5347(05)64151-1

40. Ker CG, Wu CC, Chen JS, Hou MF, Lee KT, Sheen PC. A study of CEA, CA 19-9 and CA 125 in biliary tract diseases. *Gaoxiong Yi Xue Ke Xue Za Zhi*. (1989) 5(2):107–13.

41. Sheen-Chen SM, Sun CK, Liu YW, Eng HL, Ko SF, Kuo CH. Extremely elevated CA19-9 in acute cholangitis. *Dig Dis Sci*. (2007) 52(11):3140–2. doi: 10.1007/s10620-006-9164-7

42. Lian YG, Zhang WT, Xu Z, Ling XF, Wang LX, Hou CS, et al. Oddi sphincter preserved cholangioplasty with hepatico-subcutaneous stoma for hepatolithiasis. *World J Gastroenterol*. (2015) 21(45):12865–72. doi: 10.3748/wjg.v21.i45.12865

43. Liang TB, Liu Y, Bai XL, Yu J, Chen W. Sphincter of oddi laxity: an important factor in hepatolithiasis. *World J Gastroenterol*. (2010) 16(8):1014–8. doi: 10.3748/wjg.v16.i8.1014

44. Chen JM, Yan XY, Zhu T, Chen ZX, Zhao YJ, Xie K, et al. T-tube drainage versus choledochojejunostomy in hepatolithiasis patients with sphincter of oddi laxity: study protocol for a randomized controlled trial. *Trials*. (2020) 21(1):586. doi: 10.1186/s13063-020-04483-z

The benefits of contrast-enhanced ultrasound in the differential diagnosis of suspicious breast lesions

Runa Liang[1,2†], Jun Lian[2†], Jinhui Zhang[1], Jiayu Jing[1], Jinxia Bian[1], Jinzhi Xu[1], Xin He[1], Shanshan Yu[1], Qi Zhou[1] and Jue Jiang[1]*

[1]Department of Ultrasound, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China,
[2]Department of Ultrasound, Ankang Central Hospital, Ankang, China

Background: Contrast-enhanced ultrasound (CEUS) shows potential for the differential diagnosis of breast lesions in general, but its effectiveness remains unclear for the differential diagnosis of lesions highly suspicious for breast cancers.

Objective: This study aimed to evaluate the diagnostic value of CEUS in differentiating pathological subtypes of suspicious breast lesions defined as category 4 of US-BI-RADS.

Methods: The dataset of 150 breast lesions was prospectively collected from 150 patients who underwent routine ultrasound and CEUS examination and were highly suspected of having breast cancers. All lesions were pathologically confirmed by US-guided needle biopsy and surgery. The qualitative features and the quantitative parameters of CEUS of these breast lesions were analyzed. The CEUS and biopsy examinations were performed after informed consent.

Results: In the qualitative features, crab clam-like enhancement, the presence of more than two enhanced vessels within lesions, and surrounding enriched vessels inserting into lesions were able to differentiate atypical fibroadenomas (FIB) and mass-like non-puerperal mastitis (NPM) from invasive ductal carcinomas (IDC) and ductal carcinomas *in situ* (DCIS) ($p < 0.05$). The enlarged scope, irregular shape, and perfusion deficiency were valuable to the differential diagnosis of FIB from the others ($p < 0.05$). In the four quantitative parameters of CEUS, only the peak intensity (IMAX) contributed to the differential diagnosis between malignant and benign tumors ($p < 0.05$, ROCAUC: 0.61, sensitivity: 60.4% and specificity: 65.9%, accuracy: 62.1%). However, IMAX did not show any difference in the paired comparison of IDC, DCIS, FIB, and NPM ($p > 0.05$). The logistic regression analysis results showed that heterogeneous perfusion, crab clam-like enhancement, and partial_ IMAX were independent risk factors for benign and malignant breast lesions *($p < 0.05$)*. The area under a receiver operating characteristic of the integrated model was 0.89. In the diagnosis of benign and malignant pathological subtypes of breast lesions, independent risk factors and integrated models had no statistical significance in the diagnosis of IDC and DCISs, FIB, and NPM ($p > 0.05$).

Conclusion: Some qualitative risk features of CEUS can distinguish malignant breast lesions from NPM and atypical FIB with a high score of US-BI-RADS, aiding physicians to reduce the misdiagnosis of suspicious breast lesions in clinical practice.

Introduction

According to the statistical results of 36 kinds of cancers worldwide in 2022, the incidence rate of female breast cancer was ranked second, and the corresponding mortality rate was ranked fourth (1). Ultrasound (US) examination is an important and often-used tool to find breast lesions and distinguish the malignancies and benignities (2, 3). Currently, ultrasonographers predict the probabilities of malignant breast lesions according to the American Colleague Radiology US Breast Imaging—Reporting and Data System (ACR US-BI-RADS) (4). However, the diagnostic specificity of the high-risk categories remains widely controversial, especially for lesions scored as category 4 of US-BI-RADS, whose risk probability ranges from 2 to 95% (4), because of the highly overlapped risk features between malignant and benign breast lesions (5, 6). The technique of contrast-enhanced ultrasound (CEUS) can visualize the distribution and pattern of the microvascular environment within or surrounding organs or lesions (7–9), which has proven useful in differentiating malignant from benign breast lesions (10–13). However, the previous studies mainly evaluated the value of CEUS in differentiating benignity/malignancy of breast lesions overall (14–18), rarely focusing on the histopathological subtypes of breast lesions, especially for the atypical benignities that are easily mistaken for breast cancers.

Thus, in this study, we focused on the suspicious breast lesions defined as category 4 of US-BI-RADS and evaluated the diagnostic value of both qualitative features and quantitative parameters of CEUS in differentiating pathological subtypes of those lesions. To improve the accuracy of early diagnosis for such lesions, reduce unnecessary biopsy procedures, and obtain practical and highly accurate diagnostic guidelines for breast ultrasound contrast imaging, providing reliable and practical imaging diagnostic support for precise clinical diagnosis and treatment.

Materials and methods

Patients

In this prospective study, 228 single breast lesions from 228 patients were identified via the routine ultrasonic examination and classified as category 4 according to ACR US-BI-RADS (Figure 1). All patients were advised to undergo a CEUS examination before US-guided coarse needle biopsy (CNB) and surgery. However, 43 patients did not perform the CEUS examination and underwent surgery directly. A total of 33 patients failed to follow up and did not obtain their pathological results. Finally, 152 breast lesions from 152 patients got both histopathological results and CEUS videos successfully, and all the patients gave informed consent. The research institute granted ethical approval (No. 2018200) to carry out the study within its facilities. In the process of data analysis, two patients were excluded for failing to acquire high-quality CEUS data. Finally, 150

Initial population & Inclusion criteria
Step 1: All patients with breast ultrasound examination. **Step 2**: Following to the ACR BI-RADS, those patients diagnosed as suspicious breast lesions were suggested to undergo contrast-enhanced ultrasound (CEUS) examination. **Step 3**: After CEUS, patients scored as 4 category were suggested underwent US-guided FNA or surgery (228 lesions in 228 patients between April.2020 to August.2021).

Exclusion criteria
1. The patients rejecting CEUS examination before FNA or surgery (n=43)
2. The patients failing in following up pathological results (n=33)
3. Patients with breast cysts without entering contrast was excluded (n=2)

Final study population: 150 lesions in 150 patients

49 benignant breast lesions in 49 patients

101 malignant breast lesions in 101 patients

FIGURE 1
Flowchart of patients with breast lesions recruited in this prospective study.

breast lesions from 150 patients were analyzed in this study (Figure 1). Of the 150 suspicious breast lesions, 101 (67.3%) were malignant (16 BI-RADS 4a, 31 BI-RADS 4b, 54 BI-RADS 4c) and 49 (32.7%) were benign (35 BI-RADS 4a, 12 BI-RADS 4b, 2 BI-RADS 4c) (Table 1).

Routine US and CEUS examination

Using the Siemens Acuson Sequoia 512 color Doppler ultrasound diagnostic system, the routine US examination was performed with an 18L6 linear transducer of frequency of 4.6–17.8 MHz, and CEUS was performed with a 10L4 linear transducer of frequency of 2.9–9.9 MHz. Patients were instructed to keep a supine position and expose the breast sufficiently. Ultrasonographers switched the routine model to the contrast model after finding a lesion. The contrast agent, 4.8 mL of SonoVue (Bracco Inc., Milan, Italy), was injected into the peripheral vein of the patient. The dynamic contrast-enhanced process within the lesion was observed and recorded for 2 min. The video data were stored automatically on the machine's hard disk. Based on the grayscale images and the contrast-enhanced videos, we recorded the location, size, shape (regular/irregular), boundary (clear/unclear), blood types, echogenic foci (macro/micro-calcifications), and axillary lymph nodes of lesions. We also recorded the perfusion patterns and the direction of contrast entering the nodules, the enlarged size and the shape of enhanced lesions, and the enhanced vessels within and surrounding lesions. Finally, the DICOM format of the contrast videos was input into a quantitative software called TomTec SonoLiver v1.1.15.0, where the region of interest (ROI) of normal breast gland tissue and lesions (including the whole lesions and partial lesions with solid components) was marked and analyzed. The software provided five parameters based on its default smoothing curves, including: quality of fit (QOF, which measures the perfusion curve fitting degree), maximum intensity (IMAX, the echo intensity of the contrast medium at its peak), rising time (RT, the time from the onset of the contrast

medium and its peak), time to peak (TTP, the time between the start of the contrast agent injection and its peak), and mean transit time (mTT, the mean local transition time of contrast media). The results of the software output in this study were completed by two highly trained radiologists. When the two radiologists disagreed, a third radiologist with over 10 years of experience made the final decision.

Statistical analysis

The qualitative features were compared via the chi-square and Fisher's exact test. The quantitative parameters were compared via a t-test, Bartlett's test, ANOVA test, and Kruskal test. A $p < 0.05$ was considered statistically significant. Single-factor and multi-factor logistic regression with forward stepwise analysis were applied to screen for independent risk factors and establish an integrated model to identify benign and malignant lesions as well as histopathological subtypes of suspicious breast lesions. The ROC curve was used to evaluate the integrated model and calculate the area under the curve (AUC). The cutoffs for the qualitative features and quantitative parameters of CEUS were determined, and the corresponding sensitivity and specificity were calculated based on the maximum value of Youden's index. The DeLong test was used to compare the differences between the AUCs of the ROC curves. All statistical analyses were executed using R software version 4.0 (R codes are shown in the Supplementary material).

Results

Patient information

Of 150 breast lesions scored as category 4 of US-BI-RADS, 101 were breast cancers (16 BI-RADS 4a, 31 BI-RADS 4b, and 54 BI-RADS 4c) and 49 were atypical benign lesions (35 BI-RADS 4a, 12 BI-RADS 4b, and 2 BI-RADS 4c). The average age of the patients with malignancies and benignities was 52.8 and 44.6, with standard deviations of 13.5 and 10.9, respectively. The 49 benign lesions included 20 fibroadenomas (FIB) with atypical grayscale ultrasonic features, 10 hyperplastic nodules, 1 hyperplastic nodule with FIB, 4 intraductal papillomas, and 14 mass-like non-puerperal mastitis (NPM). The 101 malignant lesions included 83 invasive ductal carcinomas (IDC), 12 ductal carcinomas in site (DCIS), 2 invasive lobular carcinomas, 2 intraductal papillary carcinomas, and one mucinous carcinoma (Table 1). The average size of 49 benign lesions is 18.2 mm with a standard deviation of 12.1 mm, and the average size of 101 malignant lesions is 23.2 mm with a standard deviation of 13.3 mm (Supplementary Table 1). Other data, including age, menopausal history, family history, location of lesions, lymphatic metastasis, and ultrasound and contrast-enhanced ultrasound characteristics, are shown in Supplementary Table 1.

Qualitative features of CEUS in differentiating the histopathological subtypes of suspicious breast lesions

Among the risk qualitative features of CEUS, enlarged scope, heterogeneous perfusion, perfusion deficiency, crab clam-like enhancement, more than two enhanced vessels within lesions, and

TABLE 1 Histopathological subtypes of 150 suspicious breast lesions scored as category 4 of US-BI-RADS.

Histopathological subtype	Count (%)
Total (n)	150
Benign	49 (32.7%)
Atypical fibroadenoma	20
Hyperplastic nodule	10
Hyperplastic nodule with fibroadenoma	1
Intraductal papilloma	4
Mass-like non-puerperal mastitis	14
Mammary duct ectasia/periductal mastitis/ serous mastitis	7
Granulomatous lobular mastitis	7
Malignant	101 (67.3%)
Invasive ductal carcinoma	83
Ductal carcinoma in site	12
Invasive lobular carcinoma	2
Intraductal papillary carcinoma	2
Medullary carcinoma	1
Mucinous carcinoma	1

surrounding enriched vessels inserting into lesions represented statistically significant differences for differentiating benign from malignant lesions ($p < 0.01$). Of these features, crab clam-like enhancement had the highest specificity of 95.9%, but the lowest sensitivity of 51.5%, with moderate accuracy of 66.0% (Table 2). The surrounding enriched vessels inserting into lesions had the highest sensitivity of 100% and the highest accuracy of 87.3%, though with moderate specificity of 61.2% (Table 2).

In a comparison of CEUS qualitative features for differentiating the four histopathological subtypes (IDC, DCIS, FIB, and NPM), except the directional perfusion, other risk features showed significant overall differences ($p < 0.01$) (Supplementary Table 2). In the paired comparison, the crab clam-like enhancement, more than two enhanced vessels within lesions, and surrounding enriched vessels inserting into lesions could differentiate IDC and DCIS from atypical FIB and NPM ($p < 0.05$) (Figures 2, 3) but showed no significant difference between IDC and DCIS, or between FIB and NPM ($p > 0.05$) (Figure 2). The enlarged scope, irregular shape, and perfusion deficiency were valuable in differentiating FIBs from the others ($p < 0.05$), although they showed no significant difference among the paired comparison of IDC, DCIS, and mass-like NPM ($p > 0.05$) (Figure 2). The heterogeneous perfusion showed a significant difference only between IDC and atypical FIB ($p < 0.05$) (Figure 2).

Quantitative features of CEUS in differentiating the histopathological subtypes of suspicious breast lesions

After the quantitative analysis, 140 of 150 breast lesions had QOFs over 50% and the QOFs values were higher in the whole enhanced lesions than in the normal enhanced breast gland tissue and partially enhanced lesions with solid composition (Supplementary Figure 1). In a comparison of four quantitative parameters (IMAX, RT, TTP, and mTT) of CEUS in differentiating the benign from malignant breast lesions, only IMAX was significantly higher for the malignancy than the benignity in both whole and part of enhanced lesions ($p < 0.05$). The other three parameters showed no significant difference between malignancies and benignities (Supplementary Figure 2 and Supplementary Table 2).

IMAX showed the highest diagnostic value for differentiating the benign from malignant breast lesions for the whole lesions (AUCROC: 0.62, 95% confidence interval: (0.52: 0.72), the cutoff value of 299.4%, sensitivity: 44.8%, specificity: 77.3%, negative predictive value: 39.1%, positive predictive value: 81.1%, accuracy: 55%, $p < 0.05$), and the partial lesions with solid components (AUCROC: 0.61, 95% confidence interval: (0.51: 0.71), cutoff value: 695.6%, sensitivity: 60.4%, specificity: 65.9%, negative predict value: 43.3%, positive predict value: 79.5%, accuracy: 62.1%, $p < 0.05$) (Table 3 and Supplementary Figure 2). In the paired comparison of the histopathological subtypes of breast lesions, none of the quantitative metrics showed a significant difference in differentiating between IDCs, DCISs, FIBs, and NPMs ($p > 0.05$) (Supplementary Figure 3 and Supplementary Table 4).

Logistic regression analysis of qualitative and quantitative features of CEUS in differentiating the histopathological subtypes of suspicious breast lesions

In the establishment of the logistic regression analysis model, heterogeneous perfusion, crab clam-like enhancement, and partial_IMAX were independent risk factors for benign and malignant breast lesions ($p < 0.05$). Of these features, the crab clam-like enhancement had the highest OR of 30.91. Heterogeneous perfusion and partial_ IMAX OR values are 5.46 and 1.01, respectively (Table 4).

The diagnostic efficacy of the integrated model for suspicious breast lesions is higher than that of independent risk factors (heterogeneous perfusion, crab clam-like enhancement, and partial_ IMAX), with an AUCROC of 0.89 (95% CI: 0.83–0.94). When the cutoff value was 0.608, the sensitivity and specificity were 83.0 and 78.0, respectively, and the accuracy was 81.3% (Supplementary Table 5 and Figure 4).

In diagnosing benign pathological subtypes of breast lesions, there was no significant difference between independent risk factors and integrated model for diagnosis of FIBs and NPMs ($p > 0.05$). In the diagnosis of malignant pathological subtypes of breast lesions, there was no significant difference between independent risk factors and integrated model in the diagnosis of IDCs and DCISs ($p > 0.05$) (Supplementary Table 5).

TABLE 2 Diagnostic performance of qualitative risk features of CEUS in 150 suspicious breast lesions.

Risk features of enhanced lesions	Benign ($n = 49$, yes/no)	Malignant ($n = 101$, yes/no)	SEN (%)	SPE (%)	ACC (%)	P-value[a]
Enlarged scope	18/31	67/34	66.3	63.3	65.3	<0.01
Irregular shape	26/23	71/30	70.3	46.9	62.7	0.06
Directed perfusion	25/24	52/49	51.5	49.0	50.7	1.00
Heterogeneous perfusion	20/29	81/20	80.2	59.2	73.3	<0.01
Perfusion deficiency	20/29	78/23	77.2	59.2	71.3	<0.01
Crab clam-like enhancement	2/47	52/49	51.5	95.9	66.0	<0.01
More than two enhanced vessels within lesions	5/44	61/40	60.4	89.8	70.0	<0.01
Surrounding enriched vessels inserting into lesions	19/30	101/0	100	61.2	87.3	<0.01

[a]Chi-square test. SEN, sensitivity; SPE, specificity; ACC, accuracy.
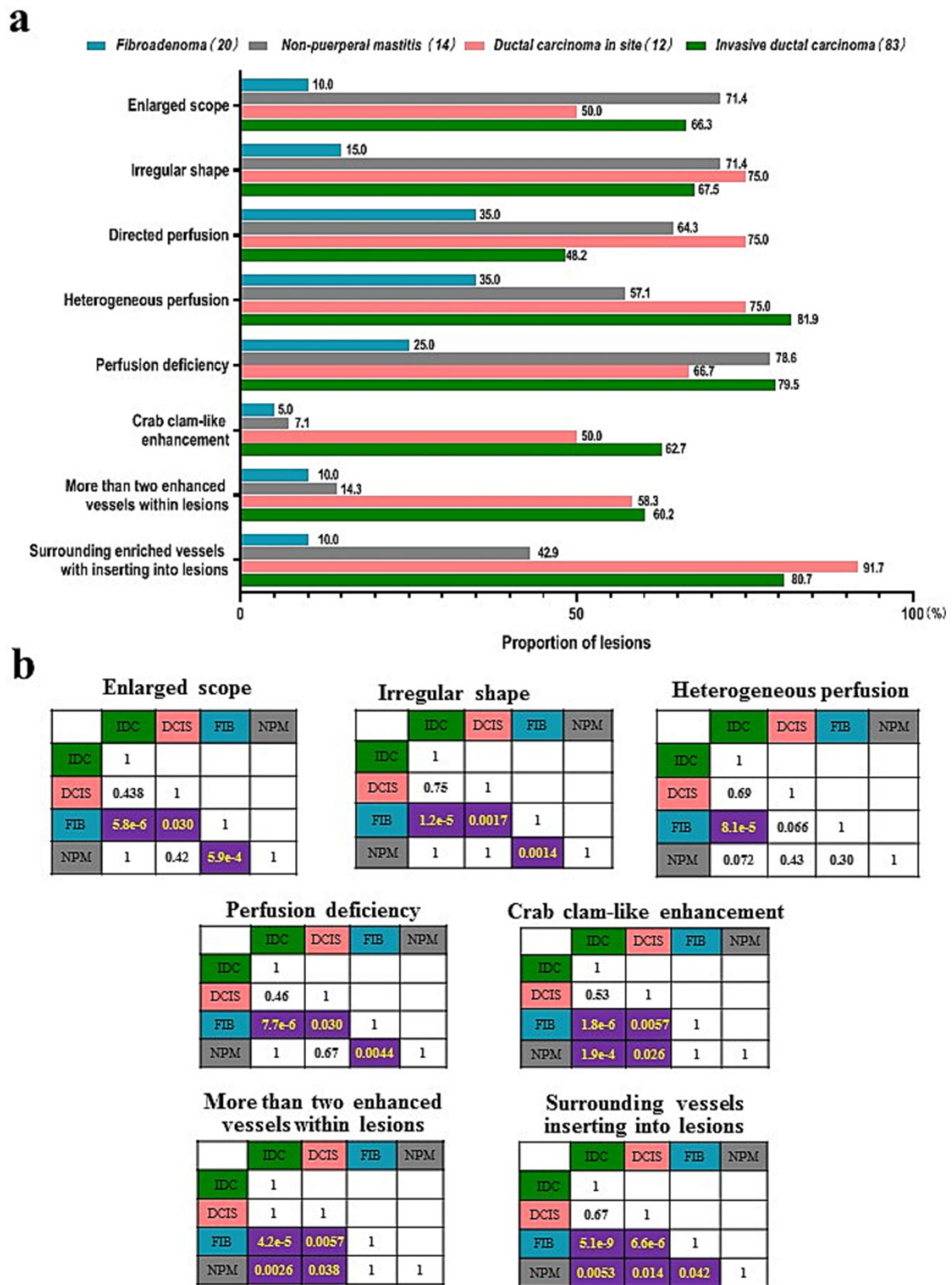
FIGURE 2
Comparison of the qualitative features risk features of CUES among four histopathological subtypes of breast lesions. **(A)** Proportion of the qualitative risk features. **(B)** *p*-values of the paired comparisons of four histopathological subtypes. IDC, invasive ductal carcinoma; DCIS, ductal carcinoma in site; FIB, fibroadenoma; NPM, non-puerperal mastitis.
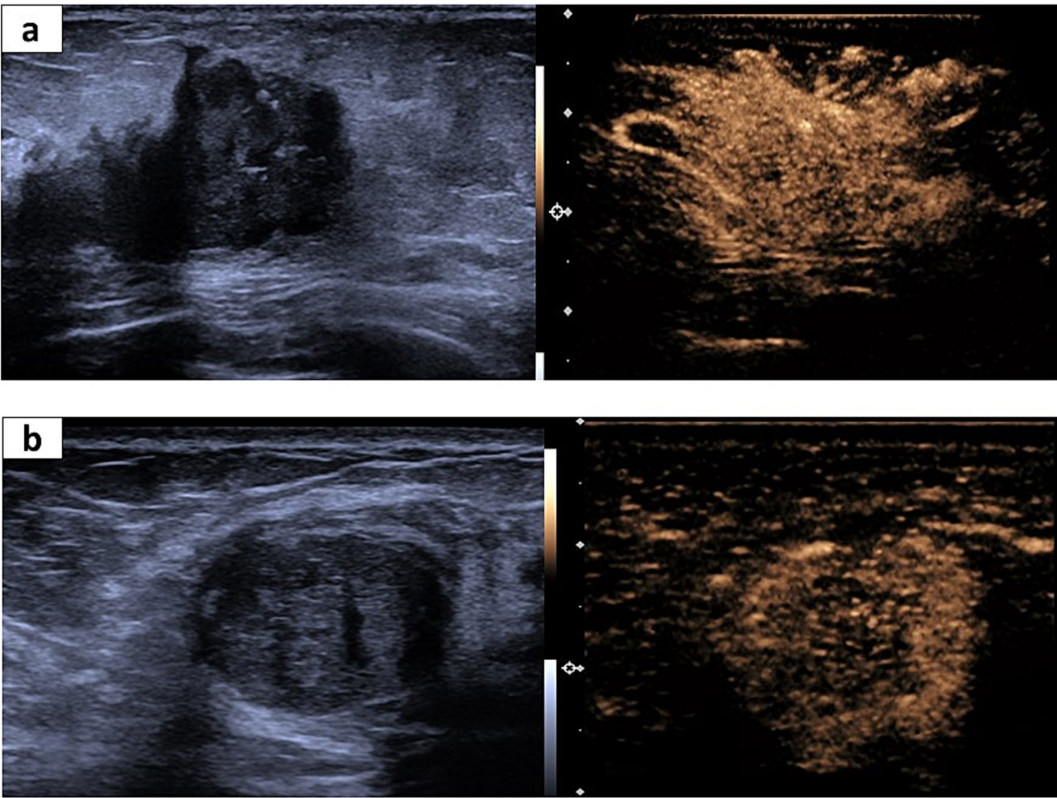
FIGURE 3
Representative US and CEUS images for invasive ductal cancer (IDC) and fibroadenoma. **(A)** A 65-year-old woman suffering from IDC. The irregular shape and partially unclear margin in the US image (left). The corresponding CEUS image shows surrounding enriched vessels inserting into the lesion, clam-like enhancement, and more than two numbers of enhanced vessels within the lesion (right). **(B)** A 56-year-old woman suffering from fibroadenoma. Regular shape and a partially unclear margin in the US image (left). The corresponding CEUS image shows surrounding enriched vessels paralleled to the lesion and <2 numbers of enhanced vessels within the lesion.

TABLE 3 Comparison of four quantitative parameters of CEUS in differentiating atypical benignties from malignant breast lesions.

Quantitative parameters	SEN (%)	SPE (%)	ACC (%)	NPV (%)	PPV (%)	AUROC (%) (95% CI)	Cutoff value
Whole lesion							
Whole_IMAX	44.8	77.3	55.0	39.1	81.1	0.62 (0.52, 0.72)	299.4%
Whole _RT	59.4	54.6	57.9	38.1	74.0	0.56 (0.46, 0.67)	9.0 s
Whole _TTP	92.7	18.2	69.3	53.3	71.2	0.51 (0.41, 0.62)	16.9 s
Whole _mTT	76.0	45.5	66.4	46.5	75.3	0.61 (0.50, 0.71)	29.6 s
Partial lesion							
Partial_IMAX	60.4	65.9	62.1	43.3	79.5	0.61 (0.51, 0.71)	267.3%
Partial _RT	57.3	52.3	55.7	35.9	72.4	0.54 (0.43, 0.64)	7.7 s
Partial _TTP	54.2	59.1	52.1	37.1	74.3	0.52 (0.41, 0.63)	10.7 s
Partial _mTT	88.5	27.3	55.7	52.2	72.6	0.56 (0.45, 0.66)	31.2 s

AUROC, area under receiver operating curve; CI, confidence interval; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; IMAX, maximum intensity; RT, rising time; TTP, time to peak; mTT, mean transit time.

Discussion

In clinical practice, the grayscale ultrasonic images of mass-like NPM and atypical FIB often present risk features similar to those of IDC and DCIS ([18](), [19]()), leading to misdiagnoses and subsequent incorrect therapies. Previous reports have shown that CEUS risk features (including enlarged scope, heterogeneous perfusion, perfusion deficiency, crab clam-like enhancement, more than two

TABLE 4 Logistic regression analysis of CEUS features.

| Variables	β	S.E	Z	P	OR (95% CI)	β	S.E	Z	P	OR (95% CI)
Enlarged scope										
0					1.00 (Ref)					
1	1.57	0.38	4.08	**<0.001**	4.79 (2.26–10.16)					
Irregular shape										
0					1.00 (Ref)					
1	1.01	0.36	2.82	**0.005**	2.75 (1.36–5.54)					
Heterogeneous perfusion										
0					1.00 (Ref)					1.00 (Ref)
1	1.77	0.38	4.62	**<0.001**	5.87 (2.77–12.44)	1.70	0.48	3.54	**<0.001**	5.46 (2.13–14.00)
Perfusion deficiency										
0					1.00 (Ref)					
1	1.59	0.38	4.25	**<0.001**	4.92 (2.36–10.26)					
Crab clam-like enhancement										
0					1.00 (Ref)					1.00 (Ref)
1	3.62	0.75	4.83	**<0.001**	37.36 (8.58–162.59)	3.43	0.77	4.43	**<0.001**	30.91 (6.78–140.83)
More than two enhanced vessels within lesions										
0					1.00 (Ref)					
1	2.60	0.51	5.05	**<0.011**	13.42 (4.90–36.75)					
Whole_ IMAX	0.01	0.00	2.53	**0.011**	1.01 (1.01–1.01)					
Partial_ IMAX	0.01	0.00	2.68	**0.007**	1.01 (1.01–1.01)	0.01	0.00	2.14	**0.032**	1.01 (1.01–1.01)

OR, odds ratio; CI, Confidence interval; Ref, reference. The bold value is $p < 0.05$.
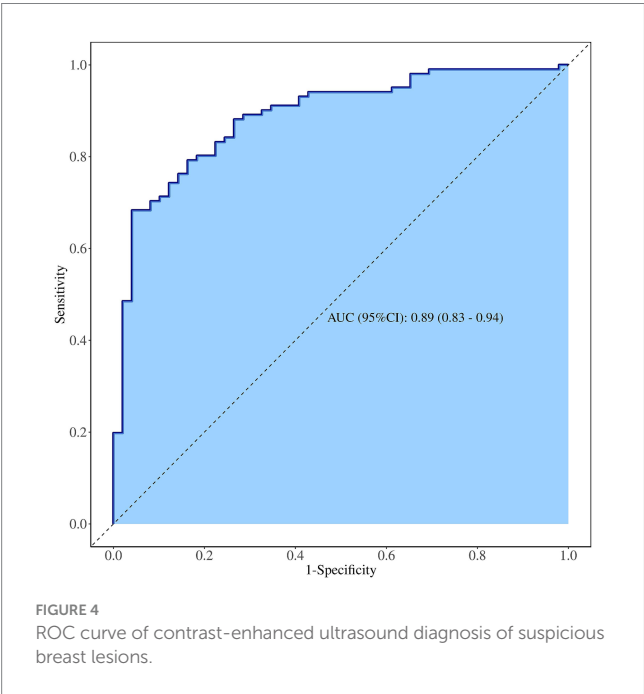
FIGURE 4
ROC curve of contrast-enhanced ultrasound diagnosis of suspicious breast lesions.

enhanced vessels within lesions, and surrounding enriched vessels inserting into lesions) are more specific to malignant than benign lesions (10, 20, 21), consistent with our findings in this study (Table 2). However, previous studies did not assess the potential value of these risk features in differentiating various histopathological subtypes of suspicious breast lesions with high US-BI-RADS scores, especially for atypical benign lesions that are easily mistaken for breast cancers. Therefore, in this study, we explored the value of these CEUS risk features in distinguishing IDC, DCIS, mass-like NPM, and atypical FIB categorized as BI-RADS 4.

Our study showed that the qualitative features of enlarged scope, irregular shape, and perfusion deficiency were less frequent in atypical FIB than IDC, DCIS, and NPM, thus enabling the distinction of atypical FIBs from other types of lesions (Figure 2). These three risk features provide physicians with valuable information for differentiating atypical FIB from breast cancers but fail to differentiate NPMs from IDC and DCIS (Figure 2). This may be related to inflammatory responses or bacterial infections occurring in NPMs, which could stimulate vascular proliferation and infiltrating into the surrounding tissue of lesions (22), subsequently leading to risk features such as an enlarged scope and irregular shape. Additionally, the specific feature of surrounding enriched vessels inserting into lesions differentiates NPMs not only from IDC and DCIS but also from atypical FIB (Figure 2). It demonstrated the highest diagnostic sensitivity of 100% and the highest accuracy of 87.3% in distinguishing between benign and malignant breast lesions (Table 2). Thus, among all the risk features of CEUS, surrounding enriched vessels inserting into lesions would be the most specific in differentiating NPM from breast cancers and atypical FIBs.

In contrast, among the quantitative parameters of CEUS, only IMAX contributed to the differential diagnosis between malignant and benign tumors ($p < 0.05$, ROCAUC: 0.61; sensitivity: 60.4%; specificity: 65.9%; accuracy: 62.1%) (Table 3 and

Supplementary Figure 2). TomTec SonoLiver software used in this study focuses on the quantitative analysis of liver lesions by CEUS. Unlike diffuse liver disease, breast lesions are often accompanied by calcification and necrotic areas, which appear as heterogeneous perfusion, potentially leading to instability in quantitative curves and parameters and ultimately resulting in low diagnostic efficiency. IMAX did not show any difference in the paired comparison of IDC, DCIS, atypical FIB, and mass-like NPM (Supplementary Table 4 and Supplementary Figure 3). These results indicate that the quantitative parameters of CEUS have limited value in differentiating the histopathological subtypes of suspicious breast lesions, consistent with the previous reports (23, 24).

Although our findings suggest that the quantitative parameters of CEUS have limited value in identifying histopathological subtypes of suspicious breast lesions, the results of integrated model constructed by combining quantitative features and quantitative parameters of CEUS show that the model has high diagnostic efficiency (ROCAUC: 0.89; sensitivity: 83.0%; and accuracy: 78.0%) for identifying category 4 of US-BI-RADS and can better distinguish suspicious breast lesions, consistent with previous reports (Supplementary Table 5) (25–29). Unfortunately, it has limited value in identifying the histopathological subtypes of suspicious breast lesions.

Conclusion

Some qualitative risk features of CEUS can distinguish malignant breast lesions from NPMs and atypical FIBs with high US-BI-RADS scores, helping physicians reduce the misdiagnosis of suspicious breast lesions in clinical practice.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by the Second Affiliated Hospital of Xi'an Jiaotong University (No. 2018200). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

RL: Software, Supervision, Writing – original draft, Writing – review & editing. JL: Data curation, Formal analysis, Investigation, Validation, Writing – review & editing. JZ: Data curation, Formal analysis, Writing – review & editing. JiJ: Data curation, Formal analysis, Writing – review & editing. JB: Data curation, Formal analysis, Writing – review & editing. JX: Data curation, Formal analysis, Writing – review & editing. XH: Data curation, Funding acquisition, Resources, Writing – review & editing. SY: Formal analysis, Project administration, Supervision, Validation, Writing – review & editing. QZ: Funding acquisition, Resources, Supervision, Writing – review & editing. JuJ: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1511200/full#supplementary-material

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2024) 74:229–63. doi: 10.3322/caac.21834

2. Geisel J, Raghu M, Hooley R. The role of ultrasound in breast Cancer screening: the case for and against ultrasound. *Semin Ultrasound CT MR*. (2018) 39:25–34. doi: 10.1053/j.sult.2017.09.006

3. Dong H, Huang Y, Song F, Dai H, Liu P, Zhu Y, et al. Improved performance of adjunctive ultrasonography after mammography screening for breast Cancer among Chinese females. *Clin Breast Cancer*. (2018) 18:e353–61. doi: 10.1016/j.clbc.2017.07.014

4. D'Orsi CJ, Sickles E, Mendelson EB. ACR BI-RADS Atlas, Breast imaging reporting and data system. Reston, VA: American College of Radiology (2013).

5. Elverici E, Barça AN, Aktaş H, Özsoy A, Zengin B, Çavuşoğlu M, et al. Nonpalpable BI-RADS 4 breast lesions: sonographic findings and pathology correlation. *Diagn Interv Radiol*. (2015) 21:189–94. doi: 10.5152/dir.2014.14103

6. Elezaby M, Li G, Bhargavan-Chatfield M, Burnside ES, DeMartini WB. ACR BI-RADS assessment category 4 subdivisions in diagnostic mammography: utilization and outcomes in the National Mammography Database. *Radiology*. (2018) 287:416–22. doi: 10.1148/radiol.2017170770

7. Strobel D, Jung EM, Ziesch M, Praktiknjo M, Link A, Dietrich CF, et al. Real-life assessment of standardized contrast-enhanced ultrasound (CEUS) and CEUS algorithms (CEUS LI-RADS®/ESCULAP) in hepatic nodules in cirrhotic patients-a prospective multicenter study. *Eur Radiol*. (2021) 31:7614–25. doi: 10.1007/s00330-021-07872-3

8. Radzina M, Ratniece M, Putrins DS, Saule L, Cantisani V. Performance of contrast-enhanced ultrasound in thyroid nodules: review of current state and future perspectives. *Cancers*. (2021) 13:5469. doi: 10.3390/cancers13215469

9. Huang R, Jiang L, Xu Y, Gong Y, Ran H, Wang Z, et al. Comparative diagnostic accuracy of contrast-enhanced ultrasound and shear wave elastography in differentiating benign and malignant lesions: a network meta-analysis. *Front Oncol*. (2019) 9:102. doi: 10.3389/fonc.2019.00102

10. Du J, Wang L, Wan CF, Hua J, Fang H, Chen J, et al. Differentiating benign from malignant solid breast lesions: combined utility of conventional ultrasound and contrast-enhanced ultrasound in comparison with magnetic resonance imaging. *Eur J Radiol*. (2012) 81:3890–9. doi: 10.1016/j.ejrad.2012.09.004

11. Wan CF, Du J, Fang H, Li FH, Zhu JS, Liu Q. Enhancement patterns and parameters of breast cancers at contrast-enhanced US: correlation with prognostic factors. *Radiology*. (2012) 262:450–9. doi: 10.1148/radiol.11110789

12. Pan J, Tong W, Luo J, Liang J, Pan F, Zheng Y, et al. Does contrast-enhanced ultrasound (CEUS) play a better role in diagnosis of breast lesions with calcification? A comparison with MRI. *Br J Radiol*. (2020) 93:20200195. doi: 10.1259/bjr.20200195

13. Zhou SC, Le J, Zhou J, Huang YX, Qian L, Chang C. The role of contrast-enhanced ultrasound in the diagnosis and pathologic response prediction in breast Cancer: a Meta-analysis and systematic review. *Clin Breast Cancer*. (2020) 20:e490–509. doi: 10.1016/j.clbc.2020.03.002

14. Janu E, Krikavova L, Little J, Dvorak K, Brancikova D, Jandakova E, et al. Prospective evaluation of contrast-enhanced ultrasound of breast BI-RADS 3-5 lesions. *BMC Med Imaging*. (2020) 20:66. doi: 10.1186/s12880-020-00467-2

15. Quan J, Hong Y, Zhang X, Mei M, You X, Huang P. The clinical role of contrast enhanced ultrasound in differential diagnosis of BI-RADS 4 breast disease. *Clin Hemorheol Microcirc*. (2019) 72:293–303. doi: 10.3233/CH-180495

16. Wan CF, Liu XS, Wang L, Zhang J, Lu JS, Li FH. Quantitative contrast-enhanced ultrasound evaluation of pathological complete response in patients with locally advanced breast cancer receiving neoadjuvant chemotherapy. *Eur J Radiol*. (2018) 103:118–23. doi: 10.1016/j.ejrad.2018.04.005

17. Qiao J, Li J, Wang L, Guo X, Bian X, Lu Z. Predictive risk factors for sentinel lymph node metastasis using preoperative contrast-enhanced ultrasound in early-stage breast cancer patients. *Gland Surg*. (2021) 10:761–9. doi: 10.21037/gs-20-867

18. An JK, Woo JJ, Lee SA. Non-puerperal mastitis masking pre-existing breast malignancy: importance of follow-up imaging. *Ultrasonography*. (2016) 35:159–63. doi: 10.14366/usg.15024

19. Gurleyik G, Aktekin A, Aker F, Karagulle H, Saglamc A. Medical and surgical treatment of idiopathic granulomatous lobular mastitis: a benign inflammatory disease mimicking invasive carcinoma. *J Breast Cancer*. (2012) 15:119–23. doi: 10.4048/jbc.2012.15.1.119

20. Wan C, Du J, Fang H, Li F, Wang L. Evaluation of breast lesions by contrast enhanced ultrasound: qualitative and quantitative analysis. *Eur J Radiol*. (2012) 81:e444–50. doi: 10.1016/j.ejrad.2011.03.094

21. Zhao YX, Liu S, Hu YB, Ge YY, Lv DM. Diagnostic and prognostic values of contrast-enhanced ultrasound in breast cancer: a retrospective study. *Onco Targets Ther*. (2017) 10:1123–9. doi: 10.2147/OTT.S124134

22. Stricker T, Navratil F, Forster I, Hürlimann R, Sennhauser FH. Nonpuerperal mastitis in adolescents. *J Pediatr*. (2006) 148:278–81. doi: 10.1016/j.jpeds.2005.08.074

23. Luo J, Chen JD, Chen Q, Yue LX, Zhou G, Lan C, et al. Predictive model for contrast-enhanced ultrasound of the breast: is it feasible in malignant risk assessment of breast imaging reporting and data system 4 lesions? *World J Radiol*. (2016) 8:600–9. doi: 10.4329/wjr.v8.i6.600

24. Liu J, Gao YH, Li DD, Gao YC, Hou LM, Xie T. Comparative study of contrast-enhanced ultrasound qualitative and quantitative analysis for identifying benign and malignant breast tumor lumps. *Asian Pac J Cancer Prev*. (2014) 15:8149–53. doi: 10.7314/apjcp.2014.15.19.8149

25. Ioannidis GS, Goumenakis M, Stefanis I, Karantanas A, Marias K. Quantification and classification of contrast enhanced ultrasound breast Cancer data: a preliminary study. *Diagnostics*. (2022) 12:425. doi: 10.3390/diagnostics12020425

26. Chen X, Yu H, Wei N, Ozcan BB, An G, Wu Q, et al. Diagnostic performance of contrast-enhanced ultrasound combined with shear wave elastography in differentiating benign from malignant breast lesions: a systematic review and meta-analysis. *Gland Surg*. (2023) 12:1610–23. doi: 10.21037/gs-23-333

27. He H, Wu X, Jiang M, Xu Z, Zhang X, Pan J, et al. Diagnostic accuracy of contrast-enhanced ultrasound synchronized with shear wave elastography in the differential diagnosis of benign and malignant breast lesions: a diagnostic test. *Gland Surg*. (2023) 12:54–66. doi: 10.21037/gs-22-684

28. Zhang N, Sun L, Chen X, Song H, Wang W, Sun H. Meta-analysis of contrast-enhanced ultrasound in differential diagnosis of breast adenosis and breast cancer. *J Clin Ultrasound*. (2024) 52:1402–18. doi: 10.1002/jcu.23803

29. Fang J, Deng Q, Zhang J, Ma Y, Jin C, Lu J, et al. A nomogram based on conventional and contrast-enhanced ultrasound for pre-operative prediction of nipple-areola complex involvement in breast Cancer: a prospective study. *Ultrasound Med Biol*. (2023) 49:2565–72. doi: 10.1016/j.ultrasmedbio.2023.08.023

Check for updates

Male-assisted training and injury patterns: hypergraph-enhanced analysis of injuries in women's water polo

Xuehui Feng[1,2]*, Zhibin Wang[1,2], Zheng Wang[1,2], Chen He[1,2], Hongxing Xun[3], Yuanfa Chen[4], Jie Ding[1,2], Gen Chen[1,2] and Zhe Liu[1,2]

[1]Key Laboratory of Sports Trauma and Rehabilitation of General Administration of Sport of the People's Republic of China, Beijing, China, [2]National Research Institute of Sports Medicine (NRISM), Beijing, China, [3]Hunan Institute of Sports Science, Hunan, China, [4]Guangxi Sports Trauma Center, Guangxi, China

Introduction: The aim of this study is to compare the injury patterns of female water polo players before and after the implementation of the Male-Assisted Female Training (MAFT) program. The study seeks to identify key factors influencing these changes and propose corresponding injury prevention measures.

Methods: We utilized pattern analysis and classification techniques to explore the injury data. A Hypergraph Neural Network (HGNN) was employed for pattern extraction, where each athlete was represented as a node in a hypergraph, with node dimensions capturing high-order relational embedding information. We applied the graph Laplacian operator to aggregate neighborhood features and visualize structural and feature differences in hypergraphs based on different influencing factors. Additionally, we introduced graph structure regularization to improve classification accuracy and prevent overfitting in the relatively small dataset, enhancing our ability to identify critical factors affecting injury types.

Results: The analysis revealed significant differences in injury patterns before and after the MAFT program, with specific influencing factors being identified through both pattern recognition and classification techniques. The classification models, supported by graph structure regularization, achieved improved accuracy in distinguishing key features that contributed to changes in injury types.

Discussion: These findings provide insights into the critical factors influencing injury patterns in female water polo players and highlight the effectiveness of the MAFT program in mitigating certain injury risks. Based on the identified features, we propose targeted preventive measures to reduce injury incidence, particularly in relation to changes brought about by the MAFT training mode. Further research is needed to refine these measures and explore their long-term effectiveness.

1 Introduction

Water polo, an aquatic sport that combines swimming, ball skills, and team tactics, has evolved into a global competitive event since its inception in Europe in the late 19th century. The Chinese women's water polo team, established in 2004, has rapidly progressed from regional competitions to the international stage. Their outstanding performance in international tournaments not only showcases the team's strength but

also reflects the advancement of the national sports sector. In recent years, to further enhance their competitive level, the Chinese women's water polo team has introduced the Male-Assisted Female Training (MAFT) program. This innovative training method involves sparring with male athletes to simulate higher-level competitive scenarios, thereby enhancing the female athletes' resilience and tactical execution.

However, the introduction of the MAFT program also presents new challenges, particularly in managing the risk of injuries. Compared to male athletes, female athletes exhibit differences in physical strength and speed, which may increase the risk of injuries during high-intensity sparring sessions. As a team-based combative sport, water polo integrates swimming, throwing, tactical skills, and physical fitness (1). The sport is characterized by intense collisions and grappling in water, lacking the stability of a land environment, leading to frequent injuries during training and matches. Current research on women's water polo primarily focuses on combat techniques, while injury-related studies are relatively scarce. To date, only one publication has analyzed injuries related to the preparation for the Rio Olympics water polo events (2). Therefore, studying the injury patterns of female water polo players under the MAFT program is crucial for developing effective training plans and injury prevention strategies.

Current research on injuries among elite female water polo players is limited and often focuses on a single body part. Studies have investigated shoulder injuries in elite female water polo players (3), analyzing the incidence of shoulder injuries among players in different positions during matches and the frequency of specific shoulder injury sites. It was found that center forwards and top shooters have the highest rates of shoulder injuries, at 88.89% and 80.95% respectively, with the majority of injuries concentrated in the joints and ligaments . Other research has investigated the impact of water polo throws on the shoulder joint (4), analyzing the effects of throwing actions on injured players, assessing the external rotation stability of injured vs. non-injured players, and providing corresponding recommendations. Lv Zhouxiang and colleagues conducted a study on female water polo players (5), analyzing the multiple injury sites and potential causes, but the study did not delve deeply into the associated movement patterns . The MAFT training model, as a newly proposed strategy, has not yet been subject to authoritative research analysis. Therefore, the research presented in this paper on the injury patterns of female water polo players under the MAFT plan is innovative and crucial for developing effective training programs, reducing injuries during training, and enhancing the performance capabilities of athletes.

The interactions among female athletes in practice are complex and diverse. To better understand and analyze these interaction patterns, this study introduces the concept of hypergraphs. A hypergraph (6–10) is a mathematical model capable of representing complex relationships among multiple nodes. It connects multiple nodes through hyperedges (11–13), which can more accurately simulate the many-to-many interactions among female water polo players. For instance, in a match, not only are there defensive and offensive confrontations between individual players, but the strategic coordination between the entire defense

and offense teams can also be represented through hyperedges. This representation method can more comprehensively reflect the tactical layout and collaboration patterns among athletes during matches.

In summary, this study aims to explore the injury patterns, characteristics, and potential coping strategies for female water polo players under the MAFT program. By thoroughly analyzing injury data from training and matches, combined with advanced analytical techniques from hypergraph neural networks (HGNN) (7), we hope to reveal the impact of the MAFT program on the injury risk of female water polo players. The findings will provide scientific evidence for reducing injury incidents, optimizing training methods, and improving athlete performance while ensuring their health and safety. Our contributions are as follows:

1. We analyzed the hypergraph patterns before and after the introduction of the MAFT training program. By comparing the overall hypergraph structure, feature patterns, and the impact of key factors on these structures and patterns, we identified the critical influencing factors associated with the introduction of MAFT.
2. From a classification perspective, we employed graph structure regularization to effectively enhance the accuracy of different injury types in our dataset. This allowed us to more precisely establish an optimized hypergraph structure, thereby identifying the key influencing factors for each injury type before and after the MAFT program.
3. Based on the comprehensive analysis from the aforementioned perspectives, our approach effectively identifies crucial features and subsequently provides recommendations for injury prevention measures.

2 Materials and methods

This study aims to explore the injury patterns, characteristics, and potential coping strategies of female water polo players by comparing injury incidents before and after the implementation of Male-Assisted Female Training (MAFT) and conducting an in-depth analysis of related indicators. To achieve this, we employ the Hypergraph Neural Network (HGNN) as our foundational model to capture the complex relationships and higher-order associative features among athletes, as shown in Figure 1. Subsequently, we perform detailed pattern recognition and factor analysis from a classification perspective to comprehensively analyze the injury characteristics of female water polo players.

The selection of Hypergraph Neural Networks (HGNN) as an analytical method is attributed to its superior capacity in handling complex interrelations and high-dimensional data. In the analysis of injury patterns among elite female water polo players, multiple factors interact and influence each other, such as training load, psychological state, and physiological indicators. Traditional network analysis methods may fail to capture these intricate associations effectively. Firstly, HGNN is a hypergraph-based neural network model capable of managing many-to-many relationships among multiple nodes. This enables HGNN to

FIGURE 1
Overview of Hypergraph-enhanced analysis of injury patterns in women's water polo under Male-Assisting-Female-Training (MAFT) mode.
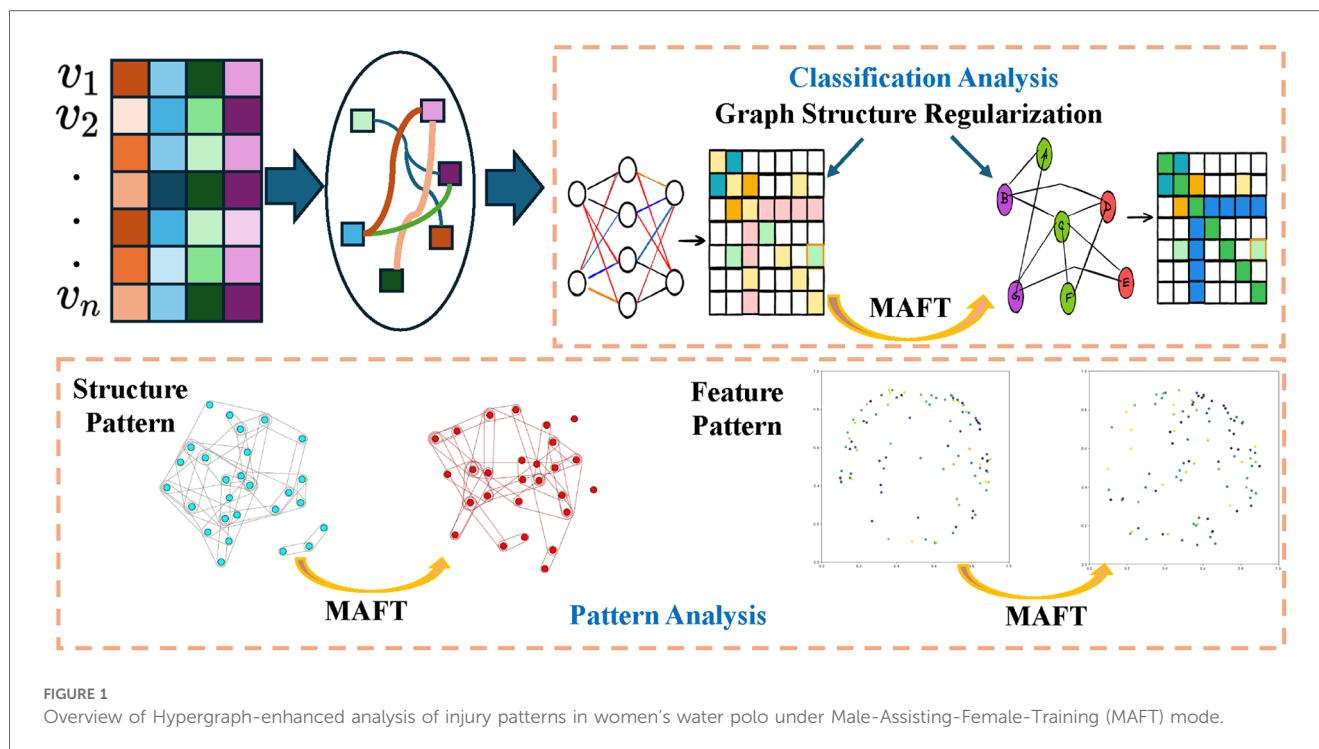
more comprehensively consider the interplay between various factors when analyzing injury patterns in water polo athletes. Secondly, by constructing hypergraphs, HGNN can visualize the association between different factors and extract key features, aiding researchers in identifying significant factors influencing changes in injury patterns. Lastly, in terms of classification analysis, the study introduces graph structural regularization techniques to enhance classification accuracy and more effectively determine key features that distinguish different types of injuries. The results demonstrate that the rHGNN model, which employs HGNN for classification analysis, exhibits excellent performance in terms of accuracy, positive predictive value, negative predictive value, sensitivity, and specificity. In summary, HGNN, as a network analysis method capable of managing complex interrelations and extracting key features, offers distinct advantages in analyzing injury patterns among water polo players. It assists researchers in gaining a more comprehensive understanding of the factors influencing injury patterns and provides scientific evidence for developing personalized training interventions and injury prevention strategies.

2.1 Study design

The study involved 26 athletes from the National Women's Water Polo Team training between February and July 2021, before and after the implementation of the MAFT mode. Among them, 12 were international-level athletes and 14 were national-level athletes. Their ages ranged from 21 to 32 years, with an average age of 24.9 years. The athletes had been participating in professional sports for 8 to 16 years, averaging 10.9 years, and

had been involved in water polo for 8 to 15 years, averaging 9.8 years. They had participated in national-level training for 1 to 14 years, averaging 5.6 years. Twelve athletes had participated in preparation for two Olympic Games. The team positions were distributed as 3 centers, 3 defensive centers, 15 perimeter players, and 5 goalkeepers.

2.2 Data collection

In this study, we employed a combination of surveys and clinical diagnoses to analyze and compare the evolution of injury characteristics before and after the implementation of the Male-Assisted Female Training (MAFT) program. The research team spent an extended period residing with the athletes, closely observing the development and progression of injuries. To gather comprehensive data, we used questionnaires to collect basic information, including names, ages, years of athletic experience, injury locations, and causes. Additionally, we recorded pre-analyzed indicators such as training frequency, intensity, duration, recovery time, and physiological metrics like heart rate, blood lactate levels, and VO2 max. The survey also aimed to identify unique injury patterns that emerged following the implementation of the MAFT program. All questionnaires were meticulously collected and processed. During this period, our team was actively involved in diagnosing and treating all athlete injuries, meticulously documenting the occurrence of common injuries and conducting thorough physical examinations. We paid particular attention to any differences in injury processes and severity compared to previous same-gender training scenarios. In summary, through the combination of survey data

and clinical records, we captured detailed information on injury locations, types, and severity, providing a comprehensive overview of the impact of the MAFT program on female water polo athletes.

2.3 Statistical analysis

Statistical analysis was conducted using SPSS software version 20.0 (SPSS Inc., USA). The proportions of injuries to the neck, shoulder, elbow-wrist, thoracic spine, lumbar-sacral region, hip, knee, ankle, forearm lateral, and sternoclavicular joint were treated as quantitative data. Injury proportions, presented as counts (percentages), were compared between groups using Fisher's exact test to assess the significance of changes in injury patterns before and after the MAFT mode, with the level of significance set at a two-tailed p-value of 0.05. There was a noticeable increase in the proportion of injuries to joints involved in confrontational activities after the implementation of the Male-Assisting-Female-Training mode. The most common injury site was the shoulder, accounting for 34.6% of injuries, which further increased to 42.3% post-implementation. This was followed by injuries to the lumbar-sacral-hip region, with an increase from 26.9% to 34.6%. Injuries to the elbow-wrist area showed a significant upward trend, rising from 7.7% to 26.9%. Conversely, injuries to the knee and ankle joints, primarily involved in non-confrontational sliding motions, showed a decreasing trend. Notably, two goalkeepers sustained injuries, including bilateral forearm ulnar side hitting injuries and a dominant hand side sternoclavicular joint injury. The details of the sports-related injuries, including the number of cases and the proportion of each injury, are summarized in Table 1. In water polo, the increase in proportional injuries to confrontational joints is particularly notable. Shoulder injuries occur at a rate of 24% to 51%, primarily due to the frequent use of the dominant hand in passing and shooting, as well as the non-dominant hand in defensive actions, leading to bilateral shoulder injuries (14, 15). In confrontational training with male athletes, female

athletes need to pass the ball more frequently and attack from various angles, increasing the burden on the dominant hand and raising the injury probability to about 57 Elbow and wrist injuries are also common, with wrist injuries occurring at a rate of 13.6% to 23.1% and elbow injuries at a rate of 6% to 18.2% (16–19). In confrontations with male athletes, female athletes need to find more flexible passing and offensive paths, increasing the burden on the wrist and elbow joints and leading to an increased risk of injury (20, 21). Conversely, the proportion of injuries to stability joints has decreased. The incidence rate of knee and ankle injuries is 4.5% to 10.8% and 6% to 18.2%, respectively (16). Confrontations with male athletes reduce the time spent "treading water," lowering the tension on the knee joint and thus reducing the injury rate. Among goalkeepers, sternal-clavicular joint injuries are a new type of injury. These injuries are usually caused by direct high-impact trauma, while indirect force-related collateral injuries can be alleviated with rest (22, 23). The inability of athletes to get adequate rest during the preparation period is the main reason for persistent pain.

2.4 Hypergraph construction

A hypergraph is a structure capable of representing complex relationships and multidimensional features, making it suitable for capturing interactions and relationships among athletes. In constructing the hypergraph, we first define nodes and hyperedges.

2.4.1 Node definition

In this study, each athlete is represented as a node v_i, with a feature vector $\mathbf{x}_i$ encoding individual attributes such as training intensity, training duration, and injury history. The set of nodes is defined as $V = v_1, v_2, \ldots, v_n$, with the node feature vector defined as Equation 1:

$$\mathbf{x}_i = [\textit{Training Intensity}, \textit{Training Duration}, \textit{Injury History}, \ldots]$$

$$(1)$$

The selection of these features is based on their ability to comprehensively reflect the athlete's condition and performance. Training intensity indicates the effort level during training, where excessive intensity may lead to overtraining and injuries. Accumulated training duration reveals the workload and fatigue accumulation of the athlete, while injury history is a critical indicator for predicting future injury risks. Understanding an athlete's injury history can help formulate more effective prevention strategies. By incorporating these features, we gain a comprehensive understanding of each athlete's training status and health condition, providing a solid foundation for subsequent analysis.

2.4.2 Hyperedge definition

In constructing hyperedges, we consider various interactions among athletes, such as passing and defensive actions. These interactions are modeled as hyperedges e_j, with weights w_j

TABLE 1 This table shows a comparison of injuries to different body parts of female water polo players in traditional training mode and male-assisted Female Training (MAFT) mode. The "Number of Injuries" column in the table shows the frequency of injuries in each part under the two training modes, while the "Injury Proportion" column shows the proportion of injuries in the corresponding part to the total number of injuries. This will provide a key basis for exploring the factors that cause such differences in hypergraph pattern recognition.

Injury	Number of injuries		Injury proportion	
Location	Tradition	MAFT	Tradition	MAFT
Neck	2	1	7.7%	3.8%
Shoulder	9	11	34.6%	42.3%
Elbow & Wrist	2	7	7.7%	26.9%
Chest & Back	1	1	3.8%	3.8%
Lumbar, sacral & hip	7	9	26.9%	34.6%
Knee	11	6	42.3%	23.1%
Ankle	4	3	15.4%	11.5%
Forearm ulnar side	0	1	0%	3.8%
Sternoclavicular joint	0	1	0%	3.8%

representing the frequency and intensity of the interactions. The set of hyperedges is defined as $E = e_1, e_2, \ldots, e_m$, where each hyperedge e_j connects a group of nodes, indicating interactions among athletes. The rationale behind this choice is that interactions among athletes significantly impact team performance and individual injuries. For instance, frequent passing interactions reflect the coordination and trust among athletes, crucial for both offensive and defensive strategies. In defensive scenarios, cooperation and coordination among athletes are equally important, with the frequency and intensity of defensive interactions revealing the execution and effectiveness of team defensive strategies. Moreover, the choice of hyperedges over simple edges is due to their ability to capture the complexity of multi-party interactions. A hyperedge can connect multiple athletes, representing their cooperation or confrontation within a training unit, whereas simple edges can only represent pairwise relationships, failing to comprehensively reflect the complexity of multi-party interactions. Through this approach, the hypergraph can thoroughly represent the complex relationships among athletes, capturing the interaction patterns and individual contributions within the team, thereby providing richer information for subsequent analysis.

2.5 Pattern analysis

To capture the complex relationships within the team, we designed a Hypergraph Neural Network (HGNN) model to learn the embeddings of nodes and hyperedges. The architecture of the HGNN model is designed to capture and represent complex higher-order relationships through a multi-layer structure.

2.5.1 Model architecture

The HGNN model is designed to learn the embeddings of nodes and hyperedges, thereby capturing complex relationships within the team. The embedding function for nodes is defined as Equation 2:

$$\mathbf{z}_i = \sigma(\mathbf{W}^T \mathbf{x}_i + \mathbf{b}) \qquad (2)$$

where $\mathbf{W}$ and $\mathbf{b}$ are learnable parameters, and σ is the activation function. This approach maps the multidimensional features of nodes to a high-dimensional space, facilitating subsequent aggregation and analysis. The activation function σ is typically chosen to be a nonlinear function, such as ReLU (Rectified Linear Unit), to introduce nonlinearity and enhance the model's expressive power.

Specifically, forward propagation is conducted through multiple hypergraph convolution layers to iteratively update the node embeddings:

$$H^{(l+1)}v = \sigma\left(\sum e \in E \frac{1}{|e|} \sum_{u \in e} \mathbf{W}^{(l)} H_u^{(l)} + \mathbf{b}^{(l)}\right) \qquad (3)$$

In Equation 3, $H_u^{(l)}$ denotes the embedding of node u in layer l, $|e|$ is the number of nodes in hyperedge e, and $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the

learnable parameters for layer l. The activation function σ introduces nonlinearity into the model. This formula updates the node embeddings by aggregating the embeddings of all nodes within the hyperedge, which allows the model to capture the intricate relationships among nodes. In each layer, the node embeddings are refined by aggregating the embeddings of all nodes within their respective hyperedges. This multi-layer structure enables the model to capture higher-order relationships and complex patterns in the data.

2.5.2 Pattern recognition

The hyperedge embedding $\mathbf{a}_j$ is aggregated as Equation 4:

$$\mathbf{a}_j = \mathrm{AGG}(\mathbf{z}_i \mid v_i \in e_j) \qquad (4)$$

where AGG is the aggregation function, typically chosen to be mean or weighted mean. The choice of aggregation function significantly impacts the effectiveness of hyperedge embeddings, as it determines how the embeddings of multiple nodes are combined. Subsequently, a multi-layer perceptron (MLP) in Equation 5 is used to identify injury-related patterns:

$$\mathbf{p}_j = \mathrm{MLP}(\mathbf{a}_j) \qquad (5)$$

During training, we simultaneously train the HGNN and MLP models to minimize the reconstruction error, defined as Equation 6:

$$\mathcal{L} = \sum_{j=1}^{m} |\mathbf{a}_j - \mathbf{p}_j|^2 \qquad (6)$$

Through this approach, the HGNN learns efficient representations of nodes and hyperedges, while the MLP identifies injury-related patterns. The minimization of reconstruction error $\mathcal{L}$ ensures that the model accurately captures and reconstructs complex higher-order relationships, thereby improving the accuracy of pattern recognition.

By employing the aforementioned methods, we can construct a higher-order hypergraph structure within the overall information and analyze the variation patterns of hypergraph structures under specific influencing factors or indicators. As the final hypergraph is capable of expressing higher-order associations, we can assess the impact of the MAFT program on female water polo athletes' injuries by comparing the changes in hypergraph structures before and after the implementation of MAFT. Specifically, by evaluating the significance of differences in hypergraph structures and node features, we can identify the key influencing factors.

2.6 Classification analysis

We preprocess the data to obtain a feature vector of total dimension $\mathbb{R}^D (D = 320)$, representing the records of all injury-

related indicators for the corresponding individual. Then, we use KNN to establish the initial hypergraph structure and utilize the features learned by HGNN for injury type prediction. To better learn the hypergraph structure, we analyzed the characteristics of the dataset and found that introducing a graph regularization mechanism preserves the local geometric structure of the hypergraph data more effectively, preventing overfitting and enhancing the model's generalization capability. This leads to better classification accuracy, which is beneficial for accurately extracting key influencing factors of injuries before and after the introduction of the MAFT model. Specifically, given an initialized hypergraph structure $\mathcal{H}(\mathcal{V}, \mathcal{E}, \mathbf{W})$, where the symbols denote a set of vertices $\mathcal{V}$, a set of hyperedges $\mathcal{E}$, and a weight matrix $\mathbf{W}$, which is a diagonal matrix representing the weights of the hyperedges. This hypergraph can be succinctly represented by an incidence matrix $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$, where each entry $h(v, e)$ is defined as Equation 7:

$$h(v, e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e, \end{cases} \quad (7)$$

Graph regularization in HGNNs aims to preserve the local geometric structure of hypergraph data and improve the model's generalization capability. This is achieved by adding a regularization term to the loss function that measures the smoothness of node embeddings. Given the $\mathcal{H} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, the incidence matrix $\mathbf{H}$, the vertex degree matrix by $\mathbf{D}_v$, and the hyperedge degree matrix by $\mathbf{D}_e$. The Laplacian matrix $\mathbf{L}$ for a hypergraph can be defined as:

$$\mathbf{L} = \mathbf{D}_e - \mathbf{H}^T \mathbf{W} \mathbf{H}$$

where $\mathbf{D}_e$ is the degree matrix of hyperedges, $\mathbf{H}$ is the incidence matrix, and $\mathbf{W}$ is the weight matrix. During training, we wish for the low-dimensional representations $\mathbf{X}$ of the vertices to preserve the local structure of the hypergraph. The regularized objective function for HGNNs can be written as Equation 8:

$$J_{\text{rHGNN}} = \frac{1}{2} \mathbf{X}^T \mathbf{L} \mathbf{X} + \lambda \Omega(\mathbf{X}) \quad (8)$$

Here, J_{rHGNN} is the regularized loss function, rHGNN represents the regularized form of HGNN in our proposed model, $\Omega(\mathbf{X})$ represents other possible regularization terms (such as weight decay), and λ is the regularization parameter that controls the strength of the regularization term. The optimization process involves minimizing J_{HGNN} through gradient descent or other optimization algorithms, thereby learning the low-dimensional representations $\mathbf{X}$ of the vertices while maintaining the local structure of the hypergraph. This regularization method helps improve the performance of HGNNs on various downstream tasks such as node classification and clustering.

Notably, The regularization term Ω is constructed to ensure that the node embeddings preserve the local structure of the hypergraph. Specifically, $\Omega(\mathbf{X})$ includes both a weight decay term

that discourages overly complex models and a graph Laplacian smoothness term that encourages nearby nodes in the hypergraph to have similar embeddings. Mathematically, $\Omega(\mathbf{X})$ is defined as Equation 9:

$$\Omega(\mathbf{X}) = \alpha \cdot \left(\sum_{i,j} \mathbf{W}_{ij} \|\mathbf{X}_i - \mathbf{X}_j\|^2 \right) + \beta \cdot \left(\sum_i \|\mathbf{X}_i\|^2 \right) \quad (9)$$

where $\mathbf{W}_{ij}$ represents the elements of the graph Laplacian $\mathbf{L}$, $\mathbf{X}_i$ and $\mathbf{X}_j$ are the embeddings of nodes i and j, respectively. The first term encourages smooth embeddings across the hypergraph, while the second term, weight decay, penalizes large embedding values. α and β are hyperparameters that control the relative importance of the smoothness and weight decay terms, respectively.

3 Experiments

3.1 Evaluation and metrics

For pattern analysis, we focus primarily on visualization differences. This involves visualizing the structure and features of the hypergraph after optimizing reconstruction loss. We visualize differences in the global hypergraph and the MAFT-induced changes under single-factor influences. Specifically, when visualizing the hypergraph structure, we highlight the clustering of nodes. For visualizing hypergraph features, we employ dimensionality reduction techniques to create 2D visualizations.

For classification analysis, to assess the model's performance, we utilize several key metrics. Firstly, Accuracy (ACC) is calculated as $\frac{TP+TN}{TP+TN+FP+FN}$, reflecting the model's overall ability to correctly classify instances across all categories. Sensitivity (SEN), given by $\frac{TP}{TP+FN}$, measures how effectively the model identifies true positive cases. Specificity (SPEC), defined as $\frac{TN}{TN+FP}$, evaluates the model's capability to accurately recognize true negative cases without misclassifying them as positive. The Positive Predictive Value (PPV), calculated as $\frac{TP}{TP+FP}$, represents the fraction of correctly identified positive cases among all cases predicted as positive. Lastly, the Negative Predictive Value (NPV), given by $\frac{TN}{TN+FN}$, indicates the proportion of true negatives among the cases predicted as negative. These metrics collectively provide a comprehensive evaluation of the model's classification performance.

3.2 Implementation details

For pattern analysis, the MLP consists of two hidden layers with 64 and 32 neurons respectively, followed by a ReLU activation function. The output layer uses a sigmoid activation function to produce the final pattern recognition output. The HGNN and MLP models are trained simultaneously using the Adam optimizer with a learning rate of 0.001. The training process includes early stopping with a patience parameter set to 10 epochs to prevent overfitting. The models are trained for a total of 100 epochs or until the validation loss stops improving.

For classification analysis, the rHGNN model is trained by minimizing the Equation 8. We employ stochastic gradient descent (SGD) as our optimization algorithm, with a learning rate of 0.01 and a decay schedule to adjust the learning rate over time. The training process is carefully monitored to ensure convergence towards a minimum of the loss function. The hyperparameters λ, α and β in the regularization term $\Omega(\mathbf{X})$ are crucial for balancing the weight decay and smoothness constraints. The default values are set as $\lambda = 0.2$, $\alpha = 0.03$, and $\beta = 0.005$. For more details on the ablation experiments, please refer to the subsequent sections.

3.3 Comparison methods

We selected several comparative methods for our evaluation, including MLP, SVM, GNN, B-GNN, HGNN, and HGNNP.

- **MLP** (24): MLP is a class of feedforward artificial neural networks consisting of multiple layers of nodes, each fully connected to the next. It is widely used for classification and regression tasks due to its ability to capture non-linear relationships.
- **SVM** (25): SVM is a supervised learning algorithm that is effective for both classification and regression challenges. It works by finding the optimal hyperplane that maximizes the margin between different classes in the feature space.
- **GNN**: GNNs (26) are designed to perform inference on data described by graphs. They leverage the graph structure to perform node classification, link prediction, and graph

classification tasks by aggregating feature information from neighboring nodes.
- **B-GNN** (27): B-GNN is a scalable graph neural network designed to handle large-scale graph data. It introduces techniques to efficiently manage large graphs while maintaining performance, making it suitable for big data applications.
- **HGNN** (7): HGNN extends traditional GNNs to hypergraphs, which can capture higher-order relationships among data points. This method is particularly effective in scenarios where interactions involve more than two entities.
- **HGNNP** (13): HGNNP is an enhanced version of HGNN that includes additional mechanisms to improve its performance. It further refines the ability to capture complex relationships in hypergraph-structured data.

Our proposed method, rHGNN, represents the regularized form of HGNN, incorporating regularization techniques to improve generalization and performance in the context of our specific application.

4 Discussion

4.1 Study on pattern analysis

Figure 2 illustrates the distribution of various features in a two-dimensional space, comparing the scenarios with and without the application of MAFT. The figure is divided into four subplots, each representing different types of features: Global Feature, Training Load Factor, Psychological State Factor, and

FIGURE 2
We visualize the positions of all features in the hypergraph within a standard 2D space. By visualizing the feature changes before and after the implementation of the MAFT program, we can observe the most impactful factors. We highlight the three most significant factors: training load, psychological state, and physiological indicators.
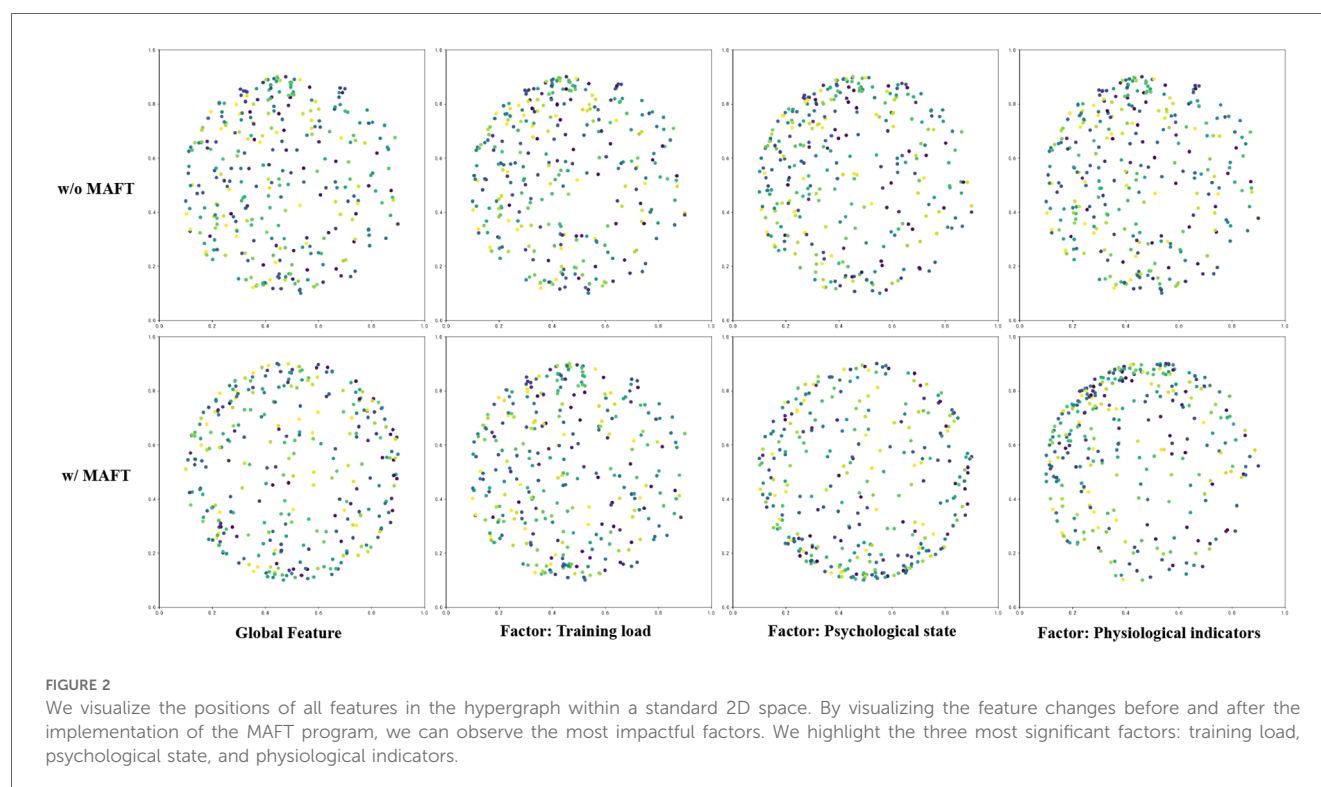
Physiological Indicators Factor. In the case of Global Features, it is observed that without MAFT, the data points are sparsely distributed with no apparent clustering pattern. However, when MAFT is applied, the data points become more densely packed and exhibit a certain degree of structural organization. This indicates that MAFT effectively enhances the correlation among global features. For Training Load Factors, a similar trend is observed. Without MAFT, the data points appear scattered and lack discernible patterns. Upon applying MAFT, the points converge into tighter clusters, suggesting that MAFT can extract more meaningful features from training load data. When examining Psychological State Factors, it is evident that without MAFT, the data points are randomly dispersed with no significant clustering. With MAFT applied, small clusters begin to form among the data points. This transformation implies that psychological state data becomes more consistent and interpretable after undergoing feature transformation through MAFT. Lastly, for Physiological Indicators Factors, the untransformed data exhibits a disordered scatter. However, post-MAFT application, the data points reveal clearer structural patterns. This suggests that physiological indicators processed through MAFT better reflect their intrinsic relationships. In summary, the comparison clearly demonstrates that applying MAFT results in more compact and organized distributions across all types of features.

Figure 3 illustrates the changes in hypergraph structures for specific factor indicators before and after implementing the MAFT program, focusing on Heart Rate Variability (HRV) and Training Frequency. Before MAFT, the HRV hypergraph shows numerous dispersed connections with some isolated nodes, indicating weak correlations. This disorganized structure may hinder effective information capture by models. After applying MAFT, the HRV hypergraph becomes more structured and cohesive, with fewer isolated nodes. Enhanced connectivity suggests improved correlation among nodes, facilitating better data utilization by models. Similarly, the pre-MAFT hypergraph for Training Frequency is characterized by scattered connections and weak node associations. This loose network structure can impede meaningful feature extraction. Post-MAFT application reveals a more organized network with tighter clusters and

stronger internal associations. This improved structure enhances data consistency and information flow, aiding models in accurately capturing training frequency impacts.

4.2 Study on classification analysis

Table 2 presents a comprehensive evaluation of various models, highlighting the performance of our proposed method, rHGNN. The results demonstrate that rHGNN consistently outperforms other methods across all evaluated metrics, including Accuracy (ACC), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity (SEN), and Specificity (SPEC). The ACC for rHGNN is notably high at 0.90635, surpassing traditional models such as MLP and SVM, which achieve ACCs of 0.68637 and 0.65817 respectively. This indicates a significant improvement in overall model accuracy. In terms of PPV and NPV, rHGNN achieves values of 0.92572 and 0.91092 respectively, reflecting its superior ability to correctly predict positive and negative cases compared to other models like GNN and B-GNN. Furthermore, the SEN value for rHGNN is the highest among all methods at 0.93216, demonstrating its exceptional sensitivity in identifying true positive cases. Similarly, SPEC is also maximized at 0.92785, indicating robust specificity in distinguishing true negatives. Overall, these results underscore the efficacy of the rHGNN model in achieving superior performance across multiple dimensions compared to conventional approaches such as HGNN and HGNNP. The consistent excellence across all metrics suggests that our method offers a highly reliable solution for the given task, setting a new benchmark for future research endeavors in this domain.

4.3 Study on graph regularization

In Table 3, we provide a detailed and comprehensive analysis of how different regularization hyperparameters affect our accuracy. It is evident that λ is optimal around the 0.1 range, α is best suited for the 0.01 range, and β performs well in the 0.001 range.

FIGURE 3
We visualize the changes in hypergraph structures under specific factor indicators before and after the implementation of the MAFT program, highlighting the significant structural differences from the heart rate variability (HRV) and training frequency indicators.
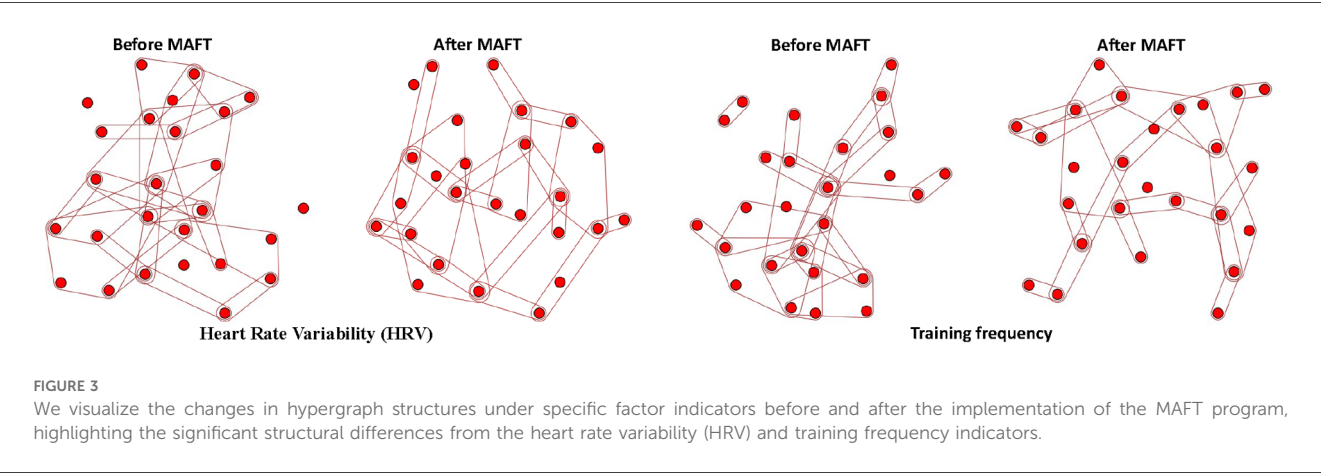
TABLE 2 We compute the accuracy of the proposed method on the testing data, and our method achieves the best results.

Method	ACC	PPV	NPV	SEN	SPEC
MLP	$0.68637_{\pm 0.0625}$	$0.68762_{\pm 0.0594}$	$0.69993_{\pm 0.0627}$	$0.68873_{\pm 0.0571}$	$0.69026_{\pm 0.0608}$
SVM	$0.65817_{\pm 0.0297}$	$0.66373_{\pm 0.0308}$	$0.65843_{\pm 0.0325}$	$0.67367_{\pm 0.0362}$	$0.66267_{\pm 0.0398}$
GNN	$0.78716_{\pm 0.0514}$	$0.77621_{\pm 0.0572}$	$0.76523_{\pm 0.0583}$	$0.77276_{\pm 0.0517}$	$0.75862_{\pm 0.0572}$
B-GNN	$0.82796_{\pm 0.0428}$	$0.82781_{\pm 0.0478}$	$0.81872_{\pm 0.0412}$	$0.82664_{\pm 0.0368}$	$0.81654_{\pm 0.0378}$
HGNN	$0.86621_{\pm 0.0381}$	$0.85163_{\pm 0.0365}$	$0.88245_{\pm 0.0327}$	$0.87833_{\pm 0.0392}$	$0.86924_{\pm 0.0461}$
HGNNP	$0.86273_{\pm 0.0367}$	$0.86252_{\pm 0.0387}$	$0.87326_{\pm 0.0352}$	$0.86237_{\pm 0.0371}$	$0.87944_{\pm 0.0308}$
rHGNN (Ours)	$\mathbf{0.90635}_{\pm 0.0288}$	$\mathbf{0.92572}_{\pm 0.0286}$	$\mathbf{0.91092}_{\pm 0.0349}$	$\mathbf{0.93216}_{\pm 0.0365}$	$\mathbf{0.92785}_{\pm 0.0367}$

Bold values represent the highest scores.

TABLE 3 Comparison of model performance under different hyperparameters.

Method	ACC	PPV	NPV	SEN	SPEC
rHGNN (Full)	$\mathbf{0.90635}_{\pm 0.0288}$	$\mathbf{0.92572}_{\pm 0.0286}$	$\mathbf{0.91092}_{\pm 0.0349}$	$\mathbf{0.93216}_{\pm 0.0365}$	$\mathbf{0.92785}_{\pm 0.0367}$
$\lambda = 0.01$	$0.89049_{\pm 0.0338}$	$0.90780_{\pm 0.0215}$	$0.89159_{\pm 0.0442}$	$0.91773_{\pm 0.0333}$	$0.91677_{\pm 0.0382}$
$\lambda = 0.1$	$0.89199_{\pm 0.0233}$	$0.91970_{\pm 0.0240}$	$0.90656_{\pm 0.0416}$	$0.93089_{\pm 0.0348}$	$0.92467_{\pm 0.0351}$
$\lambda = 1.0$	$0.87687_{\pm 0.0427}$	$0.89560_{\pm 0.0207}$	$0.88212_{\pm 0.0314}$	$0.90082_{\pm 0.0274}$	$0.91050_{\pm 0.0273}$
$\alpha = 0.001$	$0.88735_{\pm 0.0236}$	$0.90289_{\pm 0.0352}$	$0.89358_{\pm 0.0452}$	$0.91173_{\pm 0.0274}$	$0.90539_{\pm 0.0415}$
$\alpha = 0.01$	$0.89949_{\pm 0.0184}$	$0.91696_{\pm 0.0311}$	$0.90673_{\pm 0.0225}$	$0.92826_{\pm 0.0511}$	$0.91930_{\pm 0.0346}$
$\alpha = 0.1$	$0.88597_{\pm 0.0147}$	$0.89671_{\pm 0.0222}$	$0.88850_{\pm 0.0397}$	$0.91942_{\pm 0.0438}$	$0.90754_{\pm 0.0235}$
$\beta = 0.0001$	$0.89944_{\pm 0.0406}$	$0.91289_{\pm 0.0352}$	$0.90275_{\pm 0.0263}$	$0.92961_{\pm 0.0441}$	$0.90086_{\pm 0.0267}$
$\beta = 0.001$	$0.88206_{\pm 0.0286}$	$0.90511_{\pm 0.0153}$	$0.90949_{\pm 0.0338}$	$0.92774_{\pm 0.0297}$	$0.91953_{\pm 0.0251}$
$\beta = 0.01$	$0.88584_{\pm 0.0321}$	$0.89636_{\pm 0.0291}$	$0.90519_{\pm 0.0490}$	$0.92634_{\pm 0.0367}$	$0.91799_{\pm 0.0385}$

Bold values represent the highest scores.

When examining the impact of varying the hyperparameter λ, it is observed that smaller values such as $\lambda = 0.01$ result in a slight decrease in ACC to 0.89049, while PPV and NPV also show minor reductions compared to the full model. As λ increases to 1.0, there is a more pronounced decline in ACC to 0.87687, indicating that larger values may negatively affect overall accuracy. Adjustments to the hyperparameter α reveal similar trends; for instance, at $\alpha = 0.001$, ACC drops to 0.88735 with corresponding decreases in other metrics such as PPV and NPV compared to the baseline model's performance. For the hyperparameter β, we observe that at $\beta = 0.0001$, ACC remains relatively high at 0.89944. However, increasing β leads to lower performance metrics, with $\beta = 0.01$ resulting in an ACC of 0.88584. Overall, the full model (rHGNN) consistently outperforms variations with different hyperparameters. This suggests that optimal tuning plays a crucial role in achieving superior model performance.

4.4 Study on MAFT

The introduction of the MAFT mode has had a significant impact on the injury patterns among female water polo players. HGNN provided valuable insights into the complex interactions within the team and identified specific patterns associated with increased injury risk. These findings suggest that tailored training interventions and injury prevention strategies should be developed, considering the unique demands of the MAFT mode. Future research should focus on validating these findings and exploring additional applications of HGNN in sports injury

prevention. We recommend paying attention to the following points when implementing the MAFT training program:

1. It is essential to adjust the frequency of training sessions to prevent overtraining and potential injuries. Sufficient rest periods should be incorporated into the training plan to allow athletes to fully recover. Excessive training frequency can lead to fatigue, decreased performance, and an increased risk of injury.
2. Continuous monitoring of heart rate is crucial to ensure that athletes maintain their heart rate within a safe range during training. Wearable devices can be used for real-time heart rate monitoring, allowing coaches to adjust training intensity as needed. This approach helps prevent cardiovascular strain and optimize performance.
3. Providing psychological support and counseling is vital for helping athletes cope with the stress and challenges of high-intensity training programs like MAFT. Establishing open communication channels enables athletes to express their needs and feedback, fostering a supportive environment.

5 Conclusion

This study investigated the impact of the Male-Assisted Female Training (MAFT) program on the injury patterns of female water polo players through hypergraph-based pattern analysis and classification perspectives. We first summarized the overall changes in injuries among female athletes under the MAFT program. Using the collected data, we conducted pattern analysis on the hypergraph structure and features, identified key influencing factors, and proposed enhanced preventive measures within the MAFT framework. Additionally, we analyzed the

impact characteristics of various injuries before and after the implementation of MAFT from a classification standpoint, incorporating graph regularization techniques to achieve the highest classification accuracy. Our main findings are as follows:

- The introduction of the MAFT program significantly altered the injury patterns among female water polo players, particularly increasing the proportion of joint injuries involved in confrontational activities.
- Hypergraph Neural Networks (HGNN) provided in-depth insights into the complex interactions within the team and identified specific patterns associated with increased injury risk.
- Our rHGNN model, enhanced by graph regularization techniques, excelled in classification accuracy, positive predictive value, negative predictive value, sensitivity, and specificity, providing reliable scientific evidence for injury prevention.

These findings have important practical implications for water polo training and injury prevention. By adjusting the frequency of MAFT training sessions, continuously monitoring heart rates, and providing psychological support and counseling, tailored training interventions and injury prevention strategies can be developed to meet the unique demands of the MAFT mode. These strategies help optimize training methods, enhance athlete performance, and ensure their health and safety. Future research directions:

- Validate the findings of this study and explore additional applications of HGNN in sports injury prevention.
- Further investigate the impact of the MAFT program on athlete performance and injury risk, especially under varying training intensities and durations.
- Research how to more effectively monitor the physiological and psychological states of athletes through technological means, such as wearable devices, to adjust training plans in real-time.
- Explore the applicability of the MAFT program to athletes of different levels and age groups, and how to adjust training methods based on individual differences.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

XF: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. ZhiW: Writing – original draft, Writing – review & editing. ZheW: Writing – original draft, Writing – review & editing. CH: Writing – original draft, Writing – review & editing. HX: Writing – review & editing. YC: Writing – review & editing. JD: Writing – original draft. GC: Writing – review & editing. ZL: Writing – original draft.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Franić M, Ivković A, Rudić R. Injuries in water polo. *Croat Med J*. (2007) 48:281–8.

2. Smith J, Doe A, Lee B. Analysis of water polo injuries during 8904 player matches at FINA World Championships and Olympic games to make the sport safer. *Br J Sports Med*. (2019) 53(5):1–10. doi: 10.1136/bjsports-2018-099349

3. Schroeder GG, McClintick DJ, Trikha R, Kremen TJ Jr. Injuries affecting intercollegiate water polo athletes: a descriptive epidemiologic study. *Orthop J Sports Med*. (2022) 10(7):23259671221110208. doi: 10.1177/23259671221110208

4. Suzuki S, Tomoda Y, Miyanaga Y, Takahashi G, Sakata I, Shiraki H. Shoulder injuries caused by throwing during water polo. *Jpn J Phys Fit Sports Med*. (1990) 39(1):60–8. doi: 10.7600/jspfsm1949.39.60

5. Xiaocui Y. *Research on shoulder injuries in high-level female water polo players in China* (Master's thesis). Chengdu Sport University (2023).

6. Di D, Shi F, Yan F, Xia L, Mo Z, Ding Z, et al. Hypergraph learning for identification of COVID-19 with CT imaging. *Med Image Anal*. (2021) 68:101910. doi: 10.1016/j.media.2020.101910

7. Feng Y, You H, Zhang Z, Ji R, Gao Y. Hypergraph neural networks. In: *AAAI 2019* (2018).

8. Gao Y, Zhang Z, Lin H, Zhao X, Du S, Zou C. Hypergraph learning: methods and practices. *IEEE Trans Pattern Anal Mach Intell*. (2020) 44:2548–66.

9. Huang H, Chen M, Duan Y. Dimensionality reduction of hyperspectral image using spatial-spectral regularized sparse hypergraph embedding. *Remote Sens*. (2019) 11:1039. doi: 10.3390/rs11091039

10. Jiang J, Wei Y, Feng Y, Cao J, Gao Y. Dynamic hypergraph neural networks. In: *IJCAI*. (2019). p. 2635–41.

11. Di D, Yang J, Luo C, Xue Z, Chen W, Yang X, et al. Hyper-3dg: text-to-3d gaussian generation via hypergraph. *arXiv* [Preprint]. *arXiv:2403.09236* (2024).

12. Di D, Zhang J, Lei F, Tian Q, Gao Y. Big-hypergraph factorization neural network for survival prediction from whole slide image. *IEEE Trans Image Process*. (2022) 31:1149–60. doi: 10.1109/TIP.2021.3139229

13. Gao Y, Feng Y, Ji S, Ji R. Hgnn+: General hypergraph neural networks. *IEEE Trans Pattern Anal Mach Intell*. (2022) 45:3181–99. doi: 10.1109/TPAMI.2022.3182052

14. Miller AH, Evans K, Adams R, Waddington G, Witchalls J. Shoulder injury in water polo: a systematic review of incidence and intrinsic risk factors. *J Sci Med Sport*. (2018) 21:368–77. doi: 10.1016/j.jsams.2017.08.015

15. Webster M, Morris M, Williams G. Shoulder pain in water polo: impact on performance in training and playing. *J Sci Med Sport*. (2010) 13:e78. doi: 10.1016/j.jsams.2010.10.627

16. Croteau F, Brown H, Pearsall D, Robbins SM. Prevalence and mechanisms of injuries in water polo: a systematic review. *BMJ Open Sport Exerc Med*. (2021) 7: e001081. doi: 10.1136/bmjsem-2021-001081

17. Junge A, Langevoort G, Pipe A, Peytavin A, Wong F, Mountjoy M, et al. Injuries in team sport tournaments during the 2004 olympic games. *Am J Sports Med*. (2006) 34:565–76. doi: 10.1177/0363546505281807

18. Mountjoy M, Junge A, Alonso JM, Engebretsen L, Dragan I, Gerrard D, et al. Sports injuries and illnesses in the 2009 fina world championships (aquatics). *Br J Sports Med*. (2010) 44:522–7. doi: 10.1136/bjsm.2010.071720

19. Mountjoy M, Junge A, Benjamen S, Boyd K, Diop M, Gerrard D, et al. Competing with injuries: injuries prior to and during the 15th fina world championships 2013 (aquatics). *Br J Sports Med*. (2015) 49:37–43. doi: 10.1136/bjsports-2014-093991

20. Stromberg JD. Care of water polo players. *Curr Sports Med Rep*. (2017) 16:363–9. doi: 10.1249/JSR.0000000000000409

21. Zhang J. *Comparative study on the defensive capabilities of the Chinese women's water polo team and the top three teams at the 2019 world championships* (Ph.D. thesis). Chengdu Sport University (2021).

22. Ingoe HMA, Mohammed K, Malone AA, Beadle G, Sharpe T, Cockfield A, et al. Traumatic posterior sternoclavicular joint dislocation – current aspects of management. *Injury*. (2023) 54(11):110983. doi: 10.1016/j.injury.2023.110983

23. Wang Y, Chen J. Advances in the treatment of instability and degenerative sternoclavicular arthritis. *Chin J Shoulder Elbow Surg*. (2014) 2:263–9.

24. Warth RJ, Lee JT, Campbell KJ, Millett PJ. Arthroscopic sternoclavicular joint resection arthroplasty: a technical note and illustrated case report. *Arthrosc Tech*. (2014) 3(1):e165–73. doi: 10.1016/j.eats.2013.09.019

25. Cortes C. Support-vector networks. *Mach Learn*. (1995) 20(3):273–97. doi: 10.1023/A:1022627411411

26. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv* [Preprint]. *arXiv:1609.02907* (2016).

27. Wang J, Wang Y, Yang Z, Yang L, Guo Y. Bi-GCN: binary graph convolutional network. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2021). p. 1561–70.

EEG-based epilepsy detection with graph correlation analysis

Chongrui Tian[1,2] and Fengbin Zhang[1]*

[1]School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China, [2]School of Information Engineering, East University of Heilongjiang, Harbin, China

Recognizing epilepsy through neurophysiological signals, such as the electroencephalogram (EEG), could provide a reliable method for epilepsy detection. Existing methods primarily extract effective features by capturing the time-frequency relationships of EEG signals but overlook the correlations between EEG signals. Intuitively, certain channel signals exhibit weaker correlations with other channels compared to the normal state. Based on this insight, we propose an EEG-based epilepsy detection method with graph correlation analysis (EEG-GCA), by detecting abnormal channels and segments based on the analysis of inter-channel correlations. Specifically, we employ a graph neural network (GNN) with weight sharing to capture target channel information and aggregate information from neighboring channels. Subsequently, Kullback-Leibler (KL) divergence regularization is used to align the distributions of target channel information and neighbor channel information. Finally, in the testing phase, anomalies in channels and segments are detected by measuring the correlation between the two views. The proposed method is the only one in the field that does not require access to seizure data during the training phase. It introduces a new state-of-the-art method in the field and outperforms all relevant supervised methods. Experimental results have shown that EEG-GCA can indeed accurately estimate epilepsy detection.

KEYWORDS

electroencephalogram, graph neural networks, correlation analysis, anomaly detection, abnormal EEG channels detection

1 Introduction

The field of affective computing has witnessed significant development, drawing attention to emotion detection, especially in medical research related to epilepsy (1). While epilepsy, as a neurological disorder, manifests symptoms that encompass seizures, it often intertwines with fluctuations in emotional states. These emotional variations, a common symptom in epilepsy patients, are crucial for accurate disease monitoring and treatment (2).

Scalp electroencephalogram (EEG) stands as the primary tool for detecting seizures, capturing voltage changes between electrodes and providing spatial-temporal insights into brain activity (3–5). However, the current approach to seizure detection in EEGs relies on manual examination by experienced EEG readers, demanding substantial time and effort. Furthermore, discrepancies in diagnostic results may emerge due to varying opinions among experts (6).

To address these challenges, there is a pressing need for the development of automated and objective methods for epileptic seizure detection. While many studies have proposed deep learning (DL)-based models for automated seizure detection, several challenges persist (7–9). These models often train in a supervised approach, necessitating labeled seizure data that is both scarce and labor-intensive to obtain in real-world applications. Additionally, existing models frequently apply deep convolutional neural networks (CNNs) directly to time-series signals or spectrograms, overlooking crucial information related to physical distance-based and functional-based connectivity between different brain regions (10).

Recent studies have introduced graph learning techniques to capture relationships between EEG electrodes (i.e., EEG nodes) (6, 11, 12). However, these approaches fall short in considering local patterns, such as local sub-graphs and sub-structures, when learning EEG graphs. The inclusion of such local information could prove effective in detecting anomalies in EEG graphs, as demonstrated in other network analysis applications. In real-world applications, an imbalance in data availability between seizure and normal classes is common. Graph-based methods addressing this issue often employ graph augmentation, but not every augmentation technique is effective in EEG graphs (10), as some may compromise underlying brain region connectivities. Therefore, identifying appropriate augmentation strategies in EEG graphs that preserve semantic information is crucial for accurate seizure detection and localization (13).

This study delves into detecting the anomaly channels of EEG signal in patients with epilepsy (14). We propose an innovative method for epilepsy detection that distinctively focuses on exploring the inter-channel relationships within EEG signals, deemed essential for understanding the patient signal variations. We introduce an anomaly detection approach for EEG channels and segments based on inter-channel correlation analysis. This method utilizes Graph Neural Networks (GNNs) (15, 16) to capture the correlation between different channels, providing a more accurate reflection of anomaly changes. To achieve precise detection of anomaly channels in an EEG signal, we propose an EEG-based epilepsy detection method with graph correlation analysis (EEG-GCA), employing a weight-sharing GNN and aligning different channel information distributions with Kullback-Leibler (KL) (17) divergence regularization. During the testing phase, we detect anomalous channels and segments by measuring the correlation between two views, thereby achieving sensitive identification of abnormalities in epilepsy. Notably, our proposed method not only performs well in experiments but is also the only training approach that does not require access to seizure data. This research holds practical significance in improving the effectiveness of epilepsy patient treatment.

- We proposed a method named EEG-GCA for inter-channel correlation analysis simulating the correlation between channels in EEG, revealing subtle differences in patient anomaly changes. This algorithm provides a new approach to EEG signal processing.
- We redefined the anomaly channel detection of EEG as the correlation between channel feature distribution and their neighbors' distribution, and we designed an Unsupervised model to verify the effectiveness.
- The performance evaluation of the proposed abnormal EEG node and region detection is conducted on the extensive and comprehensive EEG seizure dataset TUSZ. The results demonstrate that EEG-GCA sets a new benchmark, achieving state-of-the-art performance on this dataset.

2 Related works

2.1 EEG analysis

Electroencephalogram analysis has become one of the prominent directions (18, 19). The following is a review of relevant work in this field, focusing on the application of different methods and technologies.

(a) Early approaches to epilepsy recognition primarily relied on traditional feature extraction techniques combined with machine learning algorithms (20, 21). Researchers extracted features from different domains, including time-domain, frequency-domain, and time-frequency-domain features, such as power spectral density and energy, to capture epilepsy-related patterns from EEG signals (22). Common machine learning models used in these early approaches included support vector machines (SVM) and decision trees (23, 24). While these methods achieved some success, their performance was often limited by the challenges of manually extracting relevant features and their inability to fully capture the complex dynamics of EEG signals.

(b) In recent years, deep learning methods have gained significant attention for their ability to enhance EEG-based epilepsy recognition (25). Architectures such as convolutional neural networks (CNNs) (26, 27) and recurrent neural networks (RNNs) (28) have been successfully applied, allowing models to learn feature representations in an end-to-end fashion. These deep learning techniques excel at capturing abstract and complex features from the raw EEG signals, significantly improving the accuracy of epilepsy recognition (24, 29). Furthermore, techniques such as transfer learning and multimodal fusion have been extensively explored to improve the generalization capabilities of these models, enabling better performance on unseen data.

(c) Beyond EEG signals, there has been growing interest in integrating data from multiple modalities for epilepsy recognition tasks, including physiological signals, speech, and images (30). Cross-modal research aims to combine information from diverse sources, thereby enhancing the robustness and comprehensiveness of epilepsy detection systems (31, 32). This approach leverages complementary data to improve model performance, offering a more holistic view of the patient's condition and enhancing the reliability of diagnosis (33).

2.2 Canonical correlation analysis

Canonical correlation analysis (CCA) (34, 35) is a method that aims to find the linear transformation for measuring the relationship between two vectors. Give two vectors $\mathbf{X}_1$ and $\mathbf{X}_2$, the

correlation $\rho = \frac{\mathbf{a}^T \Sigma_{\mathbf{X}_1 \mathbf{X}_2} \mathbf{b}}{\sqrt{\mathbf{a}^T \Sigma_{\mathbf{X}_1 \mathbf{X}_1} \mathbf{a}} \sqrt{\mathbf{b}^T \Sigma_{\mathbf{X}_2 \mathbf{X}_2} \mathbf{b}}}$ is maximized by optimizing the objective :

$$\max_{\mathbf{a},\mathbf{b}} \mathbf{a}^T \Sigma_{\mathbf{X}_1 \mathbf{X}_2} \mathbf{b}, s.t. \mathbf{a}^T \Sigma_{\mathbf{X}_1 \mathbf{X}_1} \mathbf{a} = \mathbf{b}^T \Sigma_{\mathbf{X}_2 \mathbf{X}_2} \mathbf{b} = \mathbf{I} \qquad (1)$$

Soft-CCA (36) considers the decorrelation constraint as a term of loss and optimizes it jointly with other terms, and the objective of Soft CCA is:

$$\max_{\theta 1, \theta 2} Tr(P_{\theta_1}^T(\mathbf{X}_1) P_{\theta_2}(\mathbf{X}_2))$$
$$s.t. P_{\theta_1}^T(\mathbf{X}_1) P_{\theta_1}(\mathbf{X}_1) = P_{\theta_2}^T(\mathbf{X}_2) P_{\theta_2}(\mathbf{X}_2) = \mathbf{I} \qquad (2)$$

where $\mathbf{I}$ is the identity matrix, and Equation 2 can be rewritten as:

$$\min_{\theta 1, \theta 2} ||P_{\theta_1}(\mathbf{X}_1) - P_{\theta_2}(\mathbf{X}_2)||_F^2 +$$
$$\lambda(\mathcal{L}_{SDL}(P_{\theta_1}(\mathbf{X}_1)) + \mathcal{L}_{SDL}(P_{\theta_2}(\mathbf{X}_2))) \qquad (3)$$

where P_{θ_1} and P_{θ_2} are the neural networks used to learn the representations of the two views. $||P_{\theta_1}(\mathbf{X}_1) - P_{\theta_2}(\mathbf{X}_2)||_F^2$ is used to maximize the correlation between the two views, and $\mathcal{L}_{SDL}$ is used to minimize the distance between $P_{\theta_i}(\mathbf{X}_i)$ and the identity matrix.

2.3 Graph learning methods

Graph data, being non-Euclidean, poses a challenge for traditional convolution methods. The effective learning of information from graph data is an actively researched problem (37). In the context of graph data, the learned representation of nodes should encapsulate both the structural information of the graph and the attributes associated with each node. Existing graph learning methods can be broadly categorized as follows:

Truncated Random Walk-Based Methods: These methods operate on the assumption that nodes with similar network structures should have similar vector representations. A notable approach in this category is DeepWalk (38), which employs random walks to generate training data and leverages Word2vec (39) to learn node representations. Node2vec (40) captures homogeneity and structural equivalence through weighted random walks.

Methods Based on k-Order Distance Between Nodes in the Graph: These approaches, exemplified by methods like LINE (41) and GraRep (42), learn node representations by capturing k-order relational structure information, aiming to achieve high-quality node embeddings.

Deep Learning-Based Methods: Distinguished by their use of deep learning, these methods (43, 44) leverage the advantages of deep neural networks to extract high-order nonlinear relationships from graph data.

Graph neural networks (GNNs) (45) represent a significant advancement as they directly operate on graph data, aggregating each node's features with those of its neighbors. Building on GNNs, certain methods (46, 47) utilize GNNs to learn node representations. They employ adversarial learning to regularize these representations and predict the likelihood of an edge existing between a pair of nodes. However, these approaches predominantly rely on graph structure information.

Moreover, methods based on dual-autoencoders, such as AnomalyDAE (45) and Dual-SVDAE (48), use Graph Convolutional Networks to capture graph structure information. They combine this with multi-layer perceptrons (MLPs) to capture node attribute information, thereby making full use of attribute network information.

3 Method

In this section, we detail the EEG-GCA in Figure 1. It consists of a graph construct module, the information mining model, and a correlation analysis module. At first, we construct the EEG graph as input for our model. Then, we introduce an identity graph that represents the identity matrix, signifying no relationships between the channels. This graph aims to capture the features of each channel in the EEG data. Then, we input the EEG graph and identity graph into a weight-sharing GCN to learn the distribution of structural information and distribution of semantic information and pull the distributions to the same prior distribution through the Kullback-Leibler (KL) divergence. Finally, we sample the network structure embedding and node embedding from the learned distribution and maximize the correlation of normal nodes on the network structure distribution and node attribute distribution by using the CCA-based objective. The correlation score is used to detect the anomaly channels.

3.1 EEG graph construction

In this paper, we first construct the EEG graph as input. The EEG graph can be defined as an attributed network $\mathcal{G} = \{\mathbf{A}, \mathbf{X}\}$. Where $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix that denotes the connection between each electrode. $\mathbf{X} \in \mathbb{R}^{N \times D}$ denotes the feature matrix. $\mathbf{X}_i$ is feature of the i-th channel. Similar to the study, given an EEG clip, we construct five types of EEG graphs (12).

- $Dist-EEG-Graph$ strives to embed the structure of electrode locations in the graph's adjacency matrix by leveraging the Euclidean distance between electrodes. Given that electrode locations remain fixed within an EEG recording cap, the same adjacency matrix is applied to all EEG clips. More precisely, the elements a_{ij} of the Dist-EEG-Graph are computed as follows:

$$a_{ij} = \begin{cases} \exp(-\frac{||v_i - v_j||}{\tau^2}), & if \; ||v_i - v_j||^2 \leq k, \\ 0, & if \; O.W. \end{cases} \qquad (4)$$

  Here, $|| \bullet ||$ represents the l_2-norm, and τ is a scaling constant. The proximity between two electrodes, v_i and v_j, is reflected by the proximity of a_{ij} to 1. In this paper, k is uniformly set to 0.9 across all EEG clips. Assigning a value of 0 to a_{ij} for distant electrodes introduces sparsity to the graph.
- $Corr-EEG-Graph$ The purpose of this graph is to capture the functional connectivity between electrodes, which is encoded

FIGURE 1
The framework of the proposed EGG-GCA.
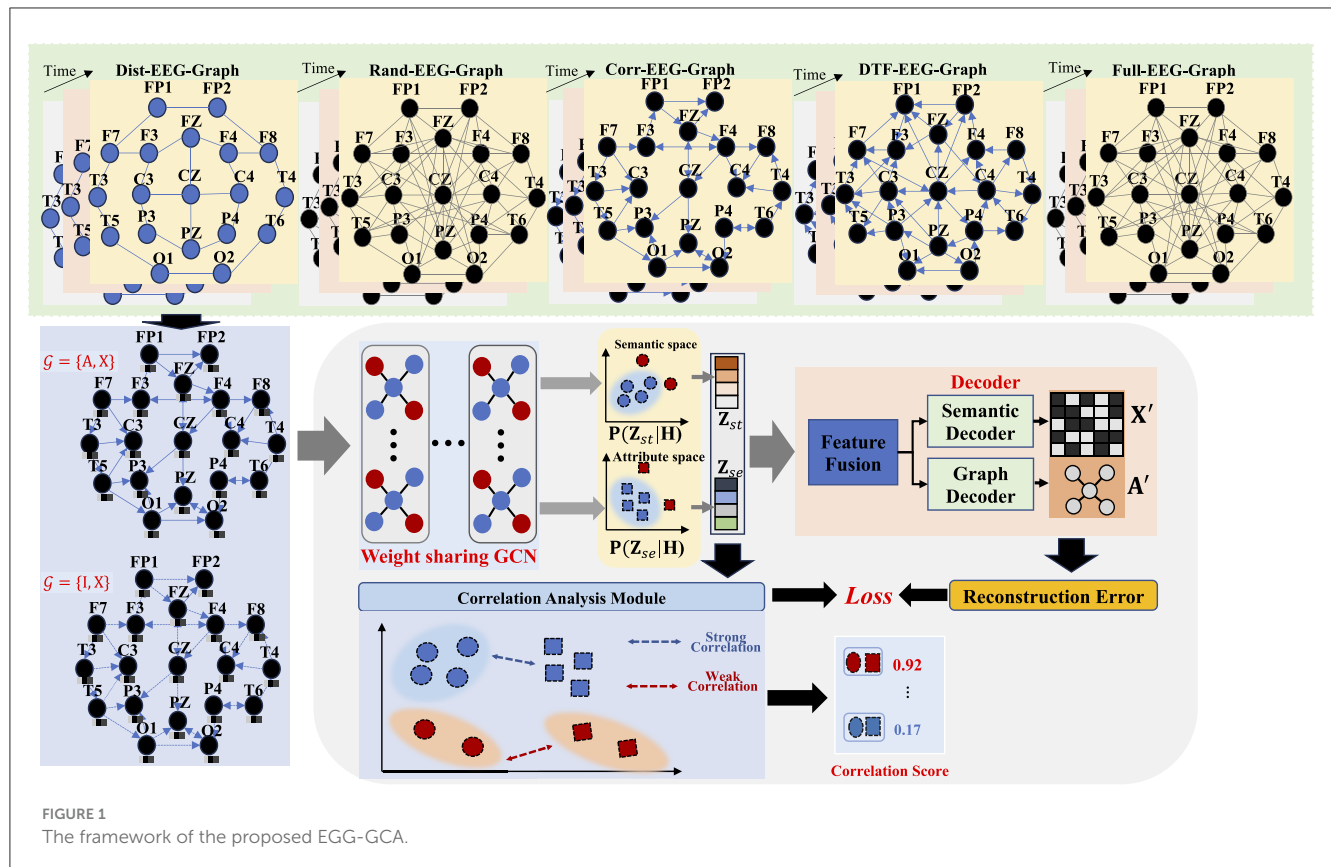
in the elements of the adjacency matrix defined as follows:

$$a_{ij} = \begin{cases} \frac{corr(\mathbf{X}_i, \mathbf{X}_j)}{||\mathbf{X}_i|| ||\mathbf{X}_j||}, & if \ v_j \in \mathcal{N}(v_i), \\ 0, & if \ O.W. \end{cases} \quad (5)$$

where $corr(\bullet)$ denotes the cross-correlation function, and v_i represents the top-3 neighborhood nodes of v_i with the highest normalized correlation. $\mathcal{N}(v_i)$ is set to the top-3 neighborhood nodes to avoid overly connected graphs. Additionally, we only keep the top three edges for each node to prevent excessively connected graphs.

- $Rand - EEG - Graph$ The construction of this graph is grounded on the assumption that all electrodes are interconnected and equally contribute to brain activities. The realization of this graph involves the formation of an adjacency matrix according to the following procedure:

$$a_{ij} = \begin{cases} 0.5, & if \ i \neq j, \\ 1, & if \ O.W. \end{cases} \quad (6)$$

- $Full - EEG - Graph$ Similar to the $Rand - EEG - Graph$, The construction of this graph is grounded on the assumption that all electrodes are interconnected and equally contribute to brain activities. But the a_{ij} is set as 1 for each connection.
- $DTF - EEG - Graph$ The Directed Transfer Function Graph aims to represent the mutual influence between EEG channels, thereby modeling the functional connectivity of different brain

regions. The adjacency matrix for this graph is defined as follows:

$$a_{ij} = \begin{cases} \frac{corr(\mathbf{X}_i, \mathbf{X}_j)}{\sqrt{\sum_{m=1, m \neq i, j}^{n} ||corr(\mathbf{X}_i, \mathbf{X}_m))||^2}}, & if \ v_j \in \mathcal{N}(v_i), \\ 0, & if \ O.W. \end{cases} \quad (7)$$

3.2 Weight-sharing GCN

To learn the correlation within the weight-sharing Graph Convolutional Network (GCN) for capturing the semantic and structural information of each node, we introduced an identity graph denoted as $\mathcal{G}' = \{I, X\}$, where I represents the identity matrix signifying no relationships between the channels. This approach enhances the similarity between the semantic information and the graph structure information of each node by transferring the learned semantic information to all node features. Consequently, each channel feature can be obtained by inputting the identity graph into the Weight-Sharing GCN.

The construct EEG-graph $\mathcal{G}$ explicitly expresses the correlations between the channels in the EEG data, therefore, to capture the relationship information (structural information) between different channels, we input the EEG graph $\mathcal{G} = \{A, X\}$ to the weight-sharing GCN.

$$GCN(\mathbf{X}, \mathbf{A}|\mathbf{W}) = \varphi((\mathbf{D})^{-\frac{1}{2}} \mathbf{A} (\mathbf{D})^{-\frac{1}{2}} \mathbf{X} \mathbf{W}) \quad (8)$$

where W is the learnable sharing weight, φ is activation function, and $\mathbf{D}$ is the diagonal degree matrix of the constructed EEG graph $\mathcal{G}$.

To extract each channel information (node semantic information), the identity aggregation is designed which inputs the identity graph $\mathcal{G}' = \{\mathbf{I}, \mathbf{X}\}$ to the weight-sharing GCN:

$$GCN(\mathbf{X}, \mathbf{I}|W) = \varphi(\mathbf{IXW}) \tag{9}$$

3.3 Distribution alignment

After obtain the node structural information embedding $\mathbf{Z}_{st}$ and the node semantic information embedding $\mathbf{Z}_{se}$, we capture the structural distribution $q(\mathbf{Z}_{st}|\mathbf{X}, \mathbf{A})$ and semantic distribution $q(\mathbf{Z}_{se}|\mathbf{X}, \mathbf{I})$ for each node by Equation 10, respectively.

$$q(\mathbf{Z}|\mathbf{X}, \mathbf{A}) = \prod_{i=0}^{N} q(\mathbf{z}_i|\mathbf{X}, \mathbf{A}) \tag{10}$$

$$q(\mathbf{z}_i|\mathbf{X}, \mathbf{A}) = \mathcal{N}(\mathbf{z}_i|\boldsymbol{\mu}_i, \mathrm{diag}(\boldsymbol{\sigma}^2)) \tag{11}$$

where $\mathbf{Z}$ is the embedding sampled from the distribution. $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\sigma}$ is the variance vector, which is learned by two different GCN layers.

$$\boldsymbol{\mu} = \mathrm{GCN}_{\boldsymbol{\mu}}(\mathbf{H}, \mathbf{A}|W) \tag{12}$$

$$\boldsymbol{\sigma} = \mathrm{GCN}_{\boldsymbol{\sigma}}(\mathbf{H}, \mathbf{A}|W) \tag{13}$$

where $\boldsymbol{\mu}_h$ and $\boldsymbol{\sigma}_h$ denote the mean and variance vectors of the structural distribution learned by Equations 12, 13. Similarly, $\boldsymbol{\mu}_f$ and $\boldsymbol{\sigma}_f$ are the mean and variance vectors of semantic distribution learned by Equations 12, 13.

To capture the correlation between the two distributions, we should align the structural distribution and semantic distribution. Due to it being harder to directly align two distributions, we use a Gaussian distribution as prior distribution p and use Kullback-Leibler (KL) divergence to align the two distributions wanting this prior distribution to achieve the desired effect.

$$\mathcal{L}_{kl} = -\mathrm{KL}[q(\mathbf{Z}_{st}|\mathbf{X}, \mathbf{A})||p(\mathbf{Z}_{st})] - \mathrm{KL}[q(\mathbf{Z}_{se}|\mathbf{X}, \mathbf{I})||p(\mathbf{Z}_{se})] \tag{14}$$

3.4 Decoder

The reconstruction of graph data is divided into two main parts, the reconstruction of the network structure and the reconstruction of the node attributes. Since nodes in graph data often have complex interactions with each other, it is necessary to fuse the features of each node with those of their neighbors.

$$\mathbf{Z}_f = \mathbf{Z}_{st} + \mathbf{Z}_{se} \tag{15}$$

Then we use an L-layers Multi-Layer Perceptron (MLP) to reconstruct the node attributes.

$$\mathbf{Z}_f^{(l)} = \sigma(\mathbf{Z}_f^{(l-1)}\mathbf{W}^{(l-1)} + \mathbf{b}^{(l-1)}) \tag{16}$$

where $\mathbf{Z}_f^{(l-1)}$, $\mathbf{Z}^{(l)}$, $\mathbf{W}^{(l-1)}$ and $\mathbf{b}^{(l-1)}$ are the input, output, the trainable weight and bias matrix of $(l-1)$-th layer respectively, $l \in \{1, 2, ..., L\}$. $\sigma(\bullet)$ is the activation function. Finally, the reconstruction of node attributes $\hat{\mathbf{X}} = \mathbf{Z}_f^{(L)}$ is obtained from the output of the last layer in MLP.

For the reconstruction of the network structure, we use an inner production of fusion embedding $\mathbf{Z}_f$ to reconstruct the network structure.

$$\hat{\mathbf{A}} = \mathbf{Z}_f\mathbf{Z}_f^T \tag{17}$$

The reconstruction loss is defined as:

$$\mathcal{L}_{dec} = ||\mathbf{X} - \hat{\mathbf{X}}|| + ||\mathbf{A} - \hat{\mathbf{A}}|| \tag{18}$$

3.5 Correlation analysis objective

The objective of correlation analysis is to discern the relationship between structural distribution and semantic distribution. Initially, we sample the embeddings of structural information, denoted as $\mathbf{Z}st$, and semantic information, denoted as $\mathbf{Z}se$, from the distributions of structural features $q(\mathbf{Z}st|\mathbf{X}, \mathbf{A})$ and semantic features $q(\mathbf{Z}se|\mathbf{X}, \mathbf{I})$. Subsequently, we normalize the node embeddings for the two perspectives using the following procedure.

$$\begin{aligned} \mathbf{Z}'_{st} &= \frac{\mathbf{Z}_{st} - \mu(\mathbf{Z}_{st})}{\sigma(\mathbf{Z}_{st}) * N^{\frac{1}{2}}} \\ \mathbf{Z}'_{se} &= \frac{\mathbf{Z}_{se} - \mu(\mathbf{Z}_{se})}{\sigma(\mathbf{Z}_{se}) * N^{\frac{1}{2}}} \end{aligned} \tag{19}$$

Subsequently, as per the formulation in Equation 3, EEG-GCA enhances the correlation between the distributions of the two views by minimizing the invariance between the network structure embedding $\mathbf{Z}_{st}$ and the node attribute embedding $\mathbf{Z}f$. The invariance loss, denoted as $\mathcal{L}inv$, is defined as:

$$\mathcal{L}_{inv} = ||\mathbf{Z}'_{st} - \mathbf{Z}'_{se}||_F^2 \tag{20}$$

To prevent collapsed solutions, we introduce the decorrelation loss, denoted as $\mathcal{L}_{dco}$, which aims to guarantee that the individual dimensions of the features are uncorrelated.

$$\mathcal{L}_{dco} = ||\mathbf{Z}'^T_{st}\mathbf{Z}'_{st} - \mathbf{I}||_F^2 + ||\mathbf{Z}'^T_{se}\mathbf{Z}'_{se} - \mathbf{I}||_F^2 \tag{21}$$

The CCA-based objective is defined as follows:

$$\mathcal{L}_{CCA} = \mathcal{L}_{inv} + \lambda * \mathcal{L}_{dco} \tag{22}$$

where λ is the trade-off between the two terms.

3.6 Loss function and anomaly score

The training objective of the proposed model involves optimizing the CCA-based loss along with minimizing the Kullback-Leibler (KL) divergence

TABLE 1 Train and test sets of TUSZ used in the supervised method and unsupervised method.

Data	Patients (% SZ)	EEG files (% SZ)	EEG clips (% SZ)
$Train_{Sup}$	591 (34.0%)	4,599 (18.9%)	38,613 (9.3%)
$Train_{ours}$	493 (0%)	4,028 (0%)	35,019 (0%)
Test	45 (77.8%)	900 (25.6%)	8,848 (14.7%)

The percentages of the seizure data (SZ) is indicated in parenthesis.

between the network structure distribution and the node attribute distribution.

$$\mathcal{L} = \mathcal{L}_{CCA} + \mathcal{L}_{KL} + \mathcal{L}_{dec} \qquad (23)$$

The anomaly score is defined as the correlation between channels with their structure information.

4 Performance evaluation

4.1 Dataset

In this study, we employed the Temple University Hospital EEG Seizure Corpus (TUSZ) v1.5.2 (12) as the benchmark dataset. This dataset stands out due to its extensive inclusion of seizure categories and patient samples, making it the dataset with the highest level of variability. Recorded over several years and by different generations of equipment, the dataset covers subjects of all ages, adding to its complexity and rendering it the most challenging for seizure detection. The EEG signals in TUSZ are captured using 19 channels based on the standard EEG 1,020 system. Table 1 provides an overview of the TUSZ dataset utilized in our experiments.

During the training phase, we employed an equal number of normal clips as other supervised methods, omitting any seizure clips. In the testing phase, we utilized an equivalent number of test clips, encompassing both seizure and normal clips, for comparison against other supervised methods and our proposed approach. To assess the model's proficiency in seizure localization, we leveraged available annotations that specify focal and generalized seizure types from 23 distinct patients. It's noteworthy that, in epilepsy patients, focal and generalized seizure types are more prevalent compared to other seizure types, making them particularly relevant for our evaluation.

4.2 Baselines

We conducted a comprehensive evaluation of our proposed EEG-GCA method by comparing it with two distinct streams of deep learning-based approaches (12). The first stream involves well-established DL models operating in the EEG time-series and spectrograms domain, including EEGNet, EEG-TL, Dense-CNN, LSTM, and CNN-LSTM. The second stream focuses on DL models specifically designed for processing EEG graph data. Notably, our method differs from the others as it is

deliberately trained without utilizing any seizure data in the training phase, ensuring a fair comparison. In addition, we compared another method, EEG-CGS (12), a graph-based method, which utilizes the constructed EEG graph and self-supervised learning to capture local structural and contextual information embedded in EEG graphs and detects the anomaly by designed anomaly scores.

In this paper, we explore six variations of EEG-GCA based on different input graph types: EEG_d-GCA, EEG_r-GCA, EEG_c-GCA, EEG_f-GCA, EEG_t-GCA, and EEG_l-GCA. These variations utilize Dist-EEG-Graph, Rand-EEG-Graph, Corr-EEG-Graph, Full-EEG-Graph, DTF-EEG-Graph, and Identity-EEG-Graph as their respective inputs. All methods were evaluated on the same dataset, with the comparative analysis focusing on assessing the robustness and generalization capabilities of EEG-GCA, particularly in scenarios where seizure data is limited or unavailable. To evaluate the performance of the models, we used three metrics: Area Under the Curve (AUC), Average Precision (AP), and Specificity (SPC). These metrics provide insights into the models' ability to distinguish between different classes, their precision in detecting positive samples, and their ability to correctly identify negative samples, respectively.

4.3 Detection of seizure clips and channels

The performance of the seizure clip detection experiment across various comparison methods is shown in Table 2. Among the supervised methods, Corr-DCRNN exhibits the highest accuracy of 0.4482, suggesting that it effectively utilizes correlation information between different EEG channels. This is a crucial feature for seizure detection, as it allows the model to capture temporal dependencies and spatial relationships within the EEG signal. However, despite its relatively high accuracy, the model still struggles with achieving high specificity, which is essential for minimizing false positives in seizure detection.

In the unsupervised methods, EEG_r-CGS, based on random graphs, performs the best with an accuracy of 0.4285. This result indicates that even without the use of labeled data, the model is still able to leverage the underlying structure in the EEG data to some extent. However, the performance gap between EEG_r-CGS and supervised methods suggests that unsupervised learning still faces challenges in achieving comparable detection accuracy, particularly when it comes to fine-tuning the decision boundaries between seizure and non-seizure clips.

When comparing our proposed methods–EEG_d-GCA, EEG_r-GCA, EEG_c-GCA, EEG_f-GCA, EEG_t-GCA, and EEG_l-GCA–it is evident that the introduction of the Graph Correlation Attention (GCA) mechanism leads to significant improvements in both accuracy and specificity. The accuracy of our methods consistently outperforms both the supervised and unsupervised methods, with EEG_d-GCA achieving the highest accuracy at 0.6812. This result is particularly noteworthy considering that EEG_d-GCA utilizes the Dist-EEG-Graph as input, which focuses on capturing the structural relationships between different EEG channels. The combination

TABLE 2 Seizure clips detection result.

Method	Acc	Precision	Spec	Method	Acc	Precision	Spec
Supervised				Unsupervised			
EEGNet	0.4742	0.298	0.9021	EEG_d-CGS	0.3076	0.3076	0.9450
EEG-TL	0.4001	0.2675	NA	EEG_r-CGS	0.4285	0.3333	0.9291
Dense-CNN	0.4143	0.2746	0.8692	EEG_c-CGS	0.2857	0.2857	0.9132
LSTM	0.3652	0.2635	0.8143	EEG_f-CGS	0.2857	0.2857	0.9211
CNN-LSTM	0.3304	0.2572	0.8574	EEG_t-CGS	0.3076	0.3076	0.9009
Dist-DCRNN	0.3414	0.2612	0.9321	–	–	–	–
Corr-DCRNN	0.4482	0.2711	0.9003	–	–	–	–
Ours							
EEG_d-GCA	0.6812	0.3469	0.9714	EEG_r-GCA	0.6636	0.3438	0.9429
EEG_c-GCA	0.6847	0.3469	0.9714	EEG_f-GCA	0.6832	0.3469	0.9714
EEG_t-GCA	0.6848	0.3469	0.9714	EEG_l-GCA	0.6625	0.3438	0.9429

of attention mechanisms with graph-based representations allows the model to selectively focus on the most informative features, leading to a more robust and accurate detection of seizure clips.

Interestingly, while EEG_d-GCA achieves the highest accuracy, the other GCA variations (EEG_r-GCA, EEG_c-GCA, EEG_f-GCA, EEG_t-GCA, EEG_l-GCA) also show consistently high performance with accuracy values close to 0.6847. This suggests that the robustness of the GCA mechanism is not highly sensitive to the specific graph input type, which makes these methods versatile across different graph representations of the EEG data. The consistently high specificity of around 0.9714 across all EEG-GCA methods indicates their effectiveness in minimizing false positives, which is a critical factor in the practical application of seizure detection systems.

4.4 Detection of synthetic anomalous channels

In this section, we focus on evaluating the performance of the proposed method EEG-GCA in reliably detecting anomalous channels. To this end, we generate a synthetic test set using normal clips from the training phase. Specifically, we average every 35 normal clips without overlap and then introduce anomalies into the averaged clips with a 3% probability. The anomalies are injected with a 0.03% probability, and at most one node is corrupted per averaged clip. The corruptions are applied both structurally and contextually. The structural corruption involves connecting the selected node to all other nodes in the average clip, while the contextual corruption alters the attribute vector of the node by replacing its feature vector with that of the node in the clip that has the largest Euclidean distance. After introducing these anomalies, we input the averaged clips, some of which contain anomalies, into the EEG-GCA networks that were trained on pure normal clips.

TABLE 3 Synthetic anomalous channels detection results.

Type	Method	AUC	AP	Spec
Supervised	EEGNet	0.6182	0.298	0.902
	EEG-TL	0.5913	0.2675	NA
	Dense-CNN	0.5877	0.2746	0.869
	LSTM	0.5198	0.2635	0.814
	CNN-LSTM	0.5412	0.2572	0.857
	Dist-DCRNN	05683	0.2612	0.932
	Corr-DCRNN	0.6122	0.2711	0.900
Unsupervised	EEG_d-CGS	0.6182	0.0845	0.9455
	EEG_r-CGS	0.8173	0.2675	0.9555
	EEG_c-CGS	0.8241	0.2887	0.9555
	EEG_f-CGS	0.8143	0.2960	0.9555
	EEG_t-CGS	0.8241	0.2887	0.9555
Ours	EEG_d-GCA	0.8903	0.4193	0.9667
	EEG_r-GCA	0.9229	0.4618	0.9722
	EEG_c-GCA	0.916	0.402	0.97
	EEG_f-GCA	0.908	0.4325	0.9689
	EEG_t-GCA	0.9101	0.4172	0.9678
	EEG_l-GCA	0.9238	0.4792	0.9722

The trained system then computes the anomaly scores for all channels.

The experimental results, as summarized in Table 3, demonstrate the effectiveness of our approach in the domain of anomaly detection. Our method outperforms both supervised and other unsupervised learning techniques across key evaluation metrics such as AUC, AP, and Specificity. Specifically, EEGNet, a supervised learning method, achieves a moderate performance with an AUC of 0.6182. However, it faces challenges when

handling imbalanced datasets, which is a critical issue in real-world anomaly detection tasks. In contrast, EEG-GCA demonstrates remarkable improvements in AUC, with EEG_r-GCA and EEG_l-GCA achieving 0.9229 and 0.9238, respectively, highlighting the effectiveness of unsupervised learning techniques in addressing imbalances in the dataset. For AP, EEG-GCA surpasses the performance of the other methods. For instance, EEG_l-GCA reaches an AP of 0.4792, significantly outperforming the supervised approaches. This indicates that our method is highly capable of accurately identifying anomalous events, which is crucial in real-world anomaly detection tasks such as sentiment recognition. Notably, EEG-GCA also excels in terms of Specificity, a metric that measures the ability to correctly identify normal samples and minimize false positives. Both EEG_r-GCA and EEG_l-GCA achieve Specificity values of 0.9722, outperforming all supervised models. This is particularly important as it demonstrates that our method can maintain high sensitivity while effectively reducing false positives, thereby improving the robustness and reliability of anomaly detection.
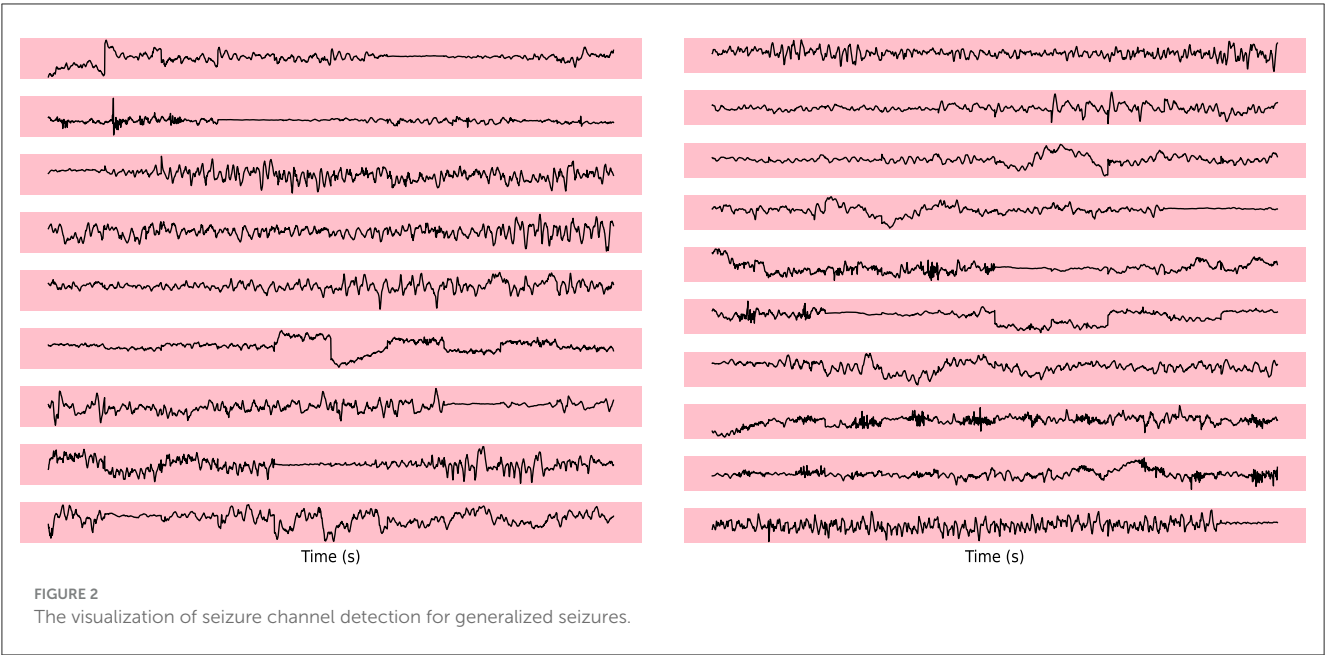
4.5 Ablation study

In the ablation study for seizure clip detection on synthetic anomalous channels, we explored two distinct approaches: Without Correlation and EEG-GCA. The results of this ablation analysis are summarized in Table 4.

In the Correlation approach, several graph construction methods were employed. Among these, EEG_c-GCA emerged as the top performer, achieving the highest AUC (0.8513) and AP (0.4227), underscoring its effectiveness in seizure detection. This result emphasizes the importance of incorporating correlation in the graph construction process for improving detection accuracy. Notably, EEG_r-GCA and EEG_t-GCA also displayed competitive results, highlighting their resilience to the absence of correlation while still maintaining reasonable performance. These findings suggest that, even without explicit correlation, the models are capable of leveraging other aspects of the data for meaningful detection.

4.6 Visualization of EEG signal

To evaluate the abnormal channels in the electroencephalogram (EEG) segments during epileptic seizures, we visualize the seizure channel for generalized seizures.

In Figure 2, which represents a case of generalized seizures, our method demonstrates a high level of accuracy in detecting all abnormal channels. This robust performance aligns with our expectations for identifying anomalies during generalized seizure events, highlighting the reliability of our approach in such scenarios. The elevated anomaly scores observed in the seizure-affected channels provide strong evidence of the discriminatory power of our model, successfully distinguishing pathological EEG patterns from normal, baseline activity. This

TABLE 4 Ablation study on seizure clips detection results.

Method	Without correlation			Ours		
	AUC	AP	Spec	AUC	AP	Spec
EEG_d-GCA	0.7579	0.3378	0.9644	0.8903	0.4193	0.9667
EEG_r-GCA	0.8377	0.3431	0.9622	0.9229	0.4618	0.9722
EEG_c-GCA	0.8513	0.4227	0.9678	0.9160	0.4020	0.9700
EEG_f-GCA	0.8628	0.4025	0.9612	0.9080	0.4325	0.9689
EEG_t-GCA	0.8355	0.3489	0.9533	0.9101	0.4172	0.9678
EEG_l-GCA	0.8331	0.3301	0.9622	0.9238	0.4792	0.9722

FIGURE 2
The visualization of seizure channel detection for generalized seizures.

underscores the potential of our approach for real-time, accurate seizure detection.

5 Conclusion

In this paper, we introduce EEG-GCA, an unsupervised graph-based model designed for EEG-based epilepsy detection. The core of the methodology is centered around computing the correlation between individual EEG channels and their neighboring channels. The process begins with the construction of a graph representation of the EEG data, which enables the exploration of correlation patterns across the channels. A weight-sharing Graph Convolutional Network is then employed to effectively capture both the semantic and structural relationships among the channels. By aligning these distributions with a prior distribution, EEG-GCA learns the underlying correlations within the EEG data. The final stage involves detecting anomalous channels based on the correlation scores, with weak correlation scores indicating potential anomalies that may signify seizures. The experimental results demonstrate that EEG-GCA outperforms existing methods, achieving superior accuracy in detecting anomalous channels. This underscores the effectiveness of leveraging graph-based correlation techniques for the detection of epilepsy in EEG signals. In the future, we exploration involves integrating multi-modal data, such as incorporating additional physiological signals or patient-specific features, to further enhance the robustness and adaptability of models.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

CT: Formal analysis, Investigation, Methodology, Software, Writing – original draft. FZ: Funding acquisition, Investigation, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Yamaguchi T, Aihara A, Mashiko S, Kurosawa E, Oizumi T, Yamagata T, et al. Exacerbation of delirium and epileptic seizures in an older man with idiopathic Parkinson's disease due to multiple prescriptions: a case report. *Front Med.* (2024) 11:1415988. doi: 10.3389/fmed.2024.1415988

2. Pradeep P, Kang H, Lee B. Glycosylation and behavioral symptoms in neurological disorders. *Transl Psychiatry.* (2023) 13:154. doi: 10.1038/s41398-023-02446-x

3. Sun Y, Wei C, Cui V, Xiu M, Wu A. Electroencephalography: clinical applications during the perioperative period. *Front Med.* (2020) 7:251. doi: 10.3389/fmed.2020.00251

4. Huang W, Yan G, Chang W, Zhang Y, Yuan Y. EEG-based classification combining Bayesian convolutional neural networks with recurrence plot for motor movement/imagery. *Pattern Recognit.* (2023) 144:109838. doi: 10.1016/j.patcog.2023.109838

5. Kim BH, Choi JW, Lee H, Jo S. A discriminative SPD feature learning approach on Riemannian manifolds for EEG classification. *Pattern Recognit.* (2023) 143:109751. doi: 10.1016/j.patcog.2023.109751

6. Behrouzi T, Hatzinakos D. Graph variational auto-encoder for deriving EEG-based graph embedding. *Pattern Recognit.* (2022) 121:108202. doi: 10.1016/j.patcog.2021.108202

7. Khurshid D, Wahid F, Ali S, Gumaei AH, Alzanin SM, Mosleh MA, et al. deep neural network-based approach for seizure activity recognition of epilepsy sufferers. *Front Med.* (2024) 11:1405848. doi: 10.3389/fmed.2024.1405848

8. Shoeibi A, Khodatars M, Ghassemi N, Jafari M, Moridian P, Alizadehsani R, et al. Epileptic seizures detection using deep learning techniques: a review. *Int J Environ Res Public Health.* (2021) 18:5780. doi: 10.3390/ijerph18115780

9. Wu M, Wan T, Wan X, Fang Z, Du Y. A new localization method for epileptic seizure onset zones based on time-frequency and clustering analysis. *Pattern Recognit.* (2021) 111:107687. doi: 10.1016/j.patcog.2020.107687

10. Thuwajit P, Rangpong P, Sawangjai P, Autthasan P, Chaisaen R, Banluesombatkul N, et al. EEGWaveNet: multiscale CNN-based spatiotemporal feature extraction for EEG seizure detection. *IEEE Trans Ind Inform.* (2021) 18:5547–57. doi: 10.1109/TII.2021.3133307

11. Wagh N, Varatharajah Y. EEG-GCNN: augmenting electroencephalogram-based neurological disease diagnosis using a domain-guided graph convolutional neural network. In: *Proceedings of the Machine Learning for Health NeurIPS Workshop, Vol. 136.* Vancouver, BC: PMLR (2020). p. 367–78.

12. Ho TKK, Armanfard N. Self-supervised learning for anomalous channel detection in EEG graphs: application to seizure analysis. *Proc AAAI Conf Artif Intell.* (2023) 37:7866–74. doi: 10.1609/aaai.v37i7.25952

13. Liu Y, Li Z, Pan S, Gong C, Zhou C, Karypis G. Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE Trans Neural Netw Learn Syst.* (2021) 33:2378–92. doi: 10.1109/TNNLS.2021.3068344

14. Kirschstein T, Köhling R. What is the source of the EEG? *Clin EEG Neurosci.* (2009) 40:146–9. doi: 10.1177/155005940904000305

15. Shi W, Rajkumar R. Point-GNN: graph neural network for 3D object detection in a point cloud. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA: IEEE (2020). p. 1711–9. doi: 10.1109/cvpr42600.2020.00178

16. Zhou K, Huang X, Song Q, Chen R, Hu X. Auto-GNN: neural architecture search of graph neural networks. *Front Big Data*. (2022) 5:1029307. doi: 10.3389/fdata.2022.1029307

17. Yang X, Yang X, Yang J, Ming Q, Wang W, Tian Q, et al. Learning high-precision bounding box for rotated object detection via Kullback-Leibler divergence. *Adv Neural Inf Process Syst*. (2021) 34:18381–94.

18. Bos DO. EEG-based emotion recognition. The influence of visual and auditory stimuli. Comput Sci. (2006) 56:1–17.

19. Suhaimi NS, Mountstephens J, Teo J. EEG-based emotion recognition: a state-of-the-art review of current trends and opportunities. *Comput Intell Neurosci*. (2020) 2020:8875426. doi: 10.1155/2020/8875426

20. Kale GV, Patil VH. A study of vision based human motion recognition and analysis. *Int J Ambient Comput Intell*. (2016) 7:75–92. doi: 10.4018/IJACI.2016070104

21. Da Gama AEF, de Menezes Chaves T, Fallavollita P, Figueiredo LS, Teichrieb V. Rehabilitation motion recognition based on the international biomechanical standards. *Expert Syst Appl*. (2019) 116:396–409. doi: 10.1016/j.eswa.2018.09.026

22. Hu Z, Lee EJ. Human motion recognition based on improved 3-dimensional convolutional neural network. In: *2019 IEEE International Conference on Computation, Communication and Engineering (ICCCE)*. Fujian: IEEE (2019), p. 154–6. doi: 10.1109/ICCCE48422.2019.9010816

23. Zhang F, Wu TY, Pan JS, Ding G, Li Z. Human motion recognition based on SVM in VR art media interaction environment. *Hum-centric Comput Inform Sci*. (2019) 9:1–15. doi: 10.1186/s13673-019-0203-8

24. Pyun KR, Kwon K, Yoo MJ, Kim KK, Gong D, Yeo WH, et al. Machine-learned wearable sensors for real-time hand motion recognition: toward practical applications in reality. *Natl Sci Rev*. (2023) 11:nwad298. doi: 10.1093/nsr/nwad298

25. Zhou H, Yang G, Wang B, Li X, Wang R, Huang X, et al. An attention-based deep learning approach for inertial motion recognition and estimation in human-robot collaboration. *J Manuf Syst*. (2023) 67:97–110. doi: 10.1016/j.jmsy.2023.01.007

26. Akinyemi TO, Omisore OM, Du W, Duan W, Chen X, Yi G, et al. Interventionalist hand motion recognition with convolutional neural network in robot-assisted coronary interventions. *IEEE Sens J*. (2023) 23:17725–36. doi: 10.1109/JSEN.2023.3281009

27. Li X, Liu Y, Zhou X, Yang Z, Tian L, Fang P, et al. Simultaneous hand/wrist motion recognition and continuous grasp force estimation based on nonlinear spectral sEMG features for transradial amputees. *Biomed Signal Process Control*. (2023) 85:105044. doi: 10.1016/j.bspc.2023.105044

28. Li X, Cao X. Human motion recognition information processing system based on LSTM recurrent neural network algorithm. *J Ambient Intell Humaniz Comput*. (2023) 14:8509–21. doi: 10.1007/s12652-021-03614-x

29. Abdullah AS, AlSaif KI. Computer vision system for backflip motion recognition in gymnastics based on deep learning. *J Al-Qadisiyah Comput Sci Math*. (2023) 15:150. doi: 10.29304/jqcm.2023.15.1.1162

30. Mocanu B, Tapu R, Zaharia T. Multimodal emotion recognition using cross modal audio-video fusion with attention and deep metric learning. *Image Vis Comput*. (2023) 133:104676. doi: 10.1016/j.imavis.2023.104676

31. Sun P, De Winne J, Zhang M, Devos P, Botteldooren D. Delayed knowledge transfer: cross-modal knowledge transfer from delayed stimulus to EEG for continuous attention detection based on spike-represented EEG signals. *Neural Netw*. (2025) 183:107003. doi: 10.1016/j.neunet.2024.107003

32. Wang J, Zhang C. Cross-modality fusion with EEG and text for enhanced emotion detection in English writing. *Front Neurorobot*. (2025) 18:1529880. doi: 10.3389/fnbot.2024.1529880

33. Xu M, Shi T, Zhang H, Liu Z, He X. A hierarchical cross-modal spatial fusion network for multimodal emotion recognition. *IEEE Trans Artif Intell*. (2025) :1–10. doi: 10.1109/TAI.2024.3523250

34. Andrew G, Arora R, Bilmes J, Livescu K. Deep canonical correlation analysis. In: *International Conference on Machine Learning*. PMLR (2013), p. 1247–55.

35. Chen Z, Liu C, Ding SX, Peng T, Yang C, Gui W, et al. A just-in-time-learning-aided canonical correlation analysis method for multimode process monitoring and fault detection. *IEEE Trans Ind Electron*. (2020) 68:5259–70. doi: 10.1109/TIE.2020.2989708

36. Chang X, Xiang T, Hospedales TM. Scalable and effective deep CCA via soft decorrelation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2018), p. 1488–97. doi: 10.1109/CVPR.2018.00161

37. Lai D, Wang S, Chong Z, Wu W, Nardini C. Task-oriented attributed network embedding by multi-view features. *Knowl-Based Syst*. (2021) 232:107448. doi: 10.1016/j.knosys.2021.107448

38. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. (2014), p. 701–10. doi: 10.1145/2623330.2623732

39. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. *arXiv*. (2013) [Preprint]. arXiv:1301.3781. doi: 10.48550/arXiv.1301.3781

40. Grover A, Leskovec J. node2vec: Scalable feature learning for networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM (2016), p. 855–64. doi: 10.1145/2939672.2939754

41. Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q. Line: large-scale information network embedding. In: : *Proceedings of the 24th International Conference on World Wide Web*. Geneva: International World Wide Web Conferences Steering Committee (2015), p. 1067–77. doi: 10.1145/2736277.2741093

42. Cao S, Lu W, Xu Q. Grarep: Learning graph representations with global structural information. In: *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. New York, NY: ACM (2015), p. 891–900. doi: 10.1145/2806416.2806512

43. Zheng C, Pan L, Wu P. Attribute augmented network embedding based on generative adversarial nets. *IEEE Trans Neural Netw Learn Syst*. (2021) 34:3473–3487. doi: 10.1109/TNNLS.2021.3116419

44. Fan H, Zhang F, Wei Y, Li Z, Zou C, Gao Y, et al. Heterogeneous hypergraph variational autoencoder for link prediction. *IEEE Trans Pattern Anal Mach Intell*. (2021) 44: 4125–138. doi: 10.1109/TPAMI.2021.3059313

45. Fan H, Zhang F, Li Z. AnomalyDAE: dual autoencoder for anomaly detection on attributed networks. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona: IEEE (2020), p. 5685–9. doi: 10.1109/ICASSP40776.2020.9053387

46. Wang H, Wang J, Wang J, Zhao M, Zhang W, Zhang F, et al. Graphgan: graph representation learning with generative adversarial nets. *Proc AAAI Conf Artif Intell*. (2018) 32:2508–15. doi: 10.1609/aaai.v32i1.11872

47. Dai Q, Li Q, Tang J, Wang D. Adversarial network embedding. *Proc AAAI Conf Artif Intell*. (2018) 32:2167–74. doi: 10.1609/aaai.v32i1.11865

48. Zhang F, Fan H, Wang R, Li Z, Liang T. Deep dual support vector data description for anomaly detection on attributed networks. *Int J Intell Syst*. (2021) 1:1. doi: 10.1002/int.22683

Early diagnosis of sepsis-associated AKI: based on destruction-replenishment contrast-enhanced ultrasonography

Zexing Yu, Xue Shi, Yang Song, Xin Li, Ling Li and Huiyu Ge*

Department of Ultrasound Medicine, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China

Objective: Establish a deep learning ultrasound radiomics model based on destruction-replenishment contrast-enhanced ultrasound (DR-CEUS) for the early prediction of acute kidney injury (SA-AKI).

Method: This paper proposes a deep learning ultrasound radiomics model (DLUR). Deep learning models were separately established using ResNet18, ResNet50, ResNext18, and ResNext50 networks. Based on the features extracted from the fully connected layers of the optimal model, a deep learning ultrasound radiomics model (DLUR) was established using three classification models (built with 3 classifiers). The predictive performance of the best DLUR model was compared with the visual assessments of two groups of ultrasound physicians with varying levels of experience. The performance of each model and the ultrasound physicians was evaluated by assessing the receiver operating characteristic (ROC) curves. The area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were subsequently calculated.

Results: Compared to the ResNet18 model, the DLUR model based on logistic regression (DLUR-LR) demonstrated the best predictive performance, showing a Net Reclassification Improvement (NRI) value of 0.210 ($p < 0.05$). The Integrated Discrimination Improvement (IDI) value for the corresponding stage was 0.169 ($p < 0.05$). Additionally, the performance of the DLUR-LR model also surpassed that of senior ultrasound physicians (AUC, 0.921 vs. 0.829, $p < 0.05$).

Conclusion: By combining deep learning and ultrasound radiomics, a deep learning ultrasound radiomics model with outstanding predictive efficiency and robustness has demonstrated excellent capability in the early prediction of acute kidney injury (SA-AKI).

KEYWORDS

destruction-replenishment contrast-enhanced ultrasound, deep learning ultrasound radiomics model, acute kidney injury, risk assessment, deep learning model

1 Introduction

According to the 2020 WHO statistics (1), there were 48.9 million cases of sepsis worldwide in 2017, resulting in 11 million deaths. Sepsis-related deaths accounted for 19.7% of all global deaths. The mortality rate of sepsis is 15–25%, and this rate increases to 30–50% in cases of septic shock. Therefore, sepsis represents a

significant public health issue worldwide due to its high incidence and mortality rates.

The kidneys are one of the organs most frequently affected by sepsis. Poston and Koyner (2) pointed out that up to 60% of sepsis patients develop secondary AKI, and the mortality rate significantly increases once sepsis is complicated by AKI. It is currently believed (3) that sepsis triggers macrocirculatory disturbances, leading to reduced renal blood flow (RBF), which causes acute tubular necrosis, thereby resulting in sepsis-associated AKI. As research into sepsis-related AKI deepens, studies (4, 5) have found that during septic shock, despite maintained or even increased RBF, the glomerular filtration rate (GFR) decreases, suggesting that the pathogenesis of sepsis-related AKI may be more complex. Hence, studying intrarenal blood perfusion has become a crucial step in understanding the pathophysiology of AKI during septic shock.

Currently, there are few methods available to assess and monitor renal cortical microcirculatory perfusion in sepsis patients (6, 7). Conventional color Doppler ultrasound, widely used for real-time monitoring of renal hemodynamics in large vessels and some small vessels in the renal parenchyma (8), lacks accuracy in evaluating microcirculatory perfusion, especially in the renal cortex. The recently developed contrast-enhanced ultrasound (CEUS) technology, which uses microbubble contrast agents (ultrasound contrast agents, UCA) much smaller than red blood cells, allows assessment of human microcirculatory perfusion by reaching any terminal small vessels via the pulmonary circulation.

Concurrently, the rapid advancements in deep learning and artificial intelligence have revolutionized medical image analysis, demonstrating exceptional capabilities in feature extraction and pattern recognition (9). Radiomics, which involves the extraction of a large number of quantitative features from medical images, combined with machine learning algorithms, has shown promise in achieving precise disease diagnosis and prognostic predictions (10). However, current radiomics research on SAKI predominantly focuses on modalities such as magnetic resonance imaging (MRI) and computed tomography (CT), with limited studies exploring deep learning-based ultrasound radiomics models for SAKI (11).

To the best of our knowledge, no study has yet confirmed the feasibility of using a DLUR model for the early prediction of acute kidney injury (SA-AKI). This study aims to establish a deep learning ultrasound radiomics model based on burst-replenishment contrast-enhanced ultrasound for the early prediction of acute kidney injury (SA-AKI).

2 Materials and methods

2.1 Study participants

The retrospective study collected data from 135 patients with sepsis at Beijing Chaoyang Hospital, Capital Medical University, from January 2023 to November 2024, including 75 SA-AKI patients and 60 SA-non-AKI patients. The inclusion criteria were: (1) meeting the diagnostic criteria of the "International Consensus on the Definition of Sepsis and Septic Shock, 3rd Edition"; (2) meeting the diagnostic criteria for acute kidney injury: Acute Kidney Injury (AKI) is defined when either of the following criteria is met: (1) Serum Creatinine Elevation Absolute increase in serum

creatinine ≥ 0.3 mg/dL (26.5 μmol/L) within 48 h, OR Serum creatinine rising to ≥ 1.5 times baseline value (i.e., $\geq 50\%$ increase from baseline) within 7 days. (2) Urine Output Reduction Sustained urine output <0.5 mL/kg/h persisting for ≥ 6 h; (3) age ≥ 18 years; (4) clear ultrasound images and complete clinical data. The exclusion criteria were: (1) patients with chronic kidney disease, renal transplantation, contraindications for SonoVue^{TM} contrast agents, or pulmonary hypertension; (2) incomplete clinical data; (3) poor quality of ultrasound images. All patients provided informed consent. Please refer to Figure 1 for detailed information. The data from 135 patients with sepsis were randomly divided into a training set ($n = 95$) and a testing set ($n = 45$) in an 7:3 ratio. Input data included burst-reperfusion ultrasound contrast agents and clinical data, while the output indicated whether the patient belonged to the septic AKI or non-AKI group.

2.2 Ultrasound data acquisition

Ultrasound diagnosis was performed by physicians with more than 5 years of relevant experience using the Mindray Resona R9 color Doppler ultrasound diagnostic system manufactured by the Chinese medical device company Mindray. The procedure utilized intravenous infusion combined with burst-replenishment contrast-enhanced ultrasound technology. The patient was placed in a supine position, and a vein in the left elbow was punctured to establish an intravenous access using a special vein tube for contrast-enhanced ultrasound. Two vials of SonoVue contrast agent were dissolved in 10 ml of saline, thoroughly shaken, and then placed in a 20 ml syringe. The syringe was installed in a specialized micro-infusion pump for contrast agents, with the speed set at 2 ml/min, and connected to the venous tube. The largest coronal section of the patient's right kidney (showing the renal hilum) was selected for observation. The injection pump was activated, and the ultrasound was used to observe the time it took for the contrast to reach the kidney, followed by continuous observation for 2 min until the contrast entered the kidney parenchyma and reached equilibrium. A fixed high mechanical index (MI > 0.7) was used to continuously burst the microbubbles within the kidney parenchyma for 6 s until all the contrast microbubbles were extinguished. Subsequently, the ultrasound probe was placed at the largest coronal section of the right kidney to continuously and dynamically observe the replenishment phase when the microbubbles re-entered the kidney for 30 s. This burst-replenishment process was repeated three times to obtain three sets of dynamic replenishment images, which were then subjected to time-intensity curve (TIC) analysis to acquire the replenishment data.

2.3 Ultrasound image annotation

After anonymizing patient information, the original dynamic ultrasound images were imported into the MedAI Darwin learning platform. The patient information labels were defined as follows: gender, age, body mass index, mean arterial pressure, arterial carbon dioxide partial pressure, hemoglobin, white blood cell

FIGURE 1
Flow diagram of the study population.
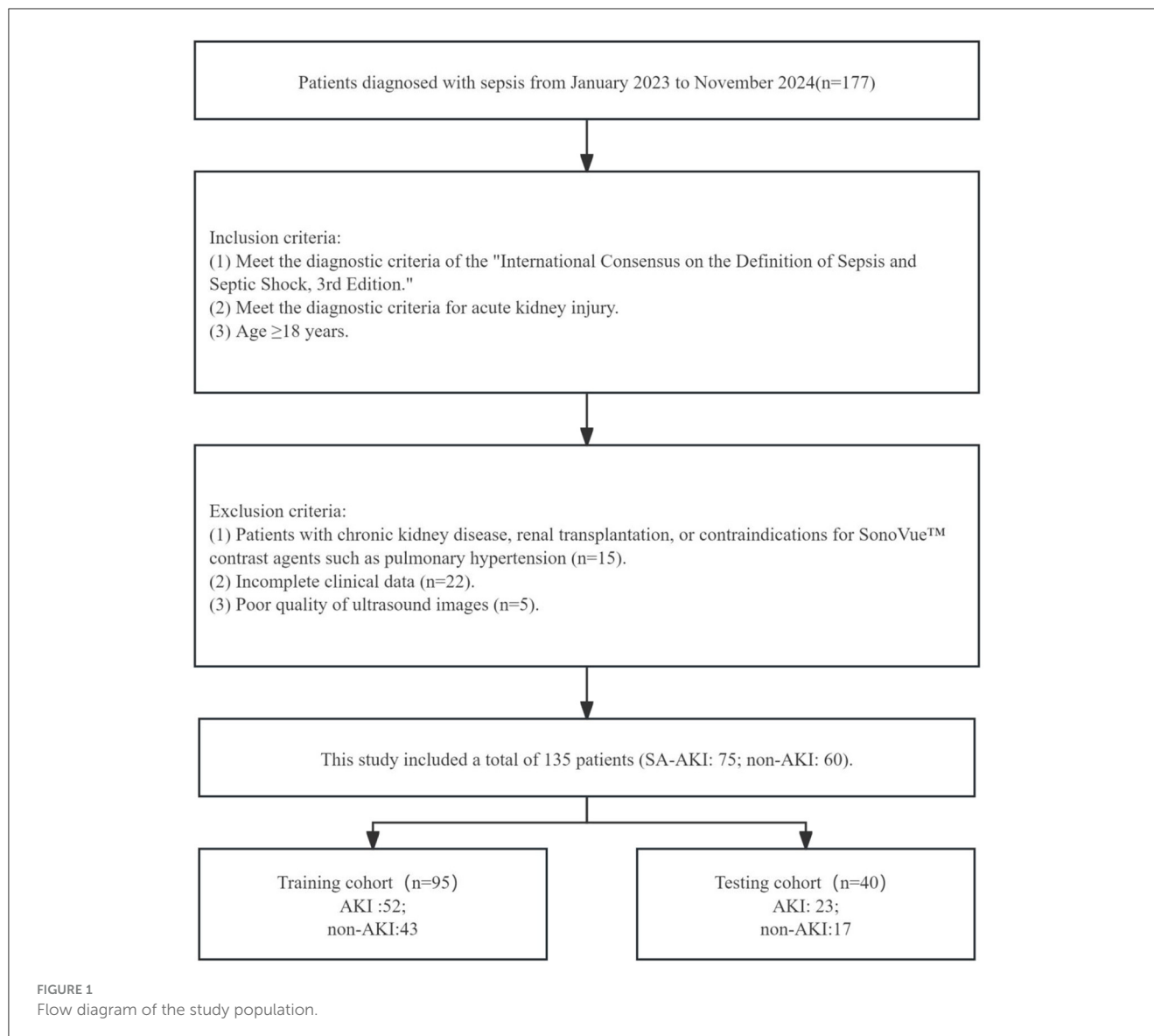
count, lactate, and serum creatinine. The lesion information label included: renal function impairment (septic AKI vs. non-AKI). Physicians with over 5 years of relevant experience manually delineated the regions of interest (ROI). In case of discrepancies, consultation with senior physicians (physicians with ≥10 years of ultrasound diagnostic experience and the title of **Associate Chief Physician** or higher) was sought for a definitive diagnosis.

2.4 Deep learning ultrasound radiomics model development

To ensure the integrity and validity of the research data, we have undertaken data preprocessing, aiming to enhance the performance and robustness of the models. The data preprocessing steps encompass data augmentation and image normalization. Considering the unique structural characteristics of the training data and the objectives of the task, we chose to build deep

learning models based on four different algorithms: ResNet18, ResNet50, ResNeXt18, and ResNeXt50. ResNet (Residual Network) and ResNeXt (Residual NeXt) are highly acclaimed deep learning models in the field of image recognition. They utilize the concept of residual learning, which enables the development of deeper networks without being hindered by issues of vanishing or exploding gradients.

ResNet addresses the issue of degradation in deep convolutional neural networks by introducing residual blocks. In these blocks, the input feature maps are combined with the subsequent layers through skip connections, allowing for the maximum preservation of the original information. Such design enables the residual blocks to learn the residual function, capturing the difference between the feature maps and the desired output. ResNeXt, an improvement upon ResNet, introduces grouped convolution within each residual block to enhance the model's expressive power. Traditional convolutional operations convolve each channel of the input feature maps with each filter, whereas grouped convolution divides the input feature maps

into multiple groups and independently convolves each group. By increasing the number of groups, ResNeXt enhances the model's expressive power without increasing the total number of parameters or computational complexity. Typical structures of ResNet and ResNeXt consist of multiple residual blocks, with variants such as ResNet-18, ResNeXt-18, ResNet-50, and ResNeXt-50 being widely used. Both ResNet and ResNeXt are composed of several residual blocks. Within each residual block, the convolutional layers are no longer ordinary convolutions but rather grouped convolutions, which divide the input feature maps into multiple groups for independent convolutional operations. The number of convolutional kernels within each group in grouped convolution is equal, and the quantity of groups is referred to as "cardinality." By increasing the cardinality, the model's non-linear expressive power can be enhanced. For instance, ResNet-18 is a relatively shallow ResNet model with approximately 11 million total parameters, while ResNet-50 is a deeper and more complex ResNet model with approximately 23 million total parameters.

In general, ResNet and ResNeXt exhibit slight differences in their model structures, but both leverage the concept of residual learning to address challenges in deep networks. These models have demonstrated outstanding performance in image recognition tasks and have become pivotal models in research and applications. The predictive performance of each model is evaluated using receiver operating characteristic (ROC) curves, and metrics such as area under the curve (AUC), sensitivity, specificity, and accuracy are calculated to select the best differentiating model for tuberculous hydronephrosis and non-tuberculous hydronephrosis. To optimize computational resources and improve training efficiency, this study uniformly employs region of interest (ROI) sub-images for model training, with the ROI sub-image size standardized to $64 \times 64 \times 64$ prior to training. Additionally, 3D image augmentation techniques, such as random flipping and random cropping, are applied to the training data.

After evaluating the deep learning modeling experiments, features were extracted from the fully connected layers of the best-performing deep learning model. These deep learning features were

FIGURE 2
The complete research process.
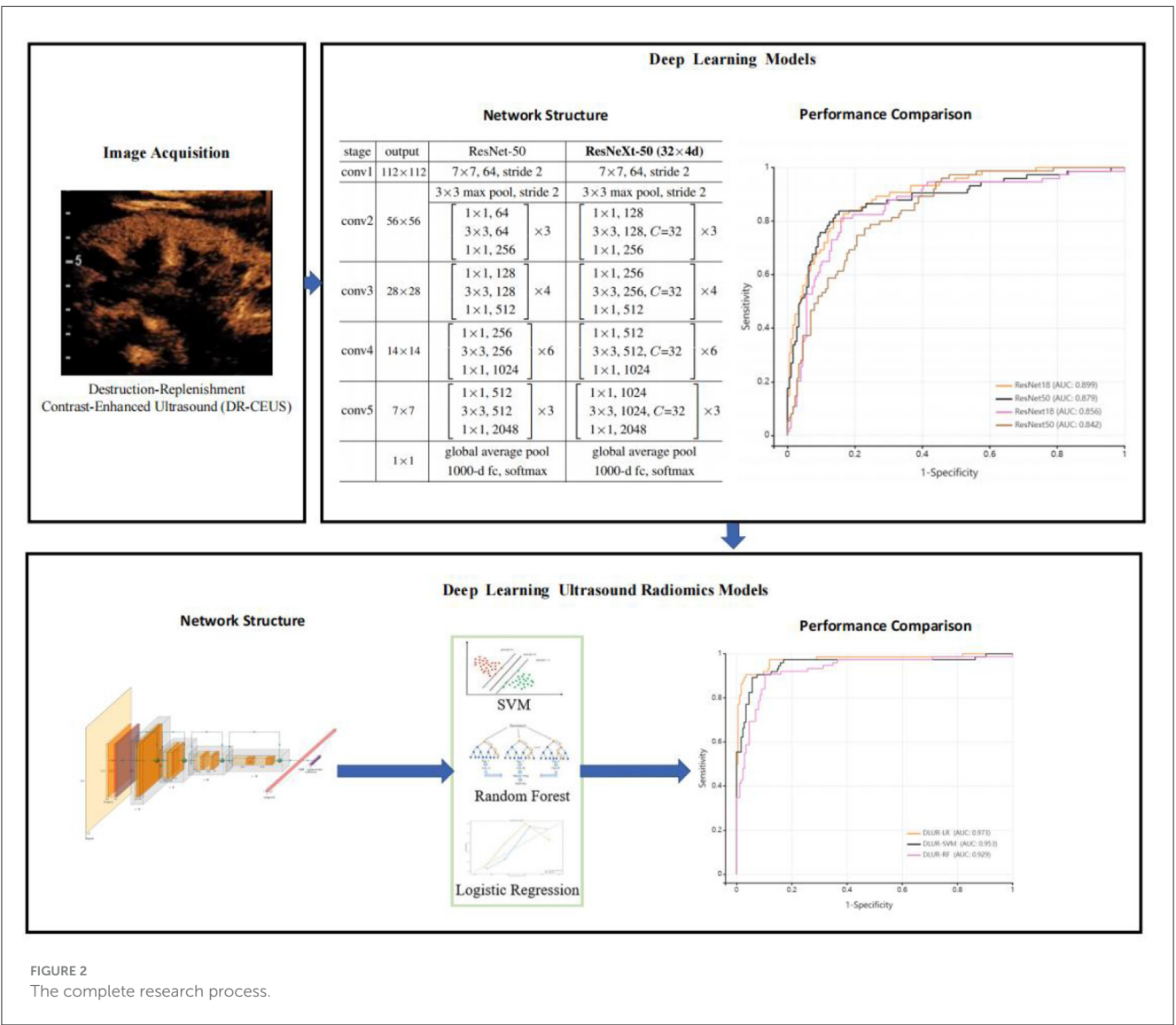
then used to build an ultrasound radiomics model using three mainstream machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). The predictive performance of each model was assessed using receiver operating characteristic (ROC) curves, and metrics such as the area under the curve (AUC), sensitivity, specificity, PPV, NPV, and accuracy were calculated. The complete research process is shown in Figure 2.

2.5 Statistical analysis

SPSS version 27.0 statistical analysis software was used to analyze the significance of each model. Categorical data were presented as actual frequencies and percentages. The classification performance of the models was assessed using the AUC, accuracy, sensitivity, specificity, PPV, and NPV derived from the receiver operating characteristic (ROC) curves. The DeLong test was used to compare the significance of the AUCs among the different models. A P-value of <0.05 was considered statistically significant, indicating a difference with practical importance.

3 Results

3.1 General clinical data

This study included 135 septic patients, who were divided into AKI group ($n = 75$) and non-AKI group ($n = 60$) based on renal function within 48 h and urine output within 24 h. There were 90 males and 45 females, with an average age of 65.3 ± 15.2 years. General clinical data are presented in Table 1. There were no statistically significant differences in age, sex, body mass index, MAP, PaCO2, and CRP between the two groups ($P > 0.05$). The

levels of Scr and Lac in the AKI group were significantly higher than those in the non-AKI group, and the differences were statistically significant ($P < 0.05$).The general clinical data of enrolled patients are shown in Table 1.

3.2 Performance of the deep learning ultrasound radiomics model

Table 2 lists four algorithm models based on deep learning technology. Compared with other deep learning models on the testing dataset, ResNet 18 exhibited superior overall performance. The AUC of ResNet 18 was 0.899 (95% CI: 0.858–0.940), with a sensitivity of 0.800, specificity of 0.857, PPV of 0.706, NPV of 0.909, and accuracy of 0.840. Comparison of performance among different deep learning models as shown in Figure 3.

Ultimately, among the three classifiers, the deep learning ultrasound radiomics model based on logistic regression demonstrated the best classification performance (see Table 3). In the testing set, the AUC of DLUR-LR was 0.973 (95% CI: 0.949–0.998), with a sensitivity of 0.905, specificity of 0.960, PPV of 0.905, NPV of 0.960, and accuracy of 0.944; the AUC of DLUR-SVM was 0.953 (95% CI: 0.918–0.988), with a sensitivity of 0.892, specificity of 0.938, PPV of 0.857, NPV of 0.954, and accuracy of 0.924; the AUC of DLUR-RF was 0.929 (95% CI: 0.890–0.968), with a sensitivity of 0.907, specificity of 0.891, PPV of 0.782, NPV of 0.957, and accuracy of 0.896. Comparison of performance among different deep learning ultrasound radiomics models as shown in Figure 4.

In our study, we utilized performance metrics such as AUC, sensitivity, specificity, PPV, and NPV due to their significant clinical importance in the context of early AKI diagnosis. These metrics were carefully chosen to align with and reflect the critical aspects of clinical outcomes. The AUC provides a comprehensive assessment of the model's overall ability to distinguish between AKI and non-AKI cases across all thresholds, offering a holistic evaluation of performance. Sensitivity and specificity are directly related to clinical priorities: high sensitivity minimizes missed true cases, ensuring timely treatment, while high specificity reduces false positives, avoiding unnecessary interventions. PPV and NPV further aid clinical decision-making by indicating the likelihood that test results accurately reflect the patient's condition, thereby supporting clinicians in making informed treatment choices. Together, these metrics not only validate the statistical performance of our model but also underscore its practical utility in improving patient outcomes by facilitating early and accurate diagnosis of sepsis-associated AKI.

4 Discussion

Sepsis-associated acute kidney injury (AKI) is a significant complication that complicates the management of septic patients and dramatically increases morbidity and mortality rates. The early diagnosis of sepsis-associated AKI is crucial for implementing timely therapeutic strategies, which can significantly improve patient outcomes. In this study, we explored the utility of destruction-replenishment contrast-enhanced ultrasonography

TABLE 1 The general clinical data of enrolled patients.

	SA-Non-AKI ($n = 60$)	SA-AKI ($n = 75$)	Significance (p)
Age	69 (17–90)	63 (54–77)	
Gender (percentage of females)	36%	27%	
Temperature	36.8 ± 0.87	36.8 ± 0.29	0.336
Pulse	82.7 ± 13.2	89.1 ± 21	0.49
Respiration (breaths per minute)	21.4 ± 4.8	18.7 ± 4.7	0.302
BMI	23.8 ± 2.98	22.9 ± 4.9	1.000
Mean arterial pressure (MAP)	72.0 ± 19.0	85.1 ± 9.7	0.193
Arterial partial pressure of carbon dioxide (PaCO2)	54.5 ± 27.5	47 ± 15.8	0.530
Hemoglobin	108.1 ± 24.3	90 ± 22	0.964
White blood cell count	11.7 ± 6.7	15.4 ± 13.3	0.151
Lactic acid	1.26 ± 0.26	2.97 ± 3.15	0.007*
ScR	60.6 ± 19.9	216.9 ± 149.5	0.025*

*Indicates that $P < 0.05$. There is a statistically significant difference between the two groups.

TABLE 2 The performance comparison of different deep learning models.

Model	AUC (95%CI)	Sensitivity	Specificity	PPV	NPV	Accuracy
ResNet18	0.899 [0.858–0.940]	0.800	0.857	0.706	0.909	0.840
ResNet50	0.879 [0.828–0.931]	0.838	0.847	0.697	0.925	0.844
ResNext18	0.856 [0.802–0.910]	0.811	0.841	0.682	0.914	0.832
ResNext50	0.842 [0.793–0.892]	0.773	0.766	0.586	0.887	0.768

FIGURE 3
Comparison of performance among different deep learning models.
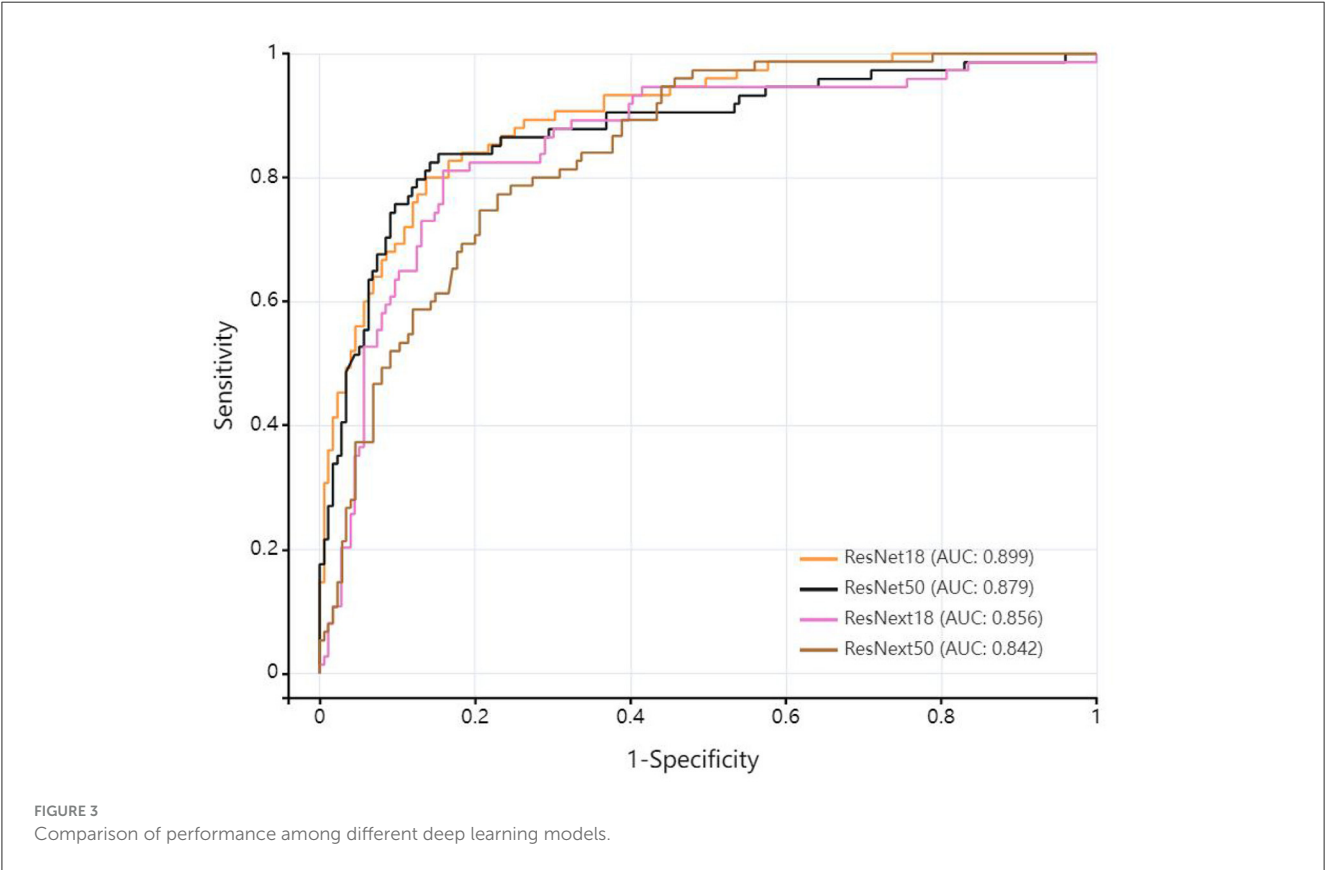
(DR-CEUS) as a novel method for the early detection of sepsis-associated AKI.

Our results indicate that DR-CEUS can detect early renal changes associated with sepsis before traditional markers show significant alterations. This is particularly important in the context of sepsis, where timely intervention is necessary to mitigate kidney injury. The ability to identify renal microcirculatory dysfunction may allow clinicians to initiate protective strategies earlier in the disease course, potentially reversing or preventing AKI progression.

We developed a deep learning ultrasound radiomics model that outperforms four different deep learning network models, namely ResNet18, ResNet50, ResNext18, and ResNext50. Compared to the best-performing model within ResNet18, our deep learning ultrasound radiomics model demonstrated superior predictive performance on the explosive-replenishment contrast-enhanced ultrasound imaging test data. The deep learning informatics model exhibited higher reliability and reproducibility in evaluating diagnostic outcomes, leveraging its inherent characteristics.

The performance differences among the various deep learning network models may be attributed to their distinct network architectures (12). In our study, we trained four different deep learning network architectures: ResNet18, ResNext18, ResNet50, and ResNext50, all of which are widely used in various clinical applications. We chose ResNet18, which exhibited the best predictive performance in our study, to extract deep learning features for constructing the deep learning ultrasound radiomics model. Among these four models, the ResNet network demonstrated more stable and superior predictive performance compared to other classical deep learning networks in the test set. The ResNet architecture maintains the integrity of information by directly passing input information to the output to learn the residual functions throughout the network. This property helps mitigate the issues of gradient vanishing and explosion, allowing the network to deepen without compromising performance (13). The ResNext network is a new architecture based on ResNet that incorporates the recurrent layer strategy of ResNet and combines it with a split-transform-merge strategy in a simple and scalable manner (14). However, the predictive results of the ResNext network were inferior to those of the ResNet network. In deep learning, dimensionality reduction, classification, and feature extraction are performed in an integrated manner. However, the quality and output of these cascading layers depend on

TABLE 3 The performance comparison of different deep learning ultrasound radiomics models.

Model	AUC (95%CI)	Sensitivity	Specificity	PPV	NPV	Accuracy
DLUR-LR	0.973 [0.949–0.998]	0.905	0.960	0.905	0.960	0.944
DLUR-SVM	0.953 [0.918–0.988]	0.892	0.938	0.857	0.954	0.924
DLUR-RF	0.929 [0.890–0.968]	0.907	0.891	0.782	0.957	0.896

FIGURE 4
Comparison of performance among different deep learning ultrasound radiomics models.
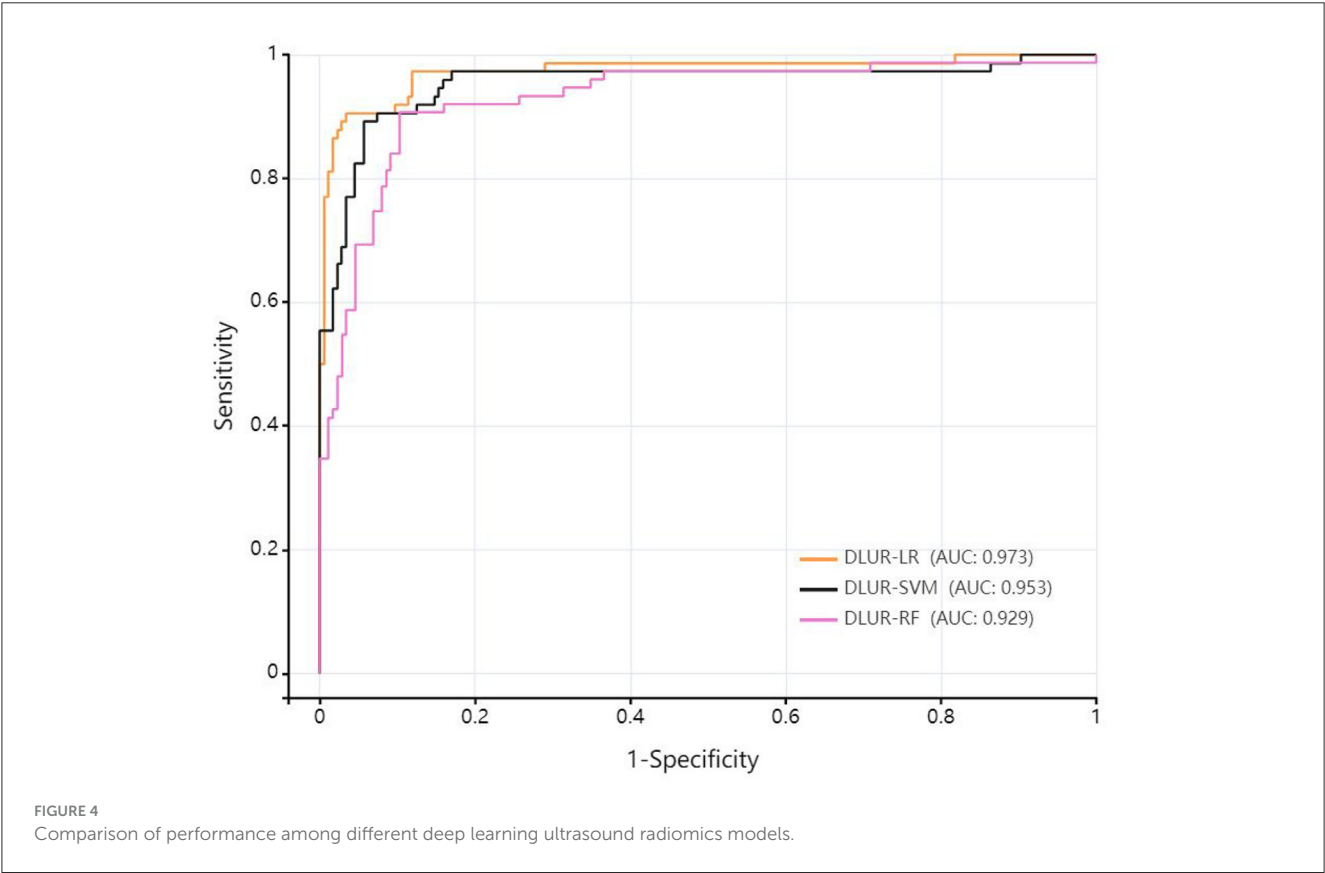
various hyperparameters such as the number of layers, feature maps, layer configurations, and structures. Different network architectures utilize different sets of hyperparameters, and the choice of these hyperparameters and architectures may impact predictive performance.

Our study reveals that the deep learning ultrasound radiomics model significantly outperforms traditional diagnostic methods and physician assessments in diagnosing sepsis-associated AKI. This advancement holds promise for improving early diagnosis in clinical settings. The integration of such AI-powered tools is increasingly feasible due to advancements in digital healthcare infrastructure, and our model can be seamlessly incorporated into existing ultrasound practices. However, potential barriers include initial investment costs, resistance to workflow changes, a need for comprehensive training, regulatory hurdles, and data privacy concerns. Despite these challenges, the clinical impact of implementing this model is substantial, offering more accurate diagnoses, timely interventions, and improved patient outcomes, while also alleviating physicians' cognitive load. Expanding our discussion to include these integration considerations and clinical benefits highlights the model's potential to enhance real-world healthcare delivery.

5 Conclusion

In this study, we propose a deep learning ultrasound imaging model based on blast-reperfusion ultrasound contrast imaging. Our method effectively integrates the technical advantages of deep learning and ultrasound imagingomics, demonstrating excellent predictive performance for the early diagnosis of sepsis-related acute kidney injury (AKI). This enables clinicians to detect renal changes earlier than traditional methods, allowing for the use of more precise interventions.

Data availability statement

The datasets presented in this article are not readily available because the dataset is subject to several restrictions, including limited access, which may require specific permissions or credentials for use. Usage of the dataset is prohibited for commercial purposes without prior authorization, and users must comply with data privacy regulations governing any personally identifiable information (PII) it contains. Additionally, proper attribution is required

when the data is cited or referenced in research, and modifications to the dataset are not allowed unless explicitly permitted. Users must carefully review these conditions to ensure compliance with all applicable legal and ethical guidelines. Requests to access the datasets should be directed to yu_zexing@126.com.

Ethics statement

The studies involving humans were approved by Medical Ethics Committee of Beijing Chaoyang Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

ZY: Conceptualization, Methodology, Writing – original draft. XS: Conceptualization, Data curation, Writing – original draft. YS: Investigation, Validation, Writing – original draft. XL: Data curation, Formal analysis, Writing – original draft. LL: Resources, Writing – original draft. HG: Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewers YZ and XM declared a shared parent affiliation with the authors to the handling editor at the time of review.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. World Health Organization. *Global Report on the Epidemiology and Burden of Sepsis: Current Evidence, Identifying Gaps and Future Directions ISBN 978-92-4-001078-9 (electronic version) ISBN 978-92-4-001079-6 (print version)* (2020). Geneva: World Health Organization.

2. Poston JT, Koyner JL. Sepsis associated acute kidney injury. *BMJ.* (2019) 364:k4891. doi: 10.1136/bmj.k4891

3. Peerapornratana S, Manrique-Caballero CL, Gómez H, Kellum JA. Acute kidney injury from sepsis: current concepts, epidemiology, pathophysiology, prevention and treatment. *Kidney Int.* (2019) 96:1083–99. doi: 10.1016/j.kint.2019.05.026

4. Maiden MJ, Otto S, Brealey JK, Finnis ME, Chapman MJ, Kuchel TR, et al. Structure and function of the kidney in septic shock. A prospective controlled experimental study. *Am J Respir Crit Care Med.* (2016) 194:692–700. doi: 10.1164/rccm.201511-2285OC

5. Burban M, Hamel JF, Tabka M, de La Bourdonnaye MR, Duveau A, Mercat A, et al. Renal macro- and microcirculation autoregulatory capacity during early sepsis and norepinephrine infusion in rats. *Crit Care.* (2013) 17:R139. doi: 10.1186/cc12818

6. Libert N, Harrois A, Duranteau J. Haemodynamic coherence in haemorrhagic shock. *Best Pract Res Clin Anaesthesiol.* (2016) 30:429–35. doi: 10.1016/j.bpa.2016.11.002

7. Gomez H, Ince C, De Backer D, Pickkers P, Payen D, Hotchkiss J, et al. A unified theory of sepsis-induced acute kidney injury: inflammation, microcirculatory dysfunction, bioenergetics, and the tubular cell adaptation to injury. *Shock.* (2014) 41:3–11. doi: 10.1097/SHK.0000000000000052

8. Le Dorze M, Bouglé A, Deruddre S, Duranteau J. Renal Doppler ultrasound: a new tool to assess renal perfusion in critical illness. *Shock.* (2012) 37:360–5. doi: 10.1097/SHK.0b013e3182467156

9. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature.* (2019) 572:116–9. doi: 10.1038/s41586-019-1390-1

10. Aerts HJWL, Velazquez ER, Leijenaar RTG, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5006

11. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med.* (2018) 46:1070–7. doi: 10.1097/CCM.0000000000003123

12. Fujima N, Andreu-Arasa VC, Onoue K, Weber PC, Hubbell RD, Setty BN, et al. Utility of deep learning for the diagnosis of otosclerosis on temporal bone CT. *Eur Radiol.* (2021) 31:5206–11. doi: 10.1007/s00330-020-07568-0

13. He K, Zhang X, Ren S, Sun J. *Deep Residual Learning for Image Recognition.* USA: IEEE (2016). p. 770–8.

14. Xie S, Girshick RB, Dollár P, Tu Z, He K. *Aggregated Residual Transformations for Deep Neural Networks.* USA: IEEE (2017). p. 5987–95.

Check for updates

Preliminary exploratory study on differential diagnosis between benign and malignant peripheral lung tumors: based on deep learning networks

Yuan Wang[1,2†], Yutong Zhang[1†], Yongxin Li[3], Tianyu She[4], Meiqing He[5], Hailing He[6], Dong Zhang[1,7]* and Jue Jiang[1]*

[1]Department of Ultrasound, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2]Department of Ultrasound, Yaozhou District People's Hospital, Tongchuan, China, [3]School of Automation and Intelligence, Beijing Jiaotong University, Beijing, China, [4]Department of Ultrasound, Xi'an Electric Power Central Hospital, Xi'an, China, [5]Department of Ultrasound, Shaanxi Provincial People's Hospital, Xi'an, China, [6]Department of Ultrasound, Tongchuan Mining Bureau Central Hospital, Tongchuan, China, [7]Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China

Background: Deep learning has shown considerable promise in the differential diagnosis of lung lesions. However, the majority of previous studies have focused primarily on X-ray, computed tomography (CT), and magnetic resonance imaging (MRI), with relatively few investigations exploring the predictive value of ultrasound imaging.

Objective: This study aims to develop a deep learning model based on ultrasound imaging to differentiate between benign and malignant peripheral lung tumors.

Methods: A retrospective analysis was conducted on a cohort of 371 patients who underwent ultrasound-guided percutaneous lung tumor procedures across two centers. The dataset was divided into a training set ($n = 296$) and a test set ($n = 75$) in an 8:2 ratio for further analysis and model evaluation. Five distinct deep learning models were developed using ResNet152, ResNet101, ResNet50, ResNet34, and ResNet18 algorithms. Receiver Operating Characteristic (ROC) curves were generated, and the Area Under the Curve (AUC) was calculated to assess the diagnostic performance of each model. DeLong's test was employed to compare the differences between the groups.

Results: Among the five models, the one based on the ResNet18 algorithm demonstrated the highest performance. It exhibited statistically significant advantages in predictive accuracy ($p < 0.05$) compared to the models based on ResNet152, ResNet101, ResNet50, and ResNet34 algorithms. Specifically, the ResNet18 model showed superior discriminatory power. Quantitative evaluation through Net Reclassification Improvement (NRI) analysis revealed that the NRI values for the ResNet18 model, when compared with ResNet152, ResNet101, ResNet50, and ResNet34, were 0.180, 0.240, 0.186, and 0.221, respectively. All corresponding p-values were less than 0.05 ($p < 0.05$ for each comparison), further confirming that the ResNet18 model significantly outperformed the other four models in reclassification ability. Moreover, its predictive outcomes led to marked improvements in risk stratification and classification accuracy.

Conclusion: The ResNet18-based deep learning model demonstrated superior accuracy in distinguishing between benign and malignant peripheral lung tumors, providing an effective and non-invasive tool for the early detection of lung cancer.

Introduction

Lung cancer remains one of the most prevalent and fatal cancers worldwide, with peripheral lung cancer (PLC) constituting a substantial proportion of these cases (1). PLC originates in the outer regions of the lungs and is often difficult to detect in its early stages due to subtle symptoms, leading to a high rate of misdiagnosis (2). Epidemiological studies indicate that peripheral lung tumors account for approximately 30–40% of all lung cancer diagnoses, underscoring the urgent need for effective early detection and accurate diagnosis (1, 3). Current clinical practices rely on low-dose computed tomography (LDCT) as the gold standard for lung cancer screening (4). However, while LDCT is highly effective, it involves the use of ionizing radiation, making it unsuitable for long-term monitoring, especially in individuals at high risk for lung cancer (5). Tissue biopsy, although definitive, is invasive and associated with potential complications, including bleeding and infection (2, 6). As a result, there is a growing interest in alternative, non-invasive diagnostic methods.

Ultrasound imaging has emerged as a promising non-invasive, radiation-free diagnostic tool for peripheral lung tumors, offering the advantage of high repeatability. This makes it particularly useful for monitoring patients over time and distinguishing between benign and malignant tumors (2, 5, 7). The application of ultrasound in lung tumor diagnosis has seen significant advancements in recent years. Recent studies have highlighted the improved accuracy of ultrasound techniques with the incorporation of elastography, which assesses tissue stiffness and provides valuable insights into tumor characterization (8). Additionally, the use of contrast-enhanced ultrasound (CEUS) has allowed for enhanced visualization of blood flow within tumors, further improving the ability to differentiate malignant from benign lesions (9). These innovations have made ultrasound a more reliable option for lung tumor diagnosis, particularly in settings where access to advanced imaging technologies such as CT or MRI may be limited.

However, despite these advancements, the accuracy of ultrasound diagnosis remains subject to operator-dependent variability, including factors such as experience, skill, and visual fatigue, which can lead to misjudgments (2, 5). To address these challenges, artificial intelligence (AI) techniques, particularly deep learning (DL) models, have been integrated into ultrasound imaging to improve diagnostic accuracy and consistency. Recent developments in AI have demonstrated substantial improvements in the automated analysis of ultrasound images, enabling more precise and reliable detection of lung tumors (7). Deep learning algorithms, especially convolutional neural networks (CNNs) like ResNet, have been shown to outperform traditional machine learning models by automatically detecting complex patterns and analyzing texture features that are often imperceptible to the human eye (7, 10). The ResNet architecture, known for its residual learning framework, helps mitigate the vanishing gradient problem and allows for the training of deeper neural networks, thus improving the robustness and accuracy of tumor detection (10, 11).

Therefore, incorporating AI into ultrasound imaging for PLC diagnosis has significantly reduced misdiagnosis rates and improved diagnostic confidence. AI-enhanced systems also provide real-time feedback, minimizing the effects of operator fatigue and variability, which are common limitations of traditional visual inspection (7). This study aims to develop and assess five deep learning models utilizing ultrasound images and clinical data of peripheral lung tumors. We hypothesize that these models will offer a highly accurate, non-invasive approach to differentiating benign from malignant tumors, thereby improving lung cancer screening and early diagnosis. The novelty of this research lies in the integration of ultrasound imaging with deep learning algorithms, addressing the limitations of conventional diagnostic methods and enhancing both diagnostic precision and clinical applicability.

Materials and methods

Study population

The study received approval from the institutional review board of The Second Affiliated Hospital of Xi'an Jiaotong University and Tongchuan mining bureau central hospital, which was conducted in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. This retrospective study collected data from 513 patients with peripheral lung tumors detected via chest CT across two centers between March 2020 and March 2024. The cohort included 438 patients from Center 1 and 75 patients from Center 2, respectively. A total of 371 lung tumors, comprising 221 malignant and 150 benign cases, were included for further analysis. The inclusion criteria were as follows: (1) adult patients aged 18 years or older; (2) peripheral lung tumors detected through chest CT imaging; (3) patients who underwent ultrasound-guided tissue biopsy; (4) pathological diagnosis confirming malignant lung tumors or inflammatory lesions; (5) clear ultrasound images of adequate quality were defined by two key criteria: a minimum resolution of 1.5 millimeters and a signal-to-noise ratio (SNR) threshold of 30 dB. These standards were established to ensure sufficient image clarity and diagnostic reliability for accurate feature extraction and tumor analysis; (6) complete clinical data. The exclusion criteria included: (1) patients with severe coagulation disorders or those unable to cooperate with ultrasound-guided tissue biopsy; (2) incomplete clinical data; (3) poor quality ultrasound images that could not provide valid data (Figure 1). All patients provided informed consent, and the key contents of the informed consent form are presented in the Supplementary materials.

Ultrasound data acquisition

Ultrasound diagnoses were performed by physicians with over 5 years of relevant experience using the Acuson Sequoia color Doppler ultrasound diagnostic system (Siemens AG, Germany), equipped with a 5C1 abdominal probe operating within a frequency range of 1.0 to 5.7 MHz. Based on lesion locations identified by CT scans, patients were positioned in supine, prone, or lateral decubitus positions to facilitate

FIGURE 1
Flow diagram of the study population.
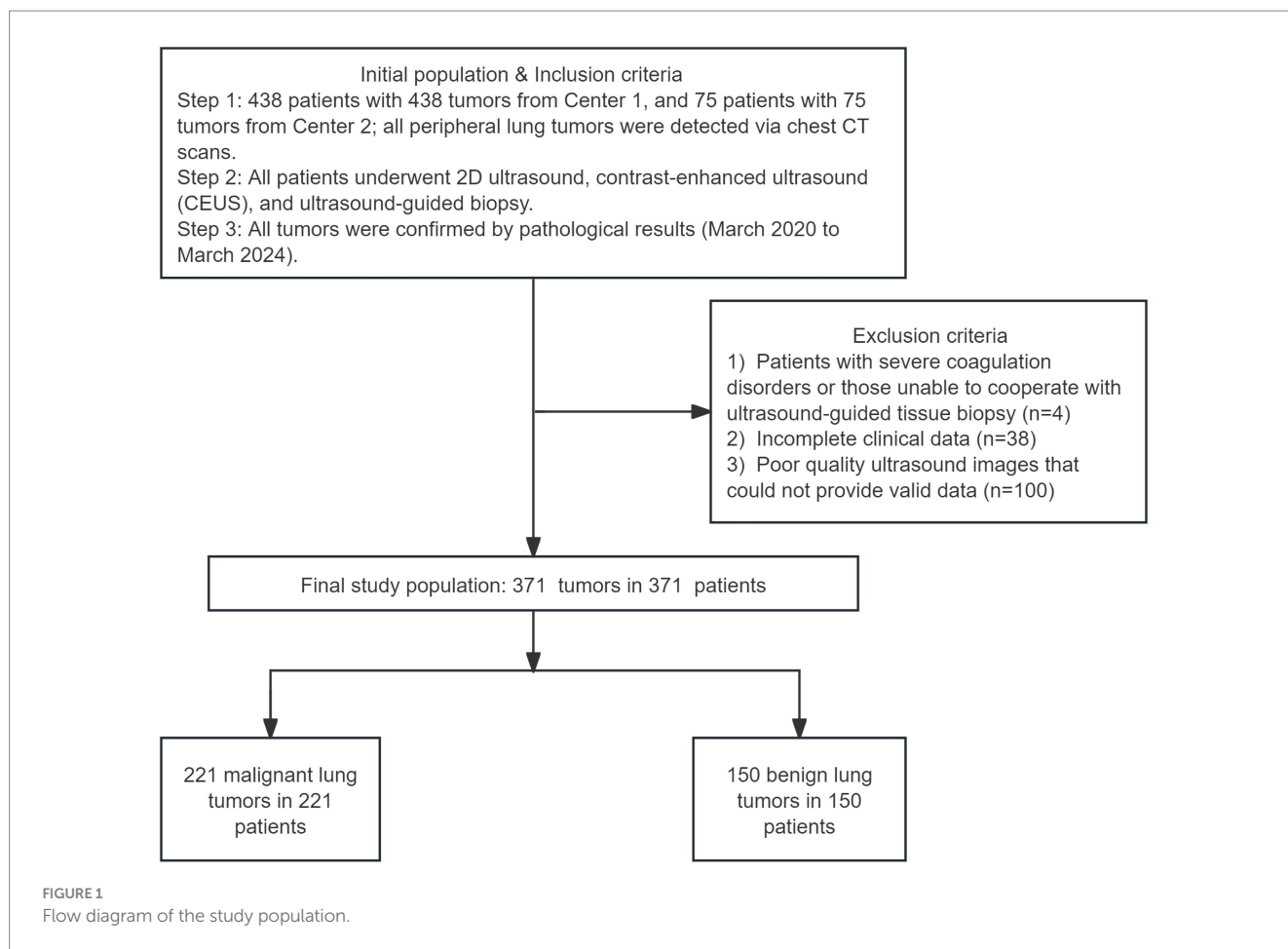
comprehensive ultrasonic examination. Clear two-dimensional ultrasound images, capturing the maximum cross-sectional area of the lesions, were retained for further analysis. To ensure consistency in data quality, all ultrasound images were acquired by trained physicians adhering to standardized imaging protocols. Images with lower resolution or insufficient signal-to-noise ratio (SNR) were excluded from the study to maintain uniformity across the dataset.

Ultrasound image analysis and modeling

After anonymizing patient information, the original ultrasound images were imported into the Darwin AI Research Platform for further processing. The patient information labels included the following: gender, age, biopsy site, lesion size, history of lung diseases, smoking history, and lung tumor markers. Lesion-related labels encompassed pathological results (benign or malignant), shape (round, quasi-round, triangular, wedge-shaped, or irregular), echogenicity (homogeneous or heterogeneous), presence of necrosis (present or absent), air bronchogram (present or absent), and boundary clarity (clear or unclear). Regions of interest (ROIs) were manually delineated by physicians with more than 5 years of relevant experience. In cases of disagreement, senior physicians were consulted for a definitive diagnosis. The dataset, comprising 371 patients with lung tumors, was randomly divided into a training set ($n = 296$) and a test set ($n = 75$) in an 8:2 ratio. Peripheral lung tumor ROI images

and corresponding clinical data were input into the system, and the output indicated whether the tumors were benign or malignant. Based on these ultrasound imaging data and annotations, deep learning models were developed to predict the benign or malignant nature of peripheral lung tumors using five distinct algorithms: ResNet152, ResNet101, ResNet50, ResNet34, and ResNet18. Receiver operating characteristic (ROC) curves were plotted, and the area under the curve (AUC) was calculated to assess the diagnostic performance of each model. The complete experimental process is illustrated in Figure 2. During model training, we optimized hyperparameters including a learning rate of 0.001, a batch size of 64, and the Adam optimizer. We used a cosine annealing scheduler with a warm-up period for learning rate variation, and employed cross-entropy loss to guide the model in minimizing prediction errors. We implemented this study on a computer equipped with an Nvidia RTX A2000 GPU and an Intel Xeon Silver 4,208 CPU using the Darwin AI Research Platform. The average inference time per sample is approximately 50 milliseconds, measured on the aforementioned hardware. This time may vary depending on the system configuration and the complexity of the input data.

Observation indicators

The sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) of the five models in

FIGURE 2
The complete experimental process.
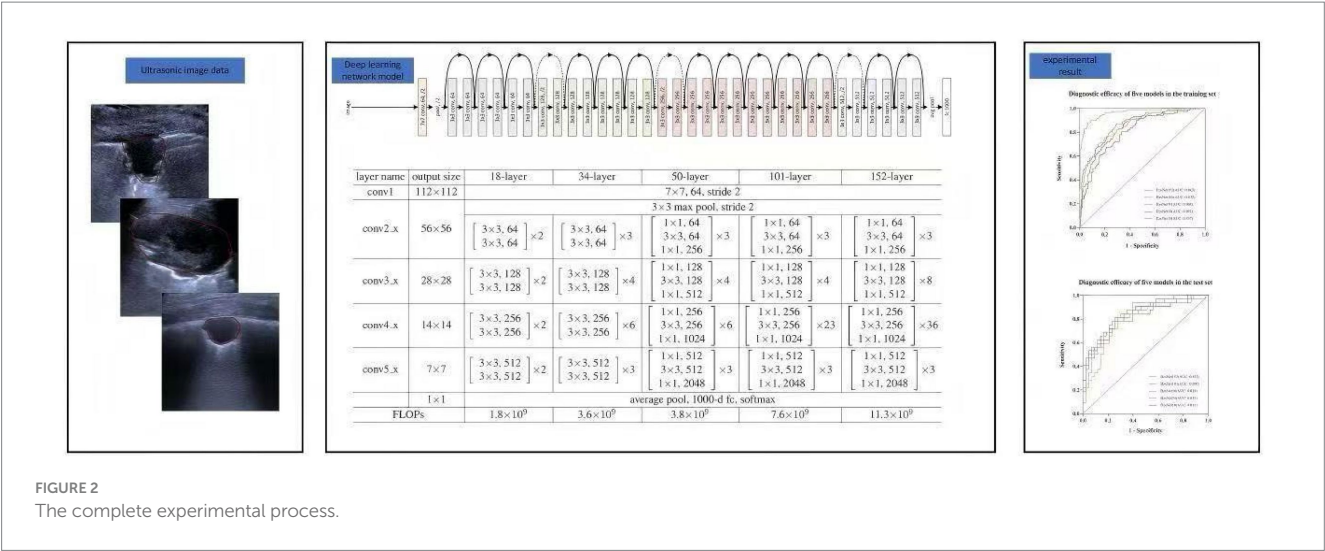
diagnosing the benignity or malignancy of peripheral lung tumors were assessed. ROC curves were plotted for each model, and the AUC was calculated to measure their diagnostic performance.

Statistical methods

SPSS version 27.0 statistical analysis software was used to evaluate the significance of each model. Categorical data were expressed as frequencies and percentages. The classification performance of the models was assessed using the AUC, accuracy, sensitivity, specificity, PPV, and NPV derived from the ROC curves. The DeLong test was employed to compare the AUCs among the five versions of ResNet. A p-value of less than 0.05 was considered statistically significant, indicating a meaningful difference in performance.

Results

Pathological results

The study included a cohort of 371 patients diagnosed with peripheral lung tumors. Pathological analysis, based on biopsy or surgical resection, identified 221 malignant and 150 benign tumors. Detailed histological classifications are provided in Table 1. Among the malignant tumors, adenocarcinoma was the most common (26.95%), followed by squamous cell carcinoma (20.75%) and small cell carcinoma (5.12%). Benign lesions were predominantly chronic inflammation of lung tissue (26.42%) and organizing pneumonia (5.39%).

Performance of the deep learning models

In the training set, the sensitivity, specificity, and diagnostic accuracy for diagnosing the benignity or malignancy of peripheral lung tumors were as follows: 87.2, 70.4, and 77.0% for Model 152; 70.1, 85.5, and 79.4% for Model 101; 88.0, 93.3, and 91.2% for Model 50; 80.3, 66.5, and 72.0% for Model 34; and 82.1, 70.4, and 75.0% for

TABLE 1 Pathological results of 371 peripheral lung tumors.

Pathological findings	Benign n (%)	Malignant n (%)
Small cell carcinoma		19 (5.12)
Squamous cell carcinoma		77 (20.75)
Adenocarcinoma		100 (26.95)
Adenosquamous carcinoma		6 (1.62)
Large cell carcinoma		6 (1.62)
Malignant pleomorphic tumor		3 (0.81)
Mesenchymal sarcoma		2 (0.54)
Choriocarcinoma		2 (0.54)
Alveolar carcinoma		1 (0.27)
Metastatic renal clear cell carcinoma		2 (0.54)
Carcinosarcoma		2 (0.54)
MALT-L		1 (0.27)
Tuberculosis	11 (2.96)	
Organizing pneumonia	20 (5.39)	
Granulomatous inflammation	17 (4.58)	
Vasculitic lung injury	1 (0.27)	
Bacterial pneumonia	1 (0.27)	
Chronic inflammation of lung tissue	98 (26.42)	
Atypical adenomatous hyperplasia	2 (0.54)	

MALT-L, Mucosa-Associated Lymphoid Tissue Lymphoma.

Model 18. In the test set, the corresponding values were 78.1, 74.4, and 76.0% for Model 152; 81.3, 72.1, and 76.0% for Model 101; 81.3, 74.4, and 77.2% for Model 50; 78.1, 74.4, and 76.0% for Model 34; and 84.4, 69.8, and 76.0% for Model 18 (Table 2). The areas under the receiver operating characteristic (ROC) curves (AUCs) for the five models in the training set were 0.865, 0.852, 0.960, 0.803, and 0.835, respectively (Figure 3). In the test set, the AUCs were 0.822, 0.800, 0.824, 0.823, and 0.831, respectively (Figure 4).

DeLong's test revealed statistically significant differences in the AUCs between the five models, with ResNet18 outperforming the

TABLE 2 Comparison of the performance of each deep learning model in the training and test sets.

	ResNet152		ResNet101		ResNet50		ResNet34		ResNet18	
	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
AUC	0.865	0.822	0.852	0.800	0.960	0.824	0.803	0.823	0.835	0.831
ACC	0.770	0.760	0.794	0.760	0.912	0.772	0.720	0.760	0.750	0.760
SEN	0.872	0.781	0.701	0.813	0.880	0.813	0.803	0.781	0.821	0.844
SPE	0.704	0.744	0.855	0.721	0.933	0.744	0.665	0.744	0.704	0.698
F1score	0.750	0.735	0.729	0.743	0.888	0.754	0.694	0.735	0.722	0.750
PPV	0.658	0.694	0.759	0.684	0.896	0.703	0.610	0.694	0.644	0.675
NPV	0.894	0.821	0.814	0.838	0.923	0.842	0.838	0.821	0.857	0.857
p value	2.648E-26	2.000E-06	1.237E-24	1.000E-05	4.249E-39	2.000E-06	2.2448E-18	2.000E-06	2.089E-22	1.000E-06
95%CI	0.823,0.906	0.725,0.919	0.808,0.896	0.701,0.899	0.937,0.983	0.724,0.924	0.750,0.851	0.726,0.921	0.788,0.881	0.738,0.925

AUC, Area Under the Curve; ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; F1score, F1 Score; PPV, Positive Predictive Value; NPV, Negative Predictive Value; 95%CI, 95% Confidence Interval.

FIGURE 3
ROC curves of each deep learning model in the training set.
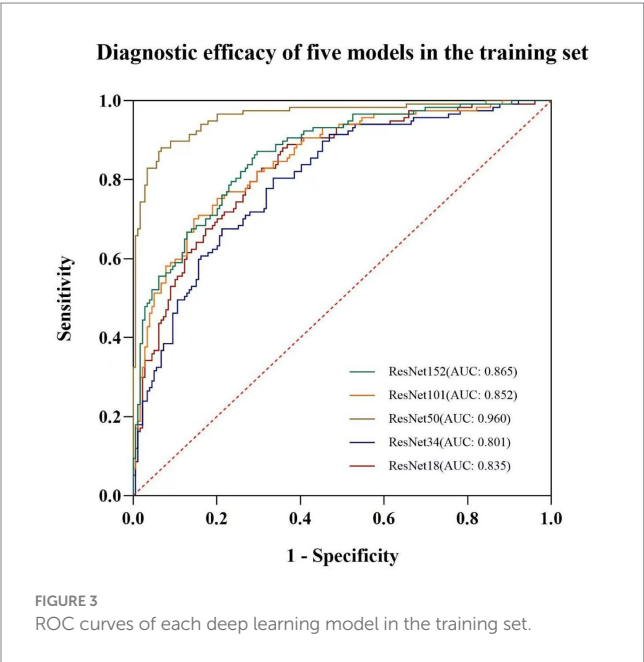
other models in terms of predictive accuracy and discriminatory power. NRI analysis revealed substantial improvements in the ResNet18 model compared to the other models. The NRI values for each model were as follows: ResNet152 (NRI = 0.180), ResNet101 (NRI = 0.240), ResNet50 (NRI = 0.186), and ResNet34 (NRI = 0.221). All NRI comparisons yielded statistically significant results, with p-values less than 0.05. These findings further substantiate that the ResNet18 model outperformed the other models in terms of reclassification ability and predictive accuracy.

Discussion

The current study successfully developed a deep learning model based on ultrasound imaging to differentiate benign and malignant peripheral lung tumors. This model, utilizing the ResNet18 architecture, demonstrated superior performance with an AUC of 0.835 in the training cohort and 0.831 in the testing cohort, compared to models based on other ResNet architectures (ResNet152, ResNet101, ResNet50, and ResNet34). The ResNet18 model significantly outperformed the other models in terms of predictive accuracy, discriminatory power, and reclassification ability, making it a promising tool for early lung cancer detection.

In recent years, deep learning techniques have significantly improved lung tumor diagnosis across imaging modalities like CT, PET/CT, and ultrasound. CT and PET/CT are commonly used in clinical settings for their high spatial resolution and detailed anatomical information. Studies have shown that deep learning can enhance the performance of these techniques in detecting malignancies. For example, Yang et al. (12) used deep convolutional neural networks (CNNs) to analyze CT scans, achieving over 90% accuracy in distinguishing between benign and malignant pulmonary nodules (12).

Despite their high diagnostic accuracy, CT and PET/CT have several limitations. Both involve ionizing radiation, which can lead to cumulative exposure risks, especially with repeated imaging. Additionally, PET/CT scanners are expensive and less accessible, limiting their use in some clinical settings (13). In contrast, ultrasound offers significant advantages, especially when combined with deep learning techniques. Unlike CT and PET/CT, ultrasound does not involve ionizing radiation, making it a safer option for patients, particularly in long-term monitoring. Ultrasound is also more cost-effective, portable, and accessible, making it ideal for resource-limited settings. Recent studies, such as Liu et al. (14), have shown that deep learning applied to ultrasound images can achieve 88% sensitivity and 85% specificity for early lung cancer detection (14). Ultrasound's real-time imaging capability provides immediate feedback, aiding quick decision-making, and it can be performed at the patient's bedside, making it a valuable tool for point-of-care diagnosis (15). However, ultrasound does have limitations. Its quality is highly dependent on the skill of the operator, which can lead to inconsistent results. Ultrasound may also struggle to visualize deeper lung tissues due to interference from air in the lungs and difficulty

FIGURE 4
ROC plots of each deep learning model in the test set.
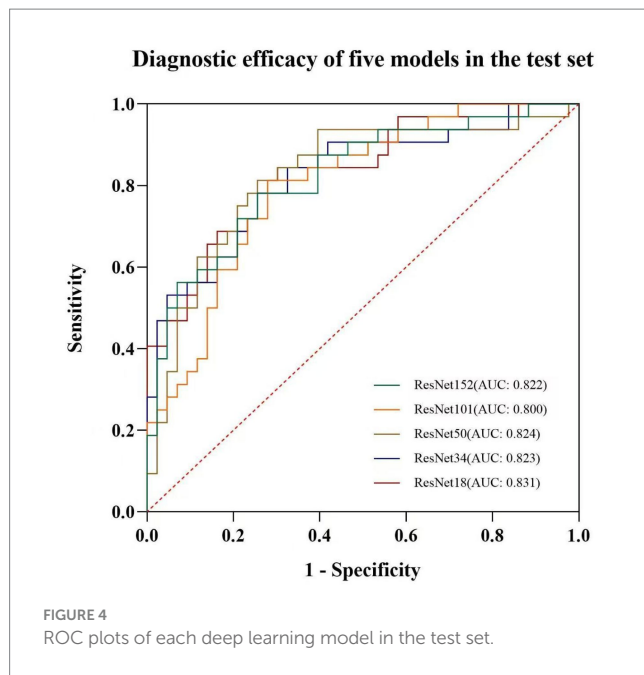
distinguishing solid tumors from surrounding structures. Additionally, deep learning algorithms for ultrasound are still underdeveloped compared to those for CT and PET/CT, which have more standardized images (16). Despite these challenges, ultrasound's non-invasive nature, lack of ionizing radiation, portability, and real-time feedback make it a promising tool for lung cancer detection, particularly when enhanced by advanced deep learning techniques (17–20).

The reasons for choosing ResNet to construct the predictive model in this study, rather than other architectures (e.g., EfficientNet, Vision Transformer), are as follows: firstly, ResNet introduces the concept of residual connections, which address the vanishing gradient problem by allowing gradients to flow more easily through deeper layers (21). This enables the training of much deeper networks without the degradation in performance typically seen in conventional deep networks (22). In comparison, while EfficientNet and ViT also achieve high performance, they do not inherently mitigate the vanishing gradient problem to the same extent, especially in very deep architectures (23, 24). Secondly, ResNet excels in feature extraction, leveraging its deep architecture and residual blocks, which enables it to capture more complex patterns and fine-grained details in images (25). This is particularly useful in medical imaging, where subtle differences in image features are crucial for accurate diagnosis (26). EfficientNet and ViT, while powerful, may not always achieve the same level of fine-grained feature extraction, particularly for highly specialized tasks such as detecting peripheral lung tumors (27, 28). Thirdly, ResNet offers a good balance between model depth and computational cost (29). Although deeper networks typically require more computation, the residual connections in ResNet allow for more efficient training and inference compared to other architectures like ViT, which can be computationally expensive due to the self-attention mechanism (30). EfficientNet, on the other hand, optimizes the trade-off between accuracy and

efficiency, but its scaling strategy might still be less computationally efficient than ResNet in certain applications (31, 32). Fourthly, ResNet has been extensively validated across a wide range of medical imaging tasks, demonstrating robustness and reliability (33). It has a proven track record in both image classification and segmentation tasks (34, 35). While architectures like EfficientNet and ViT also show great promise, ResNet's long-standing success in medical imaging, along with its established frameworks for fine-tuning, makes it a reliable choice for clinical applications (36). In summary, ResNet's advantages lie in its deep network capability, residual learning to avoid gradient issues, efficient feature extraction, and computational practicality, all of which make it particularly suitable for medical image analysis compared to other architectures like EfficientNet and Vision Transformer (37).

This study evaluated the performance of various ResNet architectures in predicting the benign or malignant nature of peripheral lung tumors. The findings revealed that the ResNet18-based model outperformed those based on ResNet152, ResNet101, ResNet50, and ResNet34. A deeper analysis, considering both the algorithmic network architecture and the dataset, provides valuable insights into the factors contributing to this result. Firstly, ResNet18, being a relatively shallow model, has fewer layers compared to deeper networks like ResNet152. This means it requires less computational power, leading to faster training times and quicker inference speeds. This can be important in resource-constrained environments, such as embedded systems or mobile devices (38). Secondly, with fewer parameters and layers, ResNet18 demands less memory for storage and computation. This is beneficial in settings where memory is limited, and it can be crucial for deployment in edge devices or situations where there is a need to optimize for power consumption and storage (39). Thirdly, in many real-world tasks, particularly when the dataset size is not large enough to fully leverage the capacity of deeper networks, a smaller model like ResNet18 can avoid overfitting. Larger models like ResNet152, due to their increased number of parameters, may overfit on smaller datasets if not properly regularized (40). Fourthly, when performing transfer learning, using a smaller model like ResNet18 may lead to easier fine-tuning, especially on datasets where the target task is relatively simpler. The smaller number of parameters also means it is easier to modify the model for specific use cases without requiring excessive computational resources (41). Fifthly, while deeper models (like ResNet152) might provide better performance on very large datasets, ResNet18 offers a good balance between performance and computational efficiency. It can achieve decent accuracy with much less computational cost, making it ideal for practical applications where speed is essential (42). Sixthly, due to fewer layers and parameters, ResNet18 is more interpretable and simpler to analyze compared to deeper architectures. This simplicity can be beneficial for debugging or understanding how the network is making decisions (43). Overall, ResNet18 is preferred in situations where computational efficiency, memory constraints, or training time are critical considerations, while deeper ResNets like ResNet152 might be more suitable for large-scale datasets and applications that demand the highest performance.

There were still several limitations in this study. Firstly, while the results achieved are promising, the sample size remains limited. Future studies should include a larger cohort of patients, and external validation using independent test datasets is essential to confirm the generalizability of the model. Secondly, all data in this study were sourced from two centers. Therefore, additional multi-center research is required to enhance the robustness and applicability of the findings in broader clinical settings. Thirdly, the model is employed solely for the classification of benign and malignant lung tumors. A more comprehensive analysis focusing on their pathological classification will be conducted in subsequent studies.

Conclusion

The deep learning model based on ResNet18 demonstrated superior performance in differentiating between benign and malignant peripheral lung tumors compared to other ResNet-based models. The ResNet18 model exhibited statistically significant improvements in predictive accuracy and discriminatory power, as evidenced by ROC analysis and NRI evaluations. These findings highlight the potential of ultrasound imaging, in combination with advanced deep learning techniques, as an effective and non-invasive approach for the early detection of lung cancer. This study supports the clinical application of ResNet18 in enhancing diagnostic accuracy and risk stratification for lung lesions, contributing to more timely and accurate diagnosis of lung cancer.

Data availability statement

The datasets presented in this article are not readily available. Requests to access the datasets should be directed to Yuan Wang, wy18329591877@163.com.

Ethics statement

The studies involving humans were approved by The Second Affiliated Hospital of Xi'an Jiaotong University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YW: Formal analysis, Supervision, Writing – review & editing. YZ: Writing – review & editing. YL: Methodology, Validation, Writing – original draft. TS: Data curation, Formal analysis, Project administration, Writing – original draft. MH: Data curation, Writing – original draft. HH: Writing – original draft. DZ: Data curation, Formal analysis, Investigation, Project administration, Writing – review & editing. JJ: Investigation, Writing – original draft.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1567545/full#supplementary-material

References

1. Saji H, Okada M, Tsuboi M, Tanaka M, Yamamoto Y, Hasegawa S. Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicenter, open-label, phase 3, randomised, controlled, non-inferiority trial. *Lancet*. (2022) 399:1607–17. doi: 10.1016/S0140-6736(21)02333-3

2. Li W, Shen M, Zhang Y, Huang S, Liu Z, Chen H. A model for predicting malignant sub-pleural solid masses using grayscale ultrasound and ultrasound elastography. *Ultrasound Med Biol*. (2021) 47:1212–8. doi: 10.1016/j.ultrasmedbio.2021.01.017

3. Gandhi Z, Gurram P, Amgai B, Patel R, Sharma N, Bhatt A. Artificial intelligence and lung cancer: impact on improving patient outcomes. *Cancers (Basel)*. (2023) 15:5236. doi: 10.3390/cancers15215236

4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2024) 74:229–63. doi: 10.3322/caac.21834

5. Lancaster HL, HeuvelmansMA OM. Low-dose computed tomography lung cancer screening: clinical evidence and implementation research. *J Intern Med*. (2022) 292:68–80. doi: 10.1111/joim.13480

6. Hu M, Wu L, Zhang X, Yuan Q, Li P, Yang S. Comparative evaluation of 2 different percutaneous techniques of simultaneous needle biopsy with microwave ablation of suspected malignant pulmonary nodules. *Technol Cancer Res Treat*. (2023) 22:15330338231168458. doi: 10.1177/15330338231168458

7. Jiang T, Chen C, Zhou Y, Zhang L, Wang J, Li S. Deep learning-assisted diagnosis of benign and malignant parotid tumors based on ultrasound: a retrospective study. *BMC Cancer*. (2024) 24:510. doi: 10.1186/s12885-024-12277-8

8. Sigrist RMS, Liau J, Kaffas AE, Chammas MC, Willmann JK. Ultrasound elastography: review of techniques and clinical applications. *Theranostics*. (2027) 7:1303–29. doi: 10.7150/thno.18650

9. Bai Z, Liu T, Liu W, Li Z, Zheng H, Li X. Application value of contrast-enhanced ultrasound in the diagnosis of peripheral pulmonary focal lesions. *Medicine (Baltimore)*. (2022) 10:e29605. doi: 10.1097/MD.0000000000029605

10. Xu W, Yang H, Li H, Zhou Y, Zhang S, Wang Y. Res net and its application to medical image processing: research progress and challenges. *Comput Methods Prog Biomed*. (2023) 240:107660. doi: 10.1016/j.cmpb.2023.107660

11. Taye MM, Zhang Y, Li W, Liu Y, Wang Y, Lin B. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Compute*. (2023) 12. doi: 10.3390/computers12050091

12. Yang X, Zhang J, Li H, Chen X, Lu S, Wang Z. Deep convolutional neural networks to predict cardiovascular risk using coronary artery calcium scoring. *Nat Commun*. (2020) 12:1–9. doi: 10.1038/s41467-021-20966-2

13. Aide N, Iravani A, Prigent K, Kottler D, Alipour R, Hicks RJ. PET/CT variants and pitfalls in malignant melanoma. *Cancer Imaging*. (2022) 22:3. doi: 10.1186/s40644-021-00440-4

14. Liu W, Liu X, Li H, Li M, Zhao X, Zhu Z. Integrating lung parenchyma segmentation and nodule detection with deep multi-task learning. *IEEE J Biomed Health Inform*. (2021) 25:3073–81. doi: 10.1109/JBHI.2021.3053023

15. Bonadia N, Carnicelli A, Piano A, Buonsenso D, Gilardi E, Kadhim C, et al. Lung ultrasound findings are associated with mortality and need for intensive care admission in COVID-19 patients evaluated in the emergency department. *Ultrasound Med Biol*. (2020) 46:2927–37. doi: 10.1016/j.ultrasmedbio.2020.07.005

16. Hashimoto F, Onishi Y, Ote K, Tashima H, Reader AJ, Yamaya T. Deep learning-based PET image denoising and reconstruction: a review. *Radiol Phys Technol*. (2024) 17:24–46. doi: 10.1007/s12194-024-00780-3

17. Ning J, Ge T, Jiang M, Jia K, Wang L, Li W, et al. Early diagnosis of lung cancer: which is the optimal choice? *Aging (Albany NY)*. (2021) 13:6214–27. doi: 10.18632/aging.202504PMC7950268

18. Luijten B, ChennakeshavaN Eldar YC, Mischi M, van Sloun RJG. Ultrasound signal processing: from models to deep learning. Ultrasound Med Biol. (2023) 49:677-698. doi: 10.1016/j.ultrasmedbio.2022.11.003

19. Akkus Z, Cai J, Boonrod A, Zeinoddini A, Weston AD, Philbrick KA, et al. A Survey of deep-learning applications in ultrasound: artificial intelligence-powered ultrasound for improving clinical workflow. *J Am Coll Radiol*. (2019) 16:1318–28. doi: 10.1016/j.jacr.2019.06.004

20. Rauby B, Xing P, Gasse M, Provost J. Deep learning in ultrasound localization microscopy: applications and perspectives. *IEEE Trans Ultrason Ferroelectr Freq Control*. (2024) 71:1765–84. doi: 10.1109/TUFFC.2024.3462299

21. Ahmad IS, Dai J, Xie Y, Liang X. Deep learning models for CT image classification: a comprehensive literature review. *Quant Imaging Med Surg*. (2025) 15:962–1011. doi: 10.21037/qims-24-1400

22. Chen W, Zhang C, Zhang W, Cai J. Class-hidden client-side watermarking in federated learning. *Entropy (Basel)*. (2025) 27:134. doi: 10.3390/e27020134

23. Hattori S, Saggar R, Heidinger E, Qi A, Mullen J, Fee B, et al. Advances in ultrasound-guided surgery and artificial intelligence applications in musculoskeletal diseases. *Diagnostics (Basel)*. (2024) 14:2008. doi: 10.3390/diagnostics14182008

24. Zhu B, Yang Y. Quality assessment of abdominal CT images: an improved ResNet algorithm with dual-attention mechanism. *Am J Transl Res*. (2024) 16:3099–107. doi: 10.62347/WKNS8633

25. Zhou B, Yang X, Curran WJ, Liu T. Artificial intelligence in quantitative ultrasound imaging: a survey. *J Ultrasound Med*. (2022) 41:1329–42. doi: 10.1002/jum.15819

26. Wang B. Data feature extraction method of wearable sensor based on convolutional neural network. *J Healthc Eng*. (2022) 2022:1580134. doi: 10.1155/2022/1580134

27. Prottasha NJ, Sami AA, Kowsher M, Murad SA, Bairagi AK, Masud M, et al. Transfer learning for sentiment analysis using BERT based supervised fine-tuning. *Sensors (Basel)*. (2022) 22:4157. doi: 10.3390/s22114157

28. Kou C, Li W, Liang W, Yu Z, Hao J. Microaneurysms segmentation with a U-Net based on recurrent residual convolutional neural network. *J Med Imaging (Bellingham)*. (2019) 6:025008. doi: 10.1117/1.JMI.6.2.025008

29. Prasad VK, Verma A, Bhattacharya P, Shah S, Chowdhury S, Bhavsar M, et al. Revolutionizing healthcare: a comparative insight into deep learning's role in medical imaging. *Sci Rep*. (2024) 14:30273. doi: 10.1038/s41598-024-71358-7

30. Lee CP, Lim KM, Song YX, Alqahtani A. Plant-CNN-ViT: Plant Classification with Ensemble of Convolutional Neural Networks and Vision Transformer. *Plants (Basel)*. (2023) 12:2642. doi: 10.3390/plants12142642

31. Venugopal V, Joseph J, Vipin Das M, Kumar NM. An EfficientNet-based modified sigmoid transform for enhancing dermatological macro-images of melanoma and nevi skin lesions. *Comput MethodsPrograms Biomed*. (2022) 222:106935. doi: 10.1016/j.cmpb.2022.106935

32. Wang J, Liu Q, Xie H, Yang Z, Zhou H. Boosted EfficientNet: detection of lymph node metastases in breast cancer using convolutional neural networks. *Cancers (Basel)*. (2021) 13:661. doi: 10.3390/cancers13040661

33. Li S, Gao P, Zhang Y, Liu X, Cai Y, Wang Z. Evaluating ResNet for detecting early-stage Alzheimer's disease in MRI scans. *Front Neurosci*. (2024) 18:121–9. doi: 10.3389/fnins.2024.1011982

34. Tan W, Yang H, Sun Y, Li J, Zhang H, Wang T. ResNet-50 for enhanced identification of cardiovascular diseases in chest X-rays. *Front Med Imaging*. (2023) 7:231–8. doi: 10.3389/fmed.2023.1098290

35. Wang YW, Kuo TT, Chou YH, Su Y, Huang SH, Chen CJ. Breast tumor classification using Short-ResNet with pixel-based tumor probability map in ultrasound images. *Ultrason Imaging*. (2023) 45:74–84. doi: 10.1177/01617346231162906

36. Rahhal MMA, Bazi Y, Jomaa RM, Zuair M, Melgani F. Contrasting EfficientNet, ViT, and gMLP for COVID-19 detection in ultrasound imagery. *J Pers Med*. (2022) 12:1707. doi: 10.3390/jpm12101707

37. Abedalla A, Abdullah M, Al-Ayyoub M, Benkhelifa E. Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures. *PeerJ Comput Sci*. (2021) 7:e607. doi: 10.7717/peerj-cs.607

38. He J, Xu J, Zhang L, Zhu J. An interpretive constrained linear model for ResNet and MgNet. *Neural Netw*. (2023) 162:384–92. doi: 10.1016/j.neunet.2023.03.011

39. Ma L, Ma J, Zelminbek M, Zhang W. High-Resolution reconstruction of temperature fields basedon improved ResNet18. *Sensors (Basel)*. (2024) 24:6564. doi: 10.3390/s24206564

40. Naz J, Sharif MI, Sharif MI, Kadry S, Rauf HT, Ragab AE. A Comparative analysis of optimization algorithms for gastrointestinal abnormalities recognition and classification based on ensemble XcepNet23 and ResNet18 features. *Biomedicines*. (2023) 11:1723. doi: 10.3390/biomedicines11061723

41. Shan L, Tang CF, Hong B, Kong M. Reconstruction of a three-dimensional temperature field in flames based on ES-ResNet18. *Appl Opt*. (2024) 63:1982–90. doi: 10.1364/AO.515383

42. Kaur P, Mahajan P. Detection of brain tumors using a transfer learning-based optimized ResNet152 model in MR images. *Comput Biol Med*. (2025) 188:109790. doi: 10.1016/j.compbiomed.2025.109790

43. Lin C, Sun G, Wu D, Xie C. Vehicle detection and tracking with roadside LiDAR using improved ResNet18 and the hungarian algorithm. *Sensors (Basel)*. (2023) 23:8143. doi: 10.3390/s23198143

Check for updates

An optimized deep learning model based on transperineal ultrasound images for precision diagnosis of female stress urinary incontinence

Ke Chen[1], Qi Chen[2], Ning Nan[2], Lu Sun[1], Miaoyan Ma[1] and Shanshan Yu[1]*

[1]Department of Ultrasound, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2]Department of Urology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

Background: Transperineal ultrasound (TPUS) is widely utilized for the evaluation of female stress urinary incontinence (SUI). However, the diagnostic accuracy of parameters related to urethral mobility and morphology remains limited and requires further optimization.

Objective: This study aims to develop and validate an optimized deep learning (DL) model based on TPUS images to improve the precision and reliability of female SUI diagnosis.

Methods: This retrospective study analyzed TPUS images from 464 women, including 200 patients with SUI and 264 controls, collected between 2020 and 2024. Three DL models (ResNet-50, ResNet-152, and DenseNet-121) were trained on resting-state and Valsalva-state images using an 8:2 training-to-testing split. Model performance was assessed using diagnostic metrics, including area under the curve (AUC), accuracy, sensitivity, and specificity. A TPUS-index model, constructed using measurement parameters assessing urethral mobility, was used for comparison. Finally, the best-performing DL model was selected to evaluate its diagnostic advantages over traditional methods.

Results: Among the three developed DL models, DenseNet-121 demonstrated the highest diagnostic performance, achieving an AUC of 0.869, an accuracy of 0.87, a sensitivity of 0.872, a specificity of 0.761, a negative predictive value (NPV) of 0.788, and a positive predictive value (PPV) of 0.853. When compared to the TPUS-index model, the DenseNet-121 model exhibited significantly superior diagnostic performance in both the training set ($z = -2.088$, $p = 0.018$) and the testing set ($z = -1.997$, $p = 0.046$).

Conclusion: This study demonstrates the potential of DL models, particularly DenseNet-121, to enhance the diagnosis of female SUI using TPUS images, providing a reliable and consistent diagnostic tool for clinical practice.

Introduction

Stress urinary incontinence (SUI) refers to the involuntary leakage of urine during activities that increase intra-abdominal pressure, such as coughing or physical exertion. It is commonly observed in female, significantly impacting the quality of their life. Studies have reported that the prevalence of SUI in postmenopausal women ranges from 10% to 40% (1). From the pathophysiological perspective, the development of SUI is primarily associated with damage to the supportive structures of the bladder neck and proximal urethra, as well as excessive urethral mobility (2).

Transperineal ultrasound (TPUS) is widely utilized in clinical practice to evaluate pelvic floor dysfunction, including SUI. This non-invasive imaging modality provides clear visualization of pelvic floor structures, such as the urethra, bladder, and vagina, and enables quantitative assessment of urethral mobility (3, 4). Measurable parameters derived from TPUS, including bladder neck descent (BND), urethral rotation angles (URA), and urethral length, hold diagnostic value for female SUI (5–7). However, current ultrasound techniques for diagnosing SUI face significant challenges. The dependence of TPUS on operator experience may lead to inconsistencies in diagnostic results, affecting clinical decision-making. Patient cooperation directly impacts the accuracy of parameters, such as the intensity and duration of the Valsalva maneuver. Moreover, the dynamic changes in the structures surrounding the urethra are complex, and existing parameters may overlook certain important functional abnormalities, resulting in incomplete diagnosis. These limitations underscore the need for innovative approaches to enhance the accuracy and reliability of SUI diagnosis.

Recent advancements in artificial intelligence (AI), particularly deep learning (DL) algorithms, have demonstrated significant potential in enhancing diagnostic accuracy in medical imaging. Unlike traditional machine learning methods, DL models automatically extract detailed structural features from raw data without requiring operator expertise or manually designed feature extraction (8). DL algorithms exhibit exceptional proficiency in segmenting pelvic floor ultrasound images and identifying pelvic floor structures. Additionally, they are capable of dynamically segmenting and automatically measuring anterior pelvic structures, such as bladder neck descent and urethral rotation angles (9–12). These capabilities make DL models potentially capable of achieving breakthroughs in addressing the limitations of conventional TPUS diagnostics for female SUI.

This study aims to develop and validate a convolutional neural network (CNN)-based DL model optimized for TPUS imaging to overcome key challenges in diagnosing SUI. By harnessing DL's capabilities for automated feature extraction and precise analysis, the model strives to enhance diagnostic accuracy, while reducing reliance on operator expertise and mitigating measurement variability. Ultimately, this study seeks to establish a diagnostic tool that can be integrated into clinical workflows, facilitating earlier detection and improved management of SUI.

Materials and methods

Objects

The study retrospectively collected data from female patients who underwent TPUS examinations at the Department of Urology and Gynecology, the Second Affiliated Hospital of Xi'an Jiaotong University from 2020 to 2024. The study was conducted in compliance with the Declaration of Helsinki and approved by the Institutional Review Board (IRB number 2020823).

All patients underwent a comprehensive clinical evaluation, which included completing the International Consultation on Incontinence Questionnaire-Urinary Incontinence Short Form (ICIQ-UI-SF), the International Consultation on Incontinence Questionnaire-Female Lower Urinary Tract Symptoms (ICIQ-FLUTS), urinalysis, uroflowmetry, and maintaining a 3-day bladder diary (13, 14). Inclusion criteria for the SUI group were: (1) age >18 years; (2) clinical diagnosis of SUI; (3) availability of complete TPUS images and clinical data. Exclusion criteria included: (1) residual urine volume >50 mL; (2) history of pelvic or pelvic floor reconstructive surgery; (3) active urinary tract infection or history of urogenital tumors; (4) unclear ultrasound images or inability to perform the Valsalva maneuver. A total of 200 patients meeting these criteria were included in the SUI group. Additionally, 264 female patients without a diagnosis of SUI during the same period, who fulfilled the inclusion criteria, were selected as the control group (non-SUI group). Transperineal ultrasound images were collected for all patients, including 464 resting-state images and 464 Valsalva-state images (Figure 1).

Ultrasound images acquisition

TPUS examinations were conducted by expert sonographer with 5–10 years of the pelvic for ultrasound experience using a Mindray Resona 8 ultrasound system equipped with a DE10-3U 3D volumetric probe (frequency range: 3–10 MHz). Patients were positioned in the lithotomy position after bladder emptying, following the protocol established by Dietz HP (15). Mid-sagittal pelvic floor static images were obtained at rest and during the Valsalva maneuver. Key parameters, including bladder symphyseal distance (BSD), urethral axis angle (α angle), and retrovesical angle (RVA), were measured. Additionally, bladder neck descent (BND) and urethral rotation angle (URA) were calculated. Ultrasound static images were exported in JPG format for subsequent analysis (Figure 2).

Deep learning model development

Ultrasound images in JPG format were imported into the MedAI Darwin learning platform (http://premium.darwin.yizhun-ai.com). Using the platform's tools, the urethra was delineated as the region of interest (ROI). The annotation process was carried out by an experienced pelvic floor ultrasound specialist, and any disagreements were resolved through consensus discussion. The
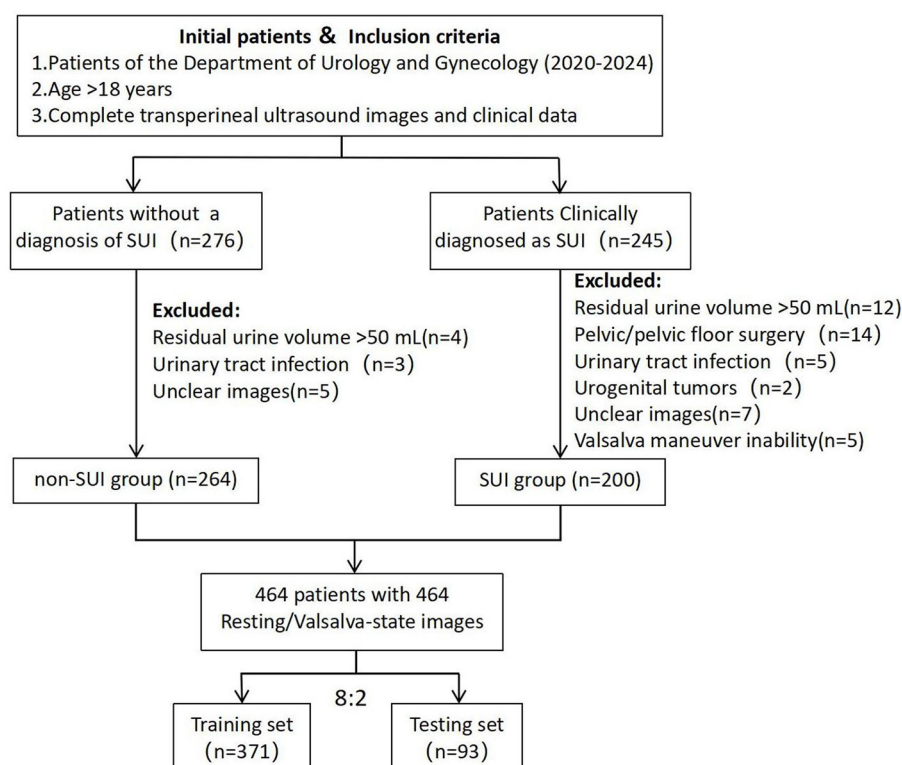
FIGURE 1
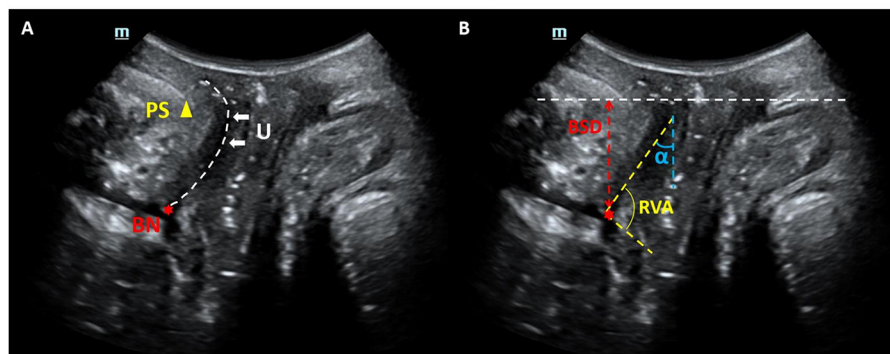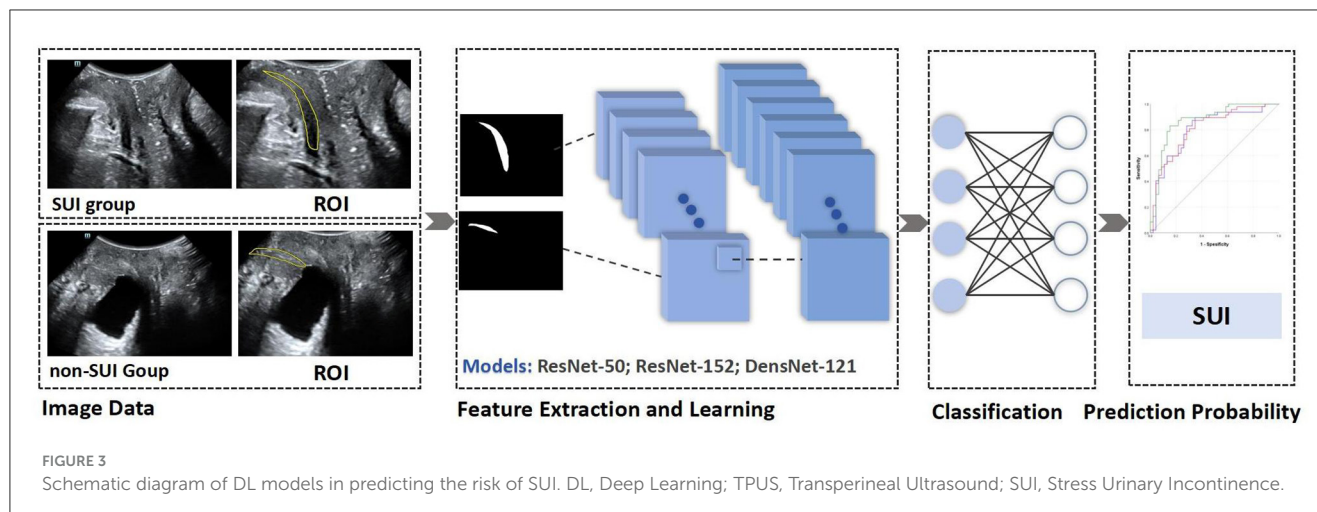The flow diagram of recruitment and grouping of research objects.

FIGURE 2
Mid-sagittal ultrasound images of the pelvic organs. **(A)** Anatomical landmarks: the public symphysis (PS), bladder neck (BN), and urethra (U) are also visualized. **(B)** Index measurement mark: bladder symphyseal distance (BSD), urethral axis angle (α angle) and retrovesical angle (RVA).

dataset included 928 images (464 resting-state and 464 Valsalva-state), with annotated ROIs. The annotated data were randomly divided into a training set ($n = 371$) and a testing set ($n = 93$) in an 8:2 ratio. Following this, preprocessing operations such as data augmentation and normalization were performed on the input ROI sub-images, including random flipping, image transposition, and pixel value normalization (16).

Three DL architectures were implemented for model development: ResNet-50, ResNet-152, and DenseNet-121. The models were trained separately on resting-state and Valsalva-state images to predict the presence of SUI. Model performance was evaluated using standard diagnostic metrics, including the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Receiver operating characteristic (ROC) curves were generated to visualize and compare the classification performance of the models. The complete experimental workflow is shown in Figure 3.

FIGURE 3
Schematic diagram of DL models in predicting the risk of SUI. DL, Deep Learning; TPUS, Transperineal Ultrasound; SUI, Stress Urinary Incontinence.

Construction of the TPUS-index model in predicting the risk of SUI

A TPUS-index model was constructed using ultrasonic measurement parameters for assessing urethral morphology and mobility as independent variables, with SUI diagnosis as the outcome variable. Binary logistic regression analyses were performed to identify significant predictors of SUI, and the model's predictive performance was evaluated using AUC, accuracy, sensitivity, and specificity.

Statistical analysis

Statistical analyses were conducted using SPSS software (version 26.0; IBM Corp). Continuous variables were expressed as mean $\pm$ standard deviation (SD) and compared using independent samples t-tests. Categorical variables were presented as counts and percentages and analyzed using chi-square tests. Univariate and multivariate logistic regression with forward stepwise analysis were applied to screen for independent risk factors and establish a TPUS-index model. The predictive performance of the model was evaluated by plotting ROC curves and calculating the AUC along with the consistency index (CI). Comparisons of AUC values between models were performed using z-tests. All statistical tests were two-tailed, and a *p-value* < 0.05 was considered statistically significant.

Results

Baseline characteristics

The baseline characteristics of the study population are summarized (Supplementary Table 1). The mean age of the participants was 47.90 ± 14.88 years, ranging from 19 to 90 years. The average parity was 1.38 ± 0.80, and the mean BMI

was 23.22 ± 1.80 kg/m². Among the participants, 233 women (50.2%) were postmenopausal. The clinical characteristics were balanced between the testing set and the training set, with no significant differences in age, parity, BMI, or menopausal status.

Diagnostic performance of DL models

The DL models based on resting-state images exhibited inferior diagnostic performance. In contrast, the DL models trained on Valsalva-state images demonstrated significantly better performance in predicting SUI. DenseNet-121 achieved the best discriminatory ability among the three models, with a well-balanced performance across multiple diagnostic metrics, including accuracy (81.7%), sensitivity (87.2%), and specificity (76.1%) (Table 1).

In the training set, significant differences were found between ResNet-50 and ResNet-152 ($z = -2.149$, $p = 0.032$) and between ResNet-50 and DenseNet-121 ($z = -3.568$, $p < 0.001$). However, no significant difference in AUC values was observed between ResNet-152 and DenseNet-121 ($z = -1.661$, $p = 0.097$). In the testing set, DenseNet-121 demonstrated a statistically significant superiority in AUC compared to ResNet-152 ($z = -2.372$, $p = 0.029$) and ResNet-50 ($z = -2.190$, $p = 0.018$). These results confirm the superior classification ability of DenseNet-121 in both training and testing datasets (Figure 4).
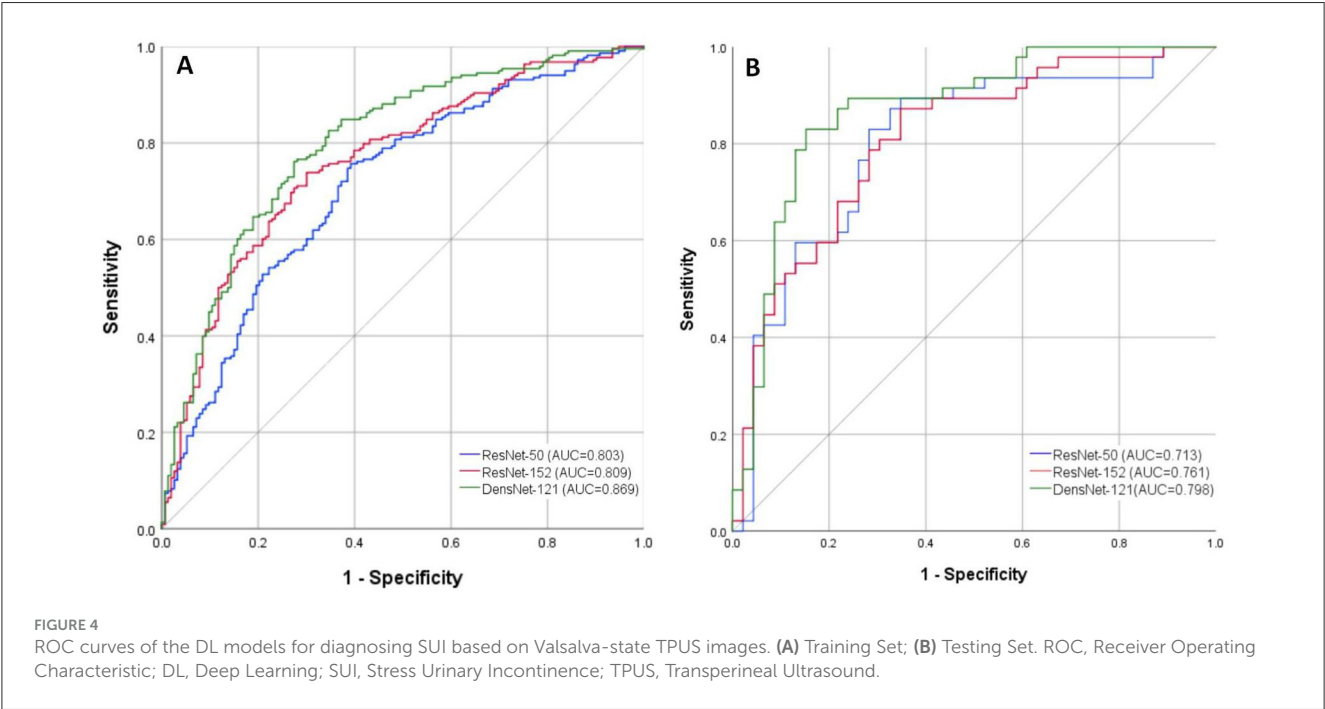
Construction of the TPUS-index model in predicting the risk of SUI

Among the TPUS measurement parameters, BSD-rest and BSD-valsalva were lower in the SUI group (OR: 0.910 and 0.894, $p < 0.001$), while α angle-valsalva, BND, and URA were higher in the SUI group (OR: 1.037, 1.124, and 1.023, $p < 0.001$).

TABLE 1 Comparison of the diagnostic performance between DL models for SUI based on TPUS images.

Image mode	DL model	Group	AUC(95%CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Resting	ResNet-50	Training set	0547 (0.487, 0.606)	0.578	0.726	0.380	0.611	0.508
		Testing set	0.605 (0.490, 0.720)	0.559	0.269	0.927	0.824	0.500
	ResNet-152	Training set	0.482 (0.421, 0.542)	0.589	0.948	0.108	0.588	0.607
		Testing set	0.553 (0.436, 0.670)	0.548	0.250	0.927	0.813	0.494
	DensNet-121	Training set	0.532 (0.472, 0.591)	0.538	0.392	0.734	0.664	0.473
		Testing set	0.490 (0.369,0.610)	0.516	0.192	0.927	0.769	0.475
Valsalva	ResNet-50	Training set	0.713 (0.659, 0.766)	0.695	0.757	0.608	0.733	0.637
		Testing set	0.803 (0.710, 0.896)	0.774	0.830	0.717	0.750	0.805
	ResNet-152	Training set	0.761 (0.712, 0.810)	0.722	0.739	0.699	0.778	0.652
		Testing set	0.809 (0.720, 0.898)	0.763	0.872	0.652	0.719	0.833
	DensNet-121	Training set	0.798 (0.752, 0.845)	0.747	0.761	0.725	0.832	0.681
		Testing set	0.869 (0.793, 0.945)	0.817	0.872	0.761	0.788	0.853

DL, Deep Learning; AUC, Area Under the ROC Curve; CI, Confidence Interval; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

FIGURE 4
ROC curves of the DL models for diagnosing SUI based on Valsalva-state TPUS images. **(A)** Training Set; **(B)** Testing Set. ROC, Receiver Operating Characteristic; DL, Deep Learning; SUI, Stress Urinary Incontinence; TPUS, Transperineal Ultrasound.

Additionally, α angle-rest was slightly lower in the SUI group (OR: 0.988, $p = 0.015$; Table 2). Based on the univariate analysis, six significant variables (BSD-rest, BSD-valsalva, α angle-rest, α angle-valsalva, BND, and URA) were included in the multivariate logistic regression analysis due to their strong statistical association with SUI ($p < 0.05$).

The Rad-score of the TPUS-index model achieved an AUC of 0.736 (95% CI: 0.629 – 0.843), with an accuracy, sensitivity, and specificity of 77.2%, 67.5%, and 75.5%, respectively (Table 3). The model formula is as follows: RadScore = −0.047*BSD(valsalva) − 0.036*BSD(rest) + 0.020*α angle(valsalva) + 0.018*BND −0.008*α angle(rest) −0.002*URA −0.003.

Comparison of the diagnostic performance between DL model and TPUS-index model

In the training set, the DL model demonstrated a significantly higher AUC compared to the TPUS-index model ($z = −2.088$, $p < 0.05$), indicating superior diagnostic performance. Similarly, in the testing set, the AUC of the DL model was also significantly higher than that of the TPUS-index model ($z = −1.997$, $p < 0.05$). Beyond the AUC, other diagnostic metrics, including accuracy, sensitivity, and specificity, was consistently better for the DL model compared to the TPUS-index model. This highlights the DL model's enhanced capability in diagnosing SUI, offering a more robust and reliable diagnostic

TABLE 2 Univariate binary logistic regression analysis of TPUS ultrasonic measurement.

TPUS measurement variable	SUI group (n = 200)	non SUI group (n = 264)	Odds ratio (95% CI)	P-value
BSD-rest(mm, mean ± SD)	22.460 ± 4.783	23.980 ± 3.465	0.910 (0.867, 0.956)	0.000
BSD-valsalva(mm, mean ± SD)	−1.310 ± 11.462	14.440 ± 7.862	0.894 (0.873, 0.916)	0.000
α angle-rest(°, mean ± SD)	24.690 ± 17.335	28.980 ± 19.646	0.988 (0.978, 0.998)	0.015
α angle-valsalva(°, mean ± SD)	45.600 ± 25.337	25.050 ± 20.535	1.037 (1.028, 1.046)	0.000
RVA-valsalva(°, mean ± SD)	137.180 ± 24.936	133.440 ± 20.467	1.008 (0.999, 1.016)	0.079
BND(mm, mean ± SD)	23.770 ± 10.552	14.440 ± 7.862	1.124 (1.095, 1.153)	0.000
URA(°, mean ± SD)	59.340 ± 29.929	40.900 ± 26.633	1.023 (1.016, 1.030)	0.000

TPUS, Transperineal Ultrasound; SUI, Stress Urinary Incontinence; CI, Confidence Interval; BSD, Bladder Symphyseal Distance; RVA, Retrovesical Angle; BND, Bladder Neck Descent; URA, Urethral Rotation Angle; SD, Standard Deviation.

TABLE 3 In comparison of the performance between DL model and TPUS-index model in predicting SUI in the training and testing sets.

Group	Diagnostic model	AUC (95% CI)	P-value	Accuracy	Sensitivity	Specificity	PPV	NPV
Training set	DL Model	0.798 (0.752, 0.845)	0.018	0.747	0.761	0.725	0.832	0.681
	TPUS-index Model	0.721 (0.666, 0.777)		0.730	0.631	0.806	0.711	0.742
Testing set	DL Model	0.869 (0.793, 0.945)	0.046	0.817	0.872	0.761	0.788	0.853
	TPUS-index Model	0.736 (0.629, 0.843)		0.772	0.675	0.755	0.675	0.775

DL, Deep Learning; TPUS, Transperineal Ultrasound; SUI, Stress Urinary Incontinence; AUC, area under the ROC curve; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

approach compared to the TPUS-index model (Table 3 and Supplementary Figures 1, 2).

Discussion

This study developed a DL model for the diagnosis of female SUI and compared its diagnostic performance with that of an ultrasound assessment model. Our findings demonstrate that the DL model outperformed the TPUS-index model in predicting the disease.

Our findings revealed that DL models trained on resting-state images exhibited poor diagnostic performance. In contrast, models trained on Valsalva-state images demonstrated significantly better performance in diagnosing SUI. The inferior performance of resting-state models can be attributed to the lower recognition rate of organs farther from the probe, such as the bladder and uterus, by CNN under resting conditions (9). Additionally, the morphology and function of the urethra change during the Valsalva maneuver. The limited anatomical changes observable under resting conditions provide insufficient diagnostic information for the model to learn effectively (17). This aligns with current clinical practice, where sonographer prioritize changes in the urethral angle and position in Valsalva-state images when assessing SUI through pelvic floor ultrasound (2). These findings indicate that DL models relying solely on resting TPUS images lack the reliability required for accurate diagnosis of female SUI.

Among the three DL models developed using Valsalva-state images, DenseNet-121 outperformed the ResNet models (ResNet-152 and ResNet-50), particularly in the testing set. This superior performance highlights DenseNet-121′s network architecture, which offers enhanced feature reuse and information flow through its dense connections, significantly reducing parameter redundancy. This optimization makes DenseNet-121 more efficient in terms of both parameter count and computational performance. The model exhibits strong classification capabilities in disease diagnosis through medical imaging (18–20). In the testing set, DenseNet-121 successfully identified four positive cases of SUI that were missed by ResNet-50 and ResNet-152. This superior sensitivity underscores DenseNet-121′s ability to capture subtle features in images. In contrast, the limitations of the ResNet models stem from their residual connections being confined to adjacent layers, which results in less effective feature reuse and greater computational resource requirements during training.

TPUS is a non-invasive and repeatable tool commonly used to assist in the assessment of female SUI. The SUI prediction models established based on TPUS ultrasound measurement data are also one of the current research hotspots. Liu and Quan (21) developed a postpartum SUI model using clinical data, bladder neck descent, and urethral funneling, achieving an AUC of 0.807 in the validation cohort. Another study on predicting SUI based on pelvic floor ultrasound data reported that combining multiple measurement parameters in the model achieved an AUC of 0.802, with sensitivity ranging from 0.542 to 0.665 and specificity from

0.867 to 0.980 (6). This study also demonstrated the predictive value of ultrasound parameters in diagnosing SUI. Indicators of Valsalva-State had a significant impact on the model's Rad-Score, consistent with previous research findings. These results further emphasize the critical role of Valsalva maneuver images in evaluating SUI. However, the limitations of such models lie in the need to collect clinical information or repeatedly measure multiple ultrasound parameters, which is time-consuming and highly dependent on the operator's technical expertise and dynamic observation skills.

This study found that the optimal DL model developed using TPUS images demonstrated higher diagnostic value for SUI compared to the TPUS-index model. The advantages of DL models lie in their ability to capture subtle imaging details that are challenging for sonographers while performing rapid automated analysis, reducing human errors and significantly shortening operational time. Notably, the optimal DL model in our study showed a significant advantage in sensitivity compared to the TPUS-index model. However, five positive cases in the testing set were misclassified by the DL model. Upon analysis, three cases involved uterine prolapse. Since uterine prolapse can cause structural changes in the adjacent urethra and posterior bladder wall, it may have affected the model's automatic segmentation of the urethra, thereby influencing the classification results. In addition, the DenseNet-121 model shows suboptimal performance with a specificity of 0.761 in ruling out non-SUI cases. Possible reasons include the limited number of non-SUI cases in the dataset, which may cause the model to be biased toward the majority class, resulting in insufficient ability to identify non-SUI cases. Additionally, non-SUI cases may exhibit diverse manifestations, making it difficult for the model to capture their complex features. To address these issues, improvements can be made by balancing the dataset and enhancing data diversity, or by fine-tuning the model or customizing layers to better capture the features of non-SUI cases. Despite these challenges, the DL model exhibited excellent performance in AUC and sensitivity for SUI diagnosis, making it a valuable tool for early detection. With an aging population, the rising incidence of SUI and its impact on quality of life have gained widespread attention. Early diagnosis and intervention are crucial in managing SUI. Accurate identification of early-stage SUI patients and the provision of timely treatments, such as lifestyle modifications, pelvic floor muscle therapy, and pharmacological interventions, can effectively slow disease progression. Compared to late-stage surgical treatments, these measures are more cost-effective, less invasive, and significantly improve patients' quality of life. Furthermore, the development of DL models facilitates large-scale screening for this condition in the general population.

Limitations of the study

First, the sample size included in this study was relatively small, necessitating further research with larger, more diverse prospective samples. Importantly, external validation in diverse populations, such as multi-center cohorts or across varying ultrasound devices, is currently absent and represents a critical limitation. Future

steps should include multi-center validation and testing across different imaging devices to ensure the model's adaptability and reliability in various clinical settings. Second, the developed DL model analyzed only ultrasound images and did not integrate clinical data. Therefore, in the next phase, we plan to combine clinical data with the DL model to design and construct a new hybrid model, further improving its diagnostic performance for female SUI.

Conclusion

In summary, this study developed a DL model for diagnosing female SUI, showing significant improvements in specificity, sensitivity, accuracy, and testing set consistency compared to the TPUS-index models, and its diagnostic performance was validated. It demonstrates the potential of DL models to enhance diagnostic accuracy and automation for female SUI.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

This retrospective study was approved by the Bioethics Committee of Xi 'an Jiaotong University Health Science Center [IRB number 2020823]. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

KC: Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. QC: Data curation, Methodology, Writing – review & editing. NN: Data curation, Formal analysis, Writing – review & editing. LS: Conceptualization, Data curation, Writing – review & editing. MM: Data curation, Methodology, Writing – review & editing. SY: Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

Research and Development Plan in the Shaanxi Province (No. 2024SF-YBXM-297).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1564446/full#supplementary-material

References

1. Schiffman M, Lamparello N. Stress incontinence in women. *N Engl J Med.* (2021) 385:e60. doi: 10.1056/NEJMc2112378

2. Falah-Hassani K, Reeves J, Shiri R, Hickling D, McLean L. The pathophysiology of stress urinary incontinence: a systematic review and meta-analysis. *Int Urogynecol J.* (2021) 32:501–52. doi: 10.1007/s00192-020-04622-9

3. Nambiar AK, Arlandis S, Bø K, Cobussen-Boekhorst H, Costantini E, de Heide M, et al. European association of urology guidelines on the diagnosis and management of female non-neurogenic lower urinary tract symptoms. Part 1: diagnostics, overactive bladder, stress urinary incontinence, and mixed urinary incontinence. *Eur Urol.* (2022) 82:49–59. doi: 10.1016/j.eururo.2022.01.045

4. Jefferson FA, Linder BJ. Evaluation and management of female stress urinary incontinence. *Mayo Clin Proc.* (2024) 11:1802–14. doi: 10.1016/j.mayocp.2024.07.003

5. Turkoglu A, Coskun ADE, Arinkan SA, Vural F. The role of transperineal ultrasound in the evaluation of stress urinary incontinence cases. *Int Braz J Urol.* (2022) 1:70–7. doi: 10.1590/s1677-5538.ibju.2020.1100

6. Xiao T, Chen Y, Gan Y, Xu J, Huang W, Zhang X. Can stress urinary incontinence be predicted by ultrasound? *AJR Am J Roentgenol.* (2019) 5:1163–9. doi: 10.2214/AJR.18.20893

7. Jamard E, Blouet M, Thubert T, Rejano-Campo M, Fauvet R, Pizzoferrato AC. Utility of 2D-ultrasound in pelvic floor muscle contraction and bladder neck mobility assessment in women with urinary incontinence. *J Gynecol Obstet Hum Reprod.* (2020) 1:101629. doi: 10.1016/j.jogoh.2019.101629

8. Brasil S, Pascoal C, Francisco R, Ferreira VDR, Videira PA, Valadão G. Artificial intelligence (AI) in rare diseases: is the future brighter? *Genes.* (2019) 10:978. doi: 10.3390/genes10120978

9. García-Mejido JA, Solis-Martín D, Martín-Morán M, Fernández-Conde C, Fernández-Palacín F, Sainz-Bueno JS. Applicability of deep learning to dynamically identify the different organs of the pelvic floor in the midsagittal plane. *Int Urogynecol J.* (2024) 35:2285–93. doi: 10.1007/s00192-024-05841-0

10. Huang X, Wang D, Li S, Yang L, Zhao J, Guo D. Advancements in artificial intelligence for pelvic floor ultrasound analysis. *Am J Transl Res.* (2024) 16:1037–43. doi: 10.62347/JXQQ5395

11. Wu S, Ren Y, Lin X, Huang Z, Zheng Z, Zhang X. Development and validation of a composite AI model for the diagnosis of levator ani muscle avulsion. *Eur Radiol.* (2022) 32:5898–906. doi: 10.1007/s00330-022-08754-y

12. Wei NY, Li XK, Lu XD, Liu XT, Sun RJ, Wang Y. Study on the consistency between automatic measurement based on convolutional neural network technology and manual visual evaluation in intracavitary ultrasonic cine of anterior pelvic. *J Ultrasound Med.* (2024) 43:671–81. doi: 10.1002/jum.16392

13. Choi EP, Lam CL, Chin WY. The test-retest reliability of the Incontinence Questionnaire-Urinary Incontinence Short Form (ICIQ-UI SF) for assessing type of urinary incontinence in males and females. *J Clin Nurs.* (2015) 24:3742–4. doi: 10.1111/jocn.12993

14. Nipa SI, Cooper D, Mostafa A, Hagen S, Abdel-Fattah M. Novel clinically meaningful scores for the ICIQ-UI-SF and ICIQ-FLUTS questionnaires in women with stress incontinence. *Int Urogynecol J.* (2023) 34:3033–40. doi: 10.1007/s00192-023-05657-4

15. Dietz HP. Pelvic floor ultrasound: a review. *Clin Obstet Gynecol.* (2017) 1:58–81. doi: 10.1097/GRF.0000000000000264

16. Zeng W. Image data augmentation techniques based on deep learning: a survey. *Math Biosci Eng.* (2024) 21:6190–224. doi: 10.3934/mbe.2024272

17. van Geelen H, Sand PK. The female urethra: urethral function throughout a woman's lifetime. *Int Urogynecol J.* (2023) 34:1175–86. doi: 10.1007/s00192-023-05469-6

18. Lu M, Zheng Y, Liu S, Zhang X, Lv J, Liu Y, et al. Deep learning model for automated diagnosis of moyamoya disease based on magnetic resonance angiography. *E Clin Med.* (2024) 77:102888. doi: 10.1016/j.eclinm.2024.102888

19. Alwakid G, Gouda W, Humayun M, Jhanjhi NZ. Deep learning-enhanced diabetic retinopathy image classification. *Digit Health.* (2023) 9:1–15. doi: 10.1177/20552076231194942

20. Tharmaseelan H, Vellala AK, Hertel A, Tollens F, Rotkopf LT, Rink J, et al. Tumor classification of gastrointestinal liver metastases using CT-based radiomics and deep learning. *Cancer Imaging.* (2023) 23:95. doi: 10.1186/s40644-023-00612-4

21. Liu W, Qian L. Establishment and validation of a risk prediction model for postpartum stress urinary incontinence based on pelvic floor ultrasound and clinical data. *Int Urogynecol J.* (2022) 33:3491–7. doi: 10.1007/s00192-022-05395-z

Check for updates

Machine learning-based fusion model for predicting HER2 expression in breast cancer by Sonazoid-enhanced ultrasound: a multicenter study

Huiting Zhang[1†], Manlin Lang[1†], Huiming Shen[2], Hang Li[3], Ning Yang[4], Bo Chen[5], Yixu Chen[6], Hong Ding[7], Weiping Yang[8], Xiaohui Ji[9], Ping Zhou[10], Ligang Cui[11], Jiandong Wang[12], Wentong Xu[12], Xiuqin Ye[13], Zhixing Liu[14], Yu Yang[15], Tianci Wei[16], Hui Wang[17], Yuanyuan Yan[18], Changjun Wu[19], Yiyun Wu[20], Jingwen Shi[21], Yaxi Wang[22], Xiuxia Fang[22], Ran Li[23], Ping Liang[1] and Jie Yu[1]*

[1]Department of Interventional Ultrasound, PLA Medical College & Chinese PLA General Hospital, Beijing, China, [2]Department of Ultrasound, Zhongda Hospital, Nanjing, China, [3]Department of Breast Surgery, Affiliated Hospital of Putian University, Putian, China, [4]Department of Ultrasound, Xingcheng People's Hospital, Xingcheng, China, [5]Department of Ultrasound Medicine, Lu' an People's Hospital of Anhui Province, Liuan, China, [6]Department of Ultrasound, The Fifth People's Hospital of Chengdu, Chengdu, China, [7]Department of Ultrasound, Huashan Hospital, Shanghai, China, [8]Department of Ultrasound, Guangxi Medical University Cancer Hospital, Nanning, China, [9]Department of Ultrasound, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China, [10]Department of Ultrasound, The Third Xiangya Hospital, Changsha, China, [11]Department of Ultrasound, Peking University Third Hospital, Beijing, China, [12]General Surgery, Chinese PLA General Hospital, Beijing, China, [13]Department of Ultrasound, First Affiliated Hospital of Southern University of Science and Technology, Second Clinical College of Jinan University, Shenzhen Medical Ultrasound Engineering Center, Shenzhen People's Hospital, Shenzhen, China, [14]Department of Ultrasound Medicine, The First Affiliated Hospital of Nanchang University, Nanchang, China, [15]Department of Ultrasound, Beijing Friendship Hospital, Beijing, China, [16]Department of Ultrasound, The 2nd Affiliated Hospital of Harbin, Harbin, China, [17]Department of Ultrasound, China-Japan Union Hospital of Jilin University, Changchun, China, [18]Department of Ultrasound, Zhengzhou Central Hospital, Zhengzhou, China, [19]Department of Ultrasound, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [20]Department of Ultrasound, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing, China, [21]Department of Ultrasound, Shengjing Hospital of China Medical University, Shenyang, China, [22]Department of Ultrasound, The Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China, [23]Department of Ultrasound, The First Affiliated Hospital of Xinxiang Medical University, Xinjiang, China

Purpose: To predict human epidermal growth factor receptor 2 (HER2) expression in breast cancer (BC) using Sonazoid-enhanced ultrasound in a machine learning-based model.

Materials and methods: Between August 2020 and February 2021, patients with breast cancer who underwent surgical treatment without neoadjuvant chemotherapy were prospectively enrolled from 17 hospitals in China. HER2 expression status was assessed by immunohistochemistry or fluorescence *in situ* hybridization (FISH). The training set contained data from 11 hospitals and the validation set contained 6 hospitals. Clinical features, B-mode ultrasound, contrast-enhanced ultrasound (CEUS), and time-intensity curve were selected by the Least Absolute Shrinkage and Selection Operator. Based on the selected features, six prediction models were established to predict HER2 3 + and 2 +/1 + expression: logistic regression (LR), support vector machine (SVM),

random forest (RF), eXtreme Gradient Boosting (XGB), XGB combined with LR, and fusion model.

Results: A total of 140 patients with breast cancer were enrolled in this study. Seven features related to HER2 3 + and six features related to HER2 2+/1 + were selected to establish prediction models. Among the six models, LR, SVM, and XGB showed the best prediction performance for both HER2 3 + and HER2 2+/1 + cases. These three models were then combined into a fusion model. In the validation, the fusion model achieved the highest value of area under the receiver operating characteristic curve as 0.869 (95%CI: 0.715−0.958) for predicting HER2 3 + and 0.747 (95%CI: 0.548−0.891) for predicting HER2 2+/1 + cases. The model could correctly upgrade HER2 2 + cases to HER2 3 + cases, consistent with the FISH test results.

Conclusion: Sonazoid-enhanced ultrasound can provide effective guidance for targeted therapy of breast cancer by predicting HER2 expression using machine learning approaches.

1 Introduction

According to the World Health Organization, breast cancer (BC) can cause 500,000 deaths, and 1.7 million new cases are diagnosed annually (1). Characterized by overexpression of the human epidermal growth factor receptor 2 (HER2) gene and its protein, HER2-positive breast cancer accounts for 20–30% of breast cancer cases and requires distinct therapeutic strategies (2, 3). Trastuzumab and pertuzumab, which are targeted by monoclonal antibody therapies, improve the survival outcomes of HER2-positive (HER2 3+) breast cancer (4–6). Recent reports have recommended HER2-targeted agents and antibody-drug conjugates (ADCs) as new clinical therapies for HER2-low expression (HER2 1+, 2+) breast cancer (7). The distinct pathological characteristics of HER2 0, HER2-low, and HER2-positive breast cancers have been the focus of research. Studies have reported that the 50% recurrence rate of HER2-positive breast cancers can be decreased by the use of HER2-targeted monoclonal antibodies (8).

For patients with HER2-positive cancers, preoperative targeted therapy could increase the chance of breast conservation and sentinel lymph node biopsy rather than mastectomy and axillary lymph node dissection (7, 9). The selection of breast cancer neoadjuvant treatment regimens (particularly monoclonal antibodies) depends on the results of preoperative core needle biopsy (CNB), especially molecular profiling tested by immunohistochemistry (IHC) and fluorescence *in situ* hybridization (FISH) (10–12). However, because of intratumoral heterogeneity, the inadequate tissue acquired from CNB may not provide complete pathological characteristics of the tumor, causing discordance between cores in 8% of HER2-positive cases and discordance between CNB and surgical pathology results for approximately 26.6% of HER2 status (11, 13, 14). Thus, HER2 expression levels in breast cancer could be underestimated, and the concomitant false-negative results may cause missed diagnosis of HER2-positive cases, affecting clinical arrangements and prognosis. Increasing the number of multi-point punctures may increase the accuracy or decrease the underestimation in the diagnosis of HER2 expression. However, it has been reported that the possibility of core

needle seeding in breast cancer varies from 2 to 63% (15–17). Adding the number of punctures to increase the amount of tissue may also increase the risk of tumor seeding (16, 17).

Contrast-enhanced ultrasound (CEUS) indicates vascular information of the tumor, which has been widely used in the diagnosis of benign and malignant breast lesions, assessing the pathological characteristics, and predicting neoadjuvant chemotherapy (NAC) response (18, 19). CEUS can improve the categorization of suspicious breast lesions, reduce unnecessary biopsies, and improve the cancer yield rate of biopsy procedures (20). SonoVue (Bracco, Milan, Italy), the most widely used ultrasound contrast agent, consisting of sulfur hexafluoride microbubbles, has shown better performance in low-intensity imaging (21). Consisting of lipid-stabilized perfluorocarbon microbubbles, Sonazoid (GE Healthcare, Oslo, Norway) is more stable for long-term imaging and has a higher resistance to ultrasound mechanical index (MI), which is more suitable for high-frequency linear array probe scanning (22, 23). Machine learning approaches have been widely applied for the early detection, diagnosis, and outcome prediction of breast cancer (24, 25). It has been reported that the diagnostic accuracy and sensitivity of CEUS in breast cancer can be improved by combining it with a machine learning approach (20).

Hence, our study aims to predict the HER2 status of breast cancer by combining B-mode ultrasound and contrast Sonazoid-enhanced ultrasound features using machine learning models.
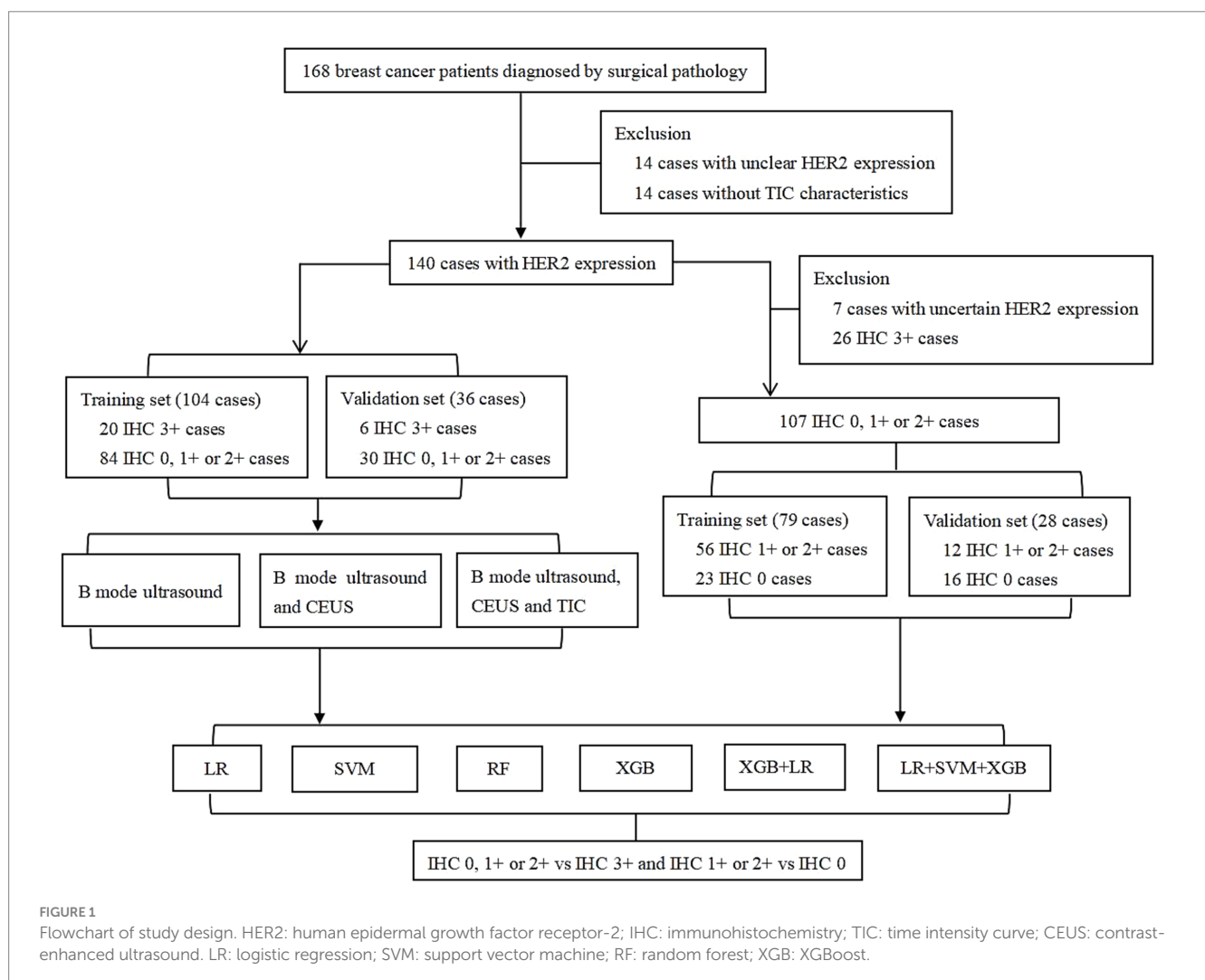
2 Materials and methods

2.1 Patients

This prospective, multicenter study was approved by the institutional ethics committee (ClinicalTrials.gov: NCT04657328). Informed written consent was obtained from all participants before the examinations. Between August 2020 and February 2021, 168 patients with breast cancer with 168 breast masses diagnosed by

surgical pathology from a multicenter cohort of 17 hospitals in China, were enrolled in this study. Patients with an unclear HER2 status and incomplete time-intensity curve (TIC) features were excluded. According to current guidelines (26), HER2 status was determined using IHC for HER2 protein expression and FISH for equivocal cases (IHC 2+). The multicenter IHC results for HER2 expression were evaluated by experienced pathologists. A total of 140 patients with HER2 status were included in the study. The exclusion criteria were (1) absence of HER2 results and (2) absence of TIC features due to substandard image acquisition. The training set contained datasets from 11 hospitals, including 104 and 79 cases in the two cohorts. The external validation set contained prospective datasets from 6 other hospitals, including 36 and 28 cases in the two cohorts. Among these cases, 104 patients from 11 hospitals were included in the training set and 36 patients from the other 6 hospitals were included in the validation set. In total, there were 26 HER2-positive (IHC 3+), 68 HER2-low (39 IHC 1 + and 29 IHC 2+), 39 HER2 0 (IHC 0), and 7 HER2-negative (IHC 0, 1+, and 2+) cases in the training and validation sets. In total, 88 patients with invasive ductal carcinoma, 3 with mucinous breast carcinoma, 1 with metaplastic breast carcinoma, and 12 with ductal carcinoma *in situ* were included.

Furthermore, to differentiate HER2-low expression cases from HER2 0 and exclude the confounding effect of HER2-positive expression levels in the analysis, 26 HER2-positive cases and 7 patients with uncertain HER2 expression status (only known as HER2-negative cases) in the cohort were excluded. Finally, 107 patients were included in the HER2-negative and low-expression group, containing 79 patients in the training cohort from the same 11 hospitals and 28 patients in the validation cohort. The study design is shown in Figure 1.

2.2 B-mode and CEUS image acquisition

B-mode ultrasound and CEUS examinations were performed by radiologists from 17 hospitals with 10 ultrasound devices (Supplementary Table 1) equipped with a linear probe. All ultrasound examinations were conducted following a uniform diagnostic consensus. Prior to image acquisition, participating radiologists in this multicenter study, with more than 3 years of experience in breast ultrasound, received systematic training in B-mode and CEUS breast examination. All radiologists in this study received standardized training in breast CEUS interpretation according to Sonazoid instructions and previous

FIGURE 1
Flowchart of study design. HER2: human epidermal growth factor receptor-2; IHC: immunohistochemistry; TIC: time intensity curve; CEUS: contrast-enhanced ultrasound. LR: logistic regression; SVM: support vector machine; RF: random forest; XGB: XGBoost.

studies. They were required to complete a minimum of 50 breast CEUS-independent case evaluations to ensure consistent diagnostic consensus prior to the study. Breast masses were first identified using a B-mode ultrasound scan. Next, 0.015 mL/kg of perfluorobutane-filled microbubble contrast agent (Sonazoid; GE Healthcare, Oslo, Norway) was injected via the catheter line (≥ 22-gauge) placed in the antecubital vein, followed by a 5 mL flush of 0.9% sodium chloride solution. The mechanical index of 0.18–0.24 was applied. When the injection was completed, the imaging timer was started simultaneously. After 1 min of continuous assessment of the whole mass, intermittent scanning (10 s each time) was arranged at the time points of 1.5 min, 2 min, 3 min, 4 min, and 5 min. For patients with multiple masses, images of the largest masses were preserved. Both B-mode and CEUS images and videos were stored in DICOM format on a hard disk at the hospital and sent to our study center. Finally, six radiologists with more than 15 years of experience in conventional breast ultrasound and breast CEUS were independently evaluated for image features at the study center (Figure 2).

In B-mode breast ultrasound, the "strip-shaped echoic" feature represents thin, elongated, and hyperechoic lines or bands within the breast tissue or mass. CEUS characteristics were evaluated, including shape (regular or not), margin (well or poorly defined), wash-in time (earlier, later, synchronous), enhancement degree (hyperenhancement, isoenhancement, hypoenhancement), complete wash-out time of lesions (≤5 min or not), uptake pattern (centripetal, centrifugal, diffuse, no enhancement), as well as exhibitions of the homogeneous pattern, rim-like enhancement, claw-shaped pattern, perfusion defects, lesion size compared with conventional ultrasound increased, and nourishing vessels. The time-intensity curve (TIC) features were evaluated using external perfusion software (VueBox™) to quantitatively evaluate the microvasculature of the tumors through the CEUS videos.

2.3 Statistical analysis

R version 3.4.4 software, SPSS Version 23.0 (IBM, Armonk, NY, United States), and MedCalc 19.5.6 were used to perform statistical analysis. Statistics are described as mean ± standard deviation or numbers with percentiles for distribution. The t-test, chi-square test, and the Least Absolute Shrinkage and Selection Operator (LASSO) were used to select the features. The regularization property of LASSO constrains the model coefficients through the penalty parameter (λ) and shrinks the coefficients of less important variables to zero to mitigate overfitting (27, 28). Logistic regression (LR), support vector machine (SVM), random forest (RF), eXtreme Gradient Boosting (XGB), late fusion model based on the voting method, and XGB combined with LR were trained to classify HER2-positive status and HER2 low expression status in the two groups. A combination of XGB (constructing new features based on existing features) and LR (classifiers) was used to establish the prediction model. Prediction models were established on the training set, and their performance was tested on the validation set (29). For internal validation, leave-one-out cross-validation (LOOCV) was performed to assess the predictive accuracy and stability of the training set. External validation was performed to test the performance of the trained models, evaluate their generalizability, and identify potential biases. The receiver operating characteristic curves (ROC) of the predictive models were analyzed. The area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, and 95%CI were assessed. The DeLong test was used to compare differences between the AUC values of the different models.

3 Results

3.1 Clinical characteristics

The clinical characteristics of 140 patients with breast cancer (mean age 52.35 ± 11.03 years, range 23–85 years) with 140 masses are shown in Table 1. In the training cohort, 104 patients were enrolled, including 20 HER2-positive cases. Of the 107 patients in the HER2 low expression group, 79 were included in the training cohort, with 56 IHC 2 + or 1 + and 23 IHC 0 cases.
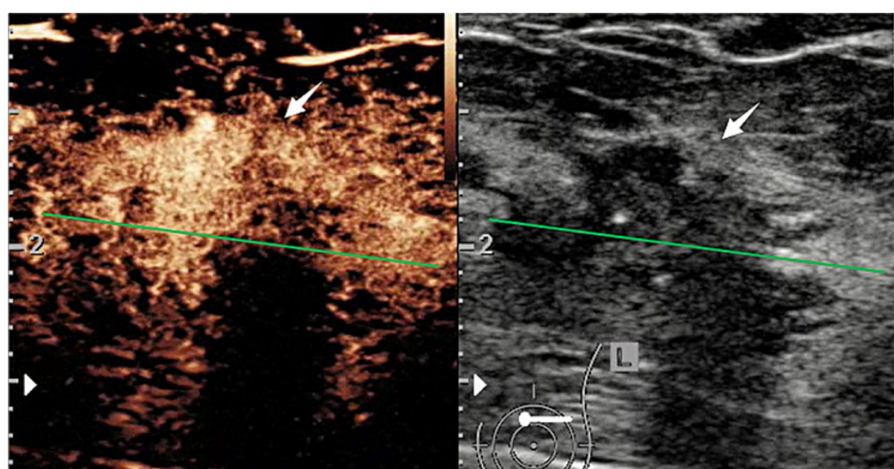

FIGURE 2
CEUS and B-mode ultrasound images of a patient with HER2-positive breast cancer.

Clinical characteristics of 140 patients in training and validation sets.

	Total ($n = 140$)	Training set ($n = 104$) (%)	Validation set ($n = 36$) (%)	p value
Age (years)	52.35 ± 1.03 (23–85)	53.13 ± 11.07	50.11 ± 10.75	0.158
BMI (kg/m²)	24.20 ± 5.04 (13.65–63.70)	24.46 ± 5.30	23.46 ± 4.18	0.308
Menopause				0.608
Premenopause	61	44 (42.3)	17 (47.2)	
Postmenopause	79	60 (57.7)	19 (52.8)	
Family history of breast cancer				1.000
No	133	99 (95.2)	34 (94.4)	
Yes	7	5 (4.8)	2 (5.6)	

BMI: body mass index.

3.2 B-mode and CEUS characteristics

In 140 patients with breast cancer, the image features of B-mode ultrasound, CEUS, and TIC were assessed (Supplementary Table 2). According to the LASSO regression in clinical B-mode with CEUS and TIC of CEUS characteristic groups, seven features related to HER-2 positive breast cancer, including tumor size (cm), echotexture, strip-shaped echoic, macrocalcifications, microcalcifications, perfusion defects, and fall time (FT) of TIC, were selected (Figure 3). No clinical characteristics were observed. The distribution of the selected characteristics is listed in Table 2.

Characteristics of B-mode imaging with CEUS.

$$0.19952497 + \left[0.01610785 * \text{Tumor size} \right]$$
$$+ \left[0.01815141 * \text{Echotexture} \right]$$
$$+ \left[-0.07019562 * \text{Strip} - \text{shaped echoic} \right]$$
$$+ \left[-0.01206527 * \text{Macrocalcifications} \right]$$
$$+ \left[0.04789061 * \text{Microcalcifications} \right]$$
$$+ \left[-0.01622952 * \text{Perfusion defect} \right].$$

Characteristics of CEUS TIC.

$$-1.21674647 + \left[-0.01299429 * \text{FT} \right].$$

In 107 cases in the HER2 low expression group, the image features of the three modalities were assessed in Supplementary Table 3. Imaging features related to HER2 low expression were selected by LASSO regression, including location, shape, strip-shaped echoic, perfusion defect, mean transit time (mTT), and FT. There were no clinical characteristics observed. The selected characteristics are listed in Supplementary Table 4.

Characteristics in B-mode.

$$0.13508836 + \left[0.16800326 * \text{Location} \right] + \left[0.31750333 * \text{Shape} \right]$$
$$+ \left[0.02871666 * \text{Strip} - \text{shaped echoic} \right].$$

Characteristics of CEUS images.

$$1.3810282 + \left[-0.6861225 * \text{Perfusion defect} \right].$$

Characteristics in TIC of CEUS.

$$1.2891179887 + \left[-0.0002615468 * \text{mTT} \right] + \left[-0.0197138051 * \text{FT} \right].$$

3.3 Machine learning models for the prediction

The prediction model was established on the training set, and its performance was tested on the validation set. The effectiveness and stability of the training set, consisting of 104 cases, were validated using LOOCV, and the accuracy and Kappa were 0.871 and 0.446, respectively. In the training set of FISH positive (IHC 3+) and negative groups, six classifiers, including logistic regression (LR), support vector machine (SVM), random forest (RF), XGB (XGBoost), decision-level fusion technique of hard voting based on LR, SVM, and XGB, as well as the XGB combined with the LR model (29, 30).

The final result of the decision-level fusion model was determined by three single classifiers: LR, SVM, and XGB (better than RF in this study). The hard-voting progression is shown in Figure 4. In the XGB combined with LR prediction model, XGB was used to construct new variables, reflecting the correlation of the selected variables. LR was used to gather the selected and new variables to construct the prediction model and to calculate the significance and weight coefficients of each variable. In the prediction of the HER2-positive breast cancer group, seven variables, including a novel feature (V11) generated by the XGB tree-based model trained on existing features, were selected for the final LR prediction model based on the feature importance rankings (Supplementary Figure 1).

Classifiers of LR, SVM, RF, and XGB were established in three imaging modalities: (1) B-mode ultrasound, (2) B-mode ultrasound combined with CEUS, and (3) B-mode ultrasound combined with CEUS and TIC. The other two types of fusion models were used in the third multi-modality to predict HER2-positive breast cancer.

The AUC, sensitivity, specificity, and accuracy of the four classifiers in three modalities are shown in Table 3. The sensitivities of SVM were increased from 0.728 (95%CI: 0.554–0.862) to 0.778 (95%CI: 0.608–0.899) by adding the CEUS modality. In the three modalities group, SVM performs the best AUC value in the four single classifiers, with an AUC of 0.806 (95%CI: 0.640–0.918), a sensitivity of 0.833 (95%CI: 0.359–0.996) and a specificity of 0.767 (95%CI: 0.577–0.901). The AUC values improved with the enrichment of the imaging modalities. In the third modality, the performances of the other two fusion models are also shown in Table 3.

According to the predictive performance of LR, SVM, RF, and XGB, the three top-performing individual classifiers for HER2 expression, LR, SVM, and XGB, were combined using hard voting to
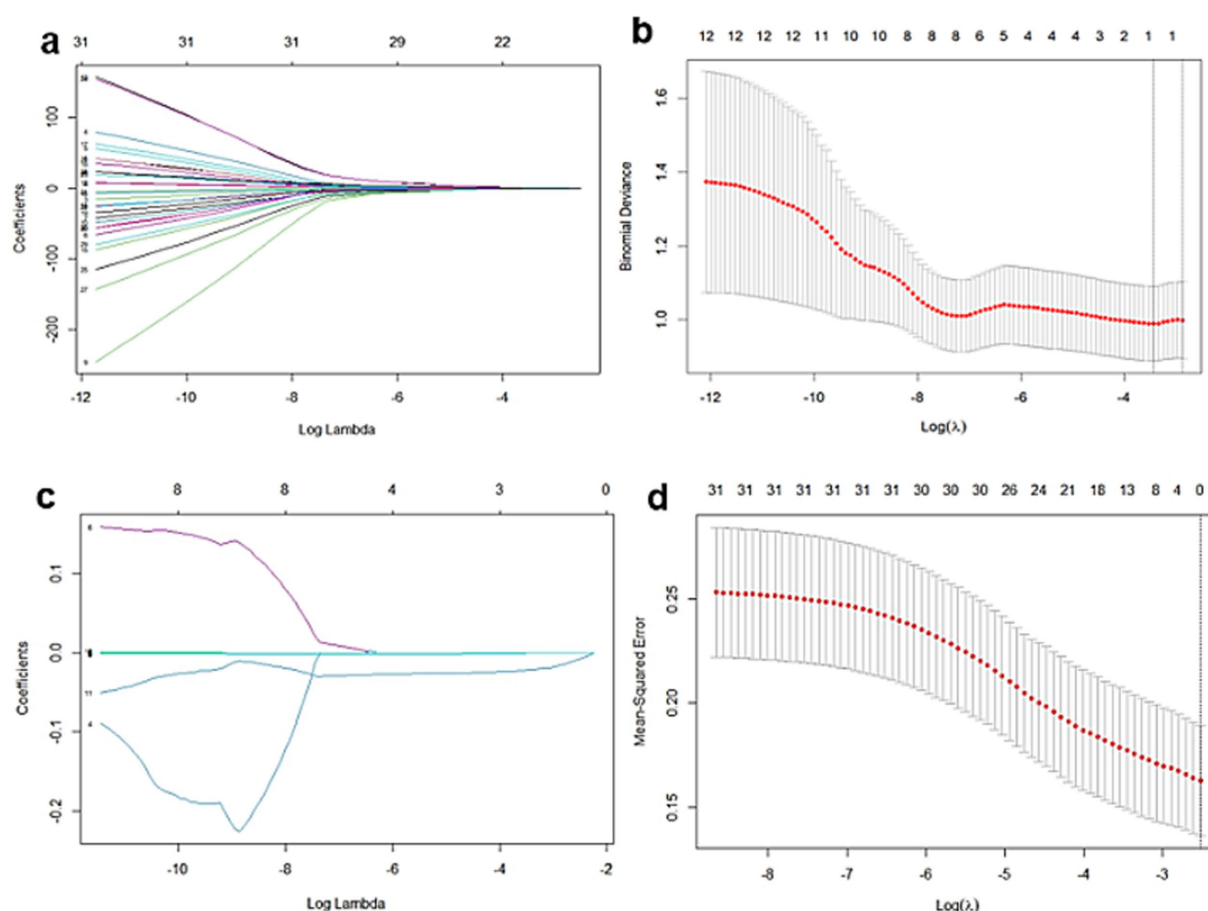
FIGURE 3
Feature selection in B-mode, CEUS, and TIC of the CEUS group by LASSO regression in 140 patients with breast cancer. **(a,b)** Selection of B-mode ultrasound and CEUS features. **(c,d)** Selection of TIC parameters.

generate a consolidated prediction result. Thus, the decision-level fusion model was constructed using hard voting based on LR, SVM, and XGB to establish the fusion model, and the weighted ratio was set as 1:1:1. In the six models, the fusion model of LR, SVM, and XGB classifiers performed best, with an AUC value of 0.869 (95%CI: 0.715–0.958), a sensitivity of 1.000 (95%CI: 0.541–1.000), and a specificity of 0.668 (95%CI: 0.472–0.827). The ROCs of the six classifiers in B-mode ultrasound combined with CEUS and TIC modalities are shown in Figure 5. In the training cohort of 104 cases, 31 cases with certain IHC results were assessed as IHC 2 + by CNB, and two of them were reclassified as IHC 3 + according to FISH results. The fusion model of LR, SVM, and XGB also predicted them as IHC 3 + cases.

In the training set of the HER2 low expression and HER2-negative groups, prediction models based on the six classifiers in the third modality were also established. In the training set of 79 participants, the accuracy and kappa values were 0.864 and 0.637, respectively. The AUC values, sensitivity, specificity, and accuracy are shown in Table 4. The decision-level fusion model was selected as the voting result of LR, SVM, and XGB, and the weighted ratio was set at 1:2:1, according to the performance of the classifiers. The fusion model of LR, SVM, RF, and XGB classifiers also gets the highest AUC value of 0.747 (95%CI: 0.548–0.891), sensitivity of 1.000 (95%CI: 0.735–1.000), and specificity of 0.438 (95%CI: 0.198–0.701). The ROCs of the six

prediction models in the HER2 low expression and negative group are shown in Figure 6. Both the AUCs for predicting HER2 status were increased using the decision-level machine learning approach.

4 Discussion

### 4.1 Key findings in the context of prior literature

HER2-targeted therapy can reduce recurrence and increase the likelihood of breast-conserving surgery in patients with HER2-positive breast cancer. In this study, the fusion model of multiple single classifiers, based on machine learning approaches, performed best in predicting HER2 3 + and HER2 2+/1 + expression, with an AUC of 0.869 (95%CI: 0.715–0.958) and 0.747 (95%CI: 0.548–0.891), respectively. It could also predict the two equivocal IHC 2 + breast cancers as HER2 3+, in concordance with the FISH results.

In this research, imaging features of multimodalities, including B-mode ultrasound, CEUS, and TIC, were obtained by assessment of radiologists. Previous studies that predicted HER2 expression using imaging features are shown in Supplementary Table 5. Compared with radiomic features acquired by software or a single ultrasound

TABLE 2 Selected features of 140 patients in training and validation sets.

	Validation set (n = 36) (%)	Training set (n = 104) (%)	Total (n = 140)
Tumor size (cm)	2.45 ± 1.18	2.11 ± 1.05	2.20 ± 1.09 (0.5–5.9)
Echotexture			
Homogeneous	11 (30.6)	26 (25.0)	37
Heterogeneous	25 (69.4)	78 (75.0)	103
Strip-shaped echoic			
Absence	11 (30.6)	32 (30.8)	43
Present	25 (69.4)	72 (69.2)	97
Macrocalcifications			
Absence	7 (19.4)	14 (13.5)	21
Present	29 (80.6)	90 (86.5)	119
Microcalcifications			
Absence	31 (86.1)	71 (68.3)	102
Present	5 (13.9)	33 (31.7)	38
Perfusion defects			
Presence	13 (36.1)	34 (32.7)	47
Absence	23 (63.9)	70 (67.3)	93
FT (s)	17.45 ± 18.46	17.49 ± 13.33	-

CEUS: contrast-enhanced ultrasound; TIC: time intensity curve; FT: fall time.

FIGURE 4
Hard voting progression of the decision-level fusion model.
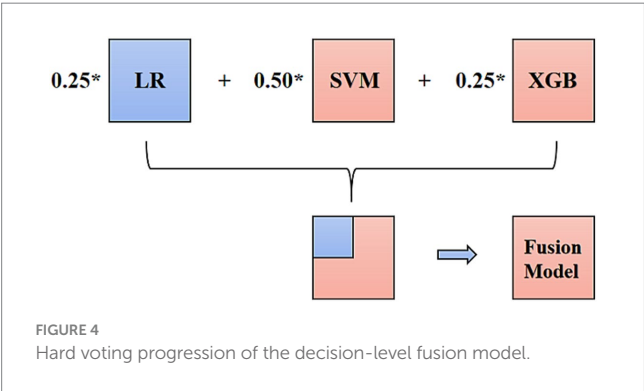
modality, these features are more available and can provide abundant vascularity information. Vasculogenic mimicry (VM), which differs from angiogenesis formed by endothelial cells, is a vascular structure formed by cancer cells that transit tumor and blood cells in a channel network and is involved in tumor neovascularization (31, 32). In breast cancer, VM is associated with HER2-positive cases, which may contribute to two anticoagulant-secreted proteins, Serpine2 and Slpi, promoting VM formation. Both of them mostly occurred in HER2-positive patients with breast cancer (33, 34). Studies have shown that CEUS can assess VM density *in vitro*, and the quantitative parameters of TIC are related to VM (35, 36). Thus, the microbubbles of CEUS may provide information on HER2-positive breast cancer neovascularization at the molecular level.

Previous studies have mostly focused on HER2 3 + expression in breast cancers using radiomic approaches. To the best of our

knowledge, this is the first study to use LR, SVM, and XGB fusion models by voting decision method to prospectively predict HER2 3 + and 2+/1 + expression levels in breast cancer based on a multicenter study of contrast Sonazoid-enhanced ultrasonography. In predicting HER2-positive and HER2-low expression BC cases, the AUC values of the fusion model in both of the two groups were the highest compared with other single machine learning models.

4.2 Clinical implications and innovations

In this study, tumor size, echotexture, strip-shaped echoic, macrocalcifications, and microcalcifications on B-mode ultrasound, perfusion defects on CEUS, and FT of TIC were predictive factors of HER2-positive breast cancer. Factors including tumor location, shape, strip-shaped echoic in B-mode ultrasound, perfusion defect in CEUS, mTT, and FT of TIC could predict HER2 low expression. Strip-shaped echogenic perfusion defects and FT are also predictors of HER2-positive expression, indicating that these features may be closely related to HER2 protein expression levels (2, 37).

Tumor size may reflect growth, indicating the prognosis of malignant tumors. Features of macrocalcifications and microcalcifications on B-mode ultrasound were associated with HER2-positive breast cancer in this study, which was also consistent with previous studies (38–41). Macrocalcification is regarded as the degeneration of the breast caused by injury and inflammation unrelated to cancer, while microcalcification is regarded as a calcium spot caused by rapid decomposition of cancer cells (38). With high aggressiveness and poor prognosis, HER2-positive breast cancer may be related to a faster growth rate than negative cases, indicating that more cell decomposition of the breast exists in positive cases (42, 43).

A strip-shaped echo mostly indicates the fibrosis inside the tumor. Malignant lesions can exhibit disordered hyperechoic strands, whereas benign lesions tend to exhibit organized linear echoes. Fibrosis in breast tumors is histologically regarded as fibroblasts and collagen fibers in the tumor center (44). Some studies have reported that fibrosis is positively related to HER2 expression and high aggressiveness of tumors (45), which is in contrast to the results of this study. In this study, fewer strip-shaped echoes were observed in HER2-positive breast cancer. A possible reason may be that most of our breast cancer cases were in stage I or II (100/104), and tumor cells were in the rapid growth phase, without undergoing necrosis and fibrosis. Further studies are still needed to determine the relationship between strip-shaped echoes and HER2 expression (45, 46).

Previous studies have also revealed that high HER2 expression might be related to the increased invasiveness of tumor cells and the formation of neovasculature (47). In some studies, perfusion defects in CEUS more frequently occurred in HER2-positive breast cancer, which might be caused by ischemic necrosis of the tumor, contributing to the slower blood vessel growth rate than the increased oxygen consumption of the tumor cells (48, 49). Other studies have also revealed that perfusion defects might be associated with uneven distribution of the contrast agent caused by heterogeneity and blood vessel distribution inside the tumor (47, 50). However, in this study, perfusion defects in Sonazoid-based CEUS were negatively associated with HER2-positive and low-expression breast cancers. In HER2 expression cases, less fibrosis was observed, indicating the presence of abundant vascularity, compared with HER2-negative cases.

TABLE 3 Diagnostic performance of the classifiers in predicting HER2-positive patients.

	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	Accuracy
B-mode ultrasound				
RF	0.567 (0.392–0.730)	0.167 (0.004–0.641)	0.967 (0.828–0.999)	0.833
B-mode ultrasound and CEUS				
SVM	0.778 (0.608–0.899)	0.667 (0.223–0.957)	0.867 (0.693–0.962)	0.833
RF	0.583 (0.408–0.745)	0.167 (0.004–0.641)	1.000 (0.884–1.000)	0.861
B-mode ultrasound, CEUS, and TIC				
LR	0.633 (0.457–0.787)	0.667 (0.223–0.957)	0.767 (0.577–0.901)	0.722
SVM	0.806 (0.640–0.918)	0.833 (0.359–0.996)	0.767 (0.577–0.901)	0.778
RF	0.583 (0.408–0.745)	0.167 (0.400–0.641)	1.000 (0.884–1.000)	0.861
XGB	0.700 (0.525–0.841)	0.500 (0.118–0.882)	0.900 (0.735–0.979)	0.833
XGB + LR	0.689 (0.513–0.832)	0.668 (0.223–0.957)	0.633 (0.439–0.801)	0.639
LR + SVM + XGB	0.869 (0.715–0.958)	1.000 (0.541–1.000)	0.668 (0.472–0.827)	0.722

CEUS: contrast-enhanced ultrasound; TIC: time intensity curve; LR: logistic regression; SVM: support vector machine; RF: random forest; XGB: XGBoost.

FIGURE 5
ROCs of the classifiers in predicting HER2-positive breast cancer based on B-mode ultrasound, CEUS, and TIC in the **(a)** training and **(b)** validation sets.
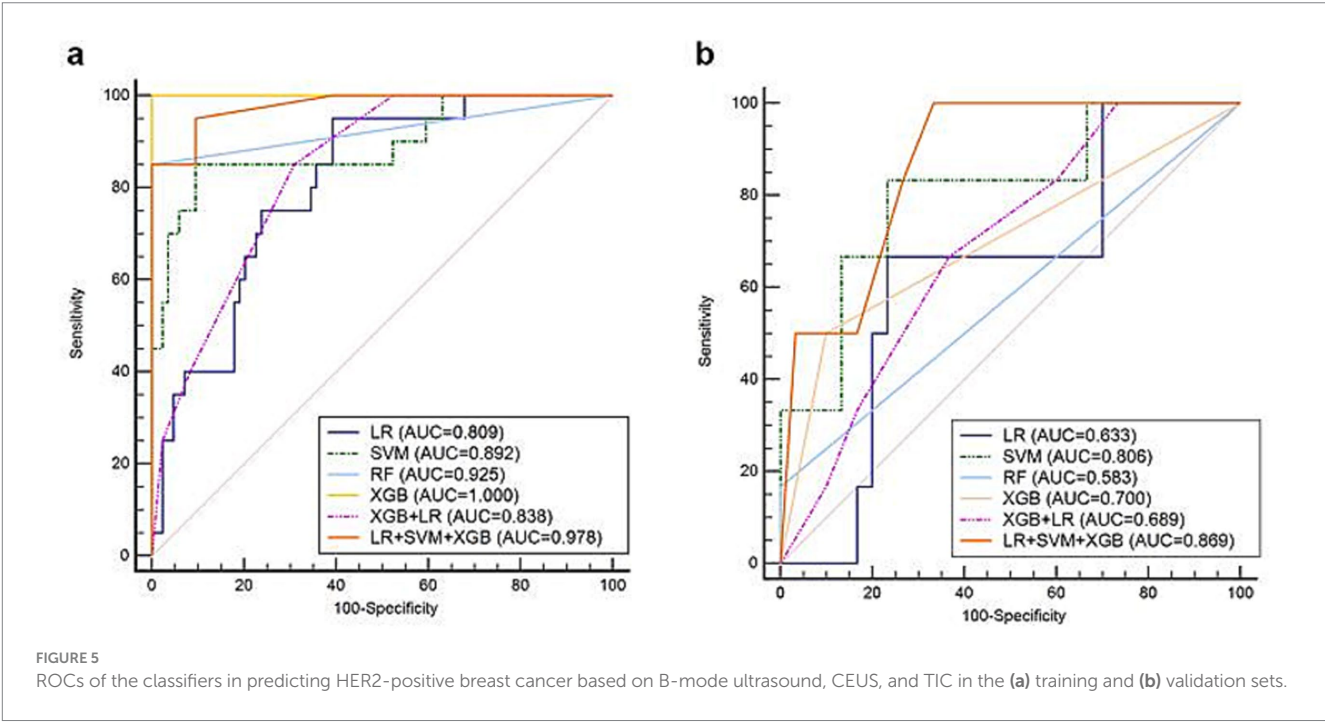
TABLE 4 Diagnostic performance of the classifiers in predicting HER2 low expression patients based on B-mode ultrasound, CEUS, and TIC characteristics.

Models	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	Accuracy
LR	0.698 (0.496–0.856)	0.750 (0.428–0.945)	0.625 (0.354–0.848)	0.679
SVM	0.687 (0.486–0.848)	0.667 (0.349–0.901)	0.813 (0.544–0.960)	0.750
RF	0.615 (0.413–0.791)	0.917 (0.615–0.998)	0.313 (0.110–0.587)	0.571
XGB	0.625 (0.423–0.799)	0.750 (0.428–0. 945)	0.500 (0.247–0.753)	0.607
XGB + LR	0.654 (0.451–0.822)	0.917 (0.615–0.998)	0.313 (0.110–0.587)	0.571
LR + SVM + XGB	0.747 (0.548–0.891)	0.917 (0.615–0.998)	0.500 (0.247–0.753)	0.679

AUC, area under curve; CI, confidence interval; LR: logistic regression; SVM: support vector machine; RF: random forest; XGB: XGBoost.
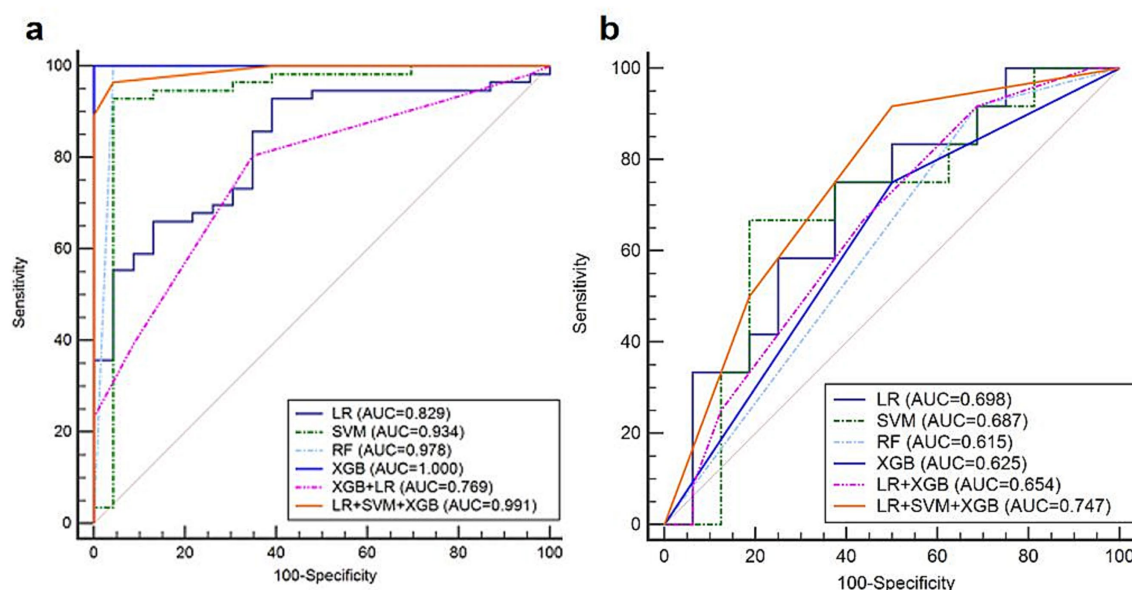
FIGURE 6
The ROCs of the six prediction models in predicting HER2 low expression patients in the **(a)** training and **(b)** validation sets.

In SVM models of three modalities, the sensitivities in predicting HER2-positive breast cancer were increased by CEUS, from 0.728 (95%CI: 0.554–0.862) to 0.778 (95%CI: 0.608–0.899). By adding the TIC feature, the sensitivity could also be increased, up to 0.806 (95%CI: 0.640–0.918). This result may indicate that the evaluation of microvasculature could improve the performance of prediction models in HER2-positive breast cancer, especially for the evaluation of TIC features. In previous studies of Sonazoid-based CEUS in liver cancer, short mTT and FT could be factors that differentiate angiomyolipoma and hepatocellular carcinoma from hepatocellular carcinoma because of the different amounts of blood vessels (51). In this study, short FT may be associated with HER2 expression (IHC3+, 2+, and 1+) in breast cancer, compared with HER2-negative expression cases. This may be related to the rapid excretion rate of Sonazoid microbubbles from intratumoral vessels in HER2 expression breast lesions. FT may be related to the number of blood vessels inside tumors because abundant vessels may contribute to a fast blood flow discharging from the draining vein and a short contrast agent staying time. Therefore, HER2-expressing breast tumors tend to exhibit higher internal vascularity.

In the 104 cases of patients with breast cancer, there were a total of 31 cases defined as IHC 2 + for the first time of CNB, with certain results of biopsy. Two of these were finally defined as IHC 3 + according to the FISH results, revealing that 6.5% (2/31) of HER2-positive cases were underestimated by IHC in this study. In the prediction results of the LR + SVM + XGB fusion model, the two cases were also predicted as IHC 3+, indicating that the fusion predictive model could improve the detection of IHC 3 + compared with the results of CNB by pathologists.

4.3 Limitations and future directions

Our study used LR, SVM, and XGB decision-level fusion models to predict three HER2 expression levels in breast cancer in two cohorts based on a prospective multicenter study of contrast Sonazoid-enhanced and B-mode ultrasound. However, this study has some limitations. First, the number of cases was limited because of the use of Sonazoid in breast CEUS multicenter studies. Second, this study only contained image features evaluated by radiologists. Radiomic features can reflect unrecognizable and quantifiable messages to the naked eye. Using radiomic approaches in multi-modal ultrasound may improve the prediction of BC biomarkers. However, radiomic features extracted by software were less available compared with the features assessed by radiologists in this study. Third, our study only included images from B-mode ultrasound and CEUS. Additional modalities, such as MRI and mammography, are expected to be included in the prediction of HER2 expression.

5 Conclusion

In conclusion, multi-mode ultrasound, including B-mode ultrasound, CEUS, and TIC, can predict HER2 expression status. Moreover, the fusion model of machine learning classifiers can improve the prediction results.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the institutional ethics Committee of "Chinese PLA General Hospital" in the center of

the principal investigator at Chinese PLA General Hospital in Beijing, China (IRB number: 2020-300). Approval was obtained on July 23rd, 2020. The research was carried out in accordance with the Declaration of Helsinki. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1585823/full#supplementary-material

References

1. Qi X, Zhang L, Chen Y, Pi Y, Chen Y, Lv Q, et al. Automated diagnosis of breast ultrasonography images using deep neural networks. *Med Image Anal.* (2019) 52:185–98. doi: 10.1016/j.media.2018.12.006

2. Diéras V, Miles D, Verma S, Pegram M, Welslau M, Baselga J, et al. Trastuzumab emtansine versus capecitabine plus lapatinib in patients with previously treated HER2-positive advanced breast cancer (EMILIA): a descriptive analysis of final overall survival results from a randomised, open-label, phase 3 trial. *Lancet Oncol.* (2017) 18:732–42. doi: 10.1016/S1470-2045(17)30312-1

3. Krop IE, Kim SB, Martin AG, LoRusso PM, Ferrero JM, Badovinac-Crnjevic T, et al. Trastuzumab emtansine versus treatment of physician's choice in patients with previously treated HER2-positive metastatic breast cancer (TH3RESA): final overall survival results from a randomised open-label phase 3 trial. *Lancet Oncol.* (2017) 18:743–54. doi: 10.1016/S1470-2045(17)30313-3

4. von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, et al. Adjuvant Pertuzumab and Trastuzumab in early HER2-positive breast Cancer. *N Engl J Med.* (2017) 377:122–31. doi: 10.1056/NEJMoa1703643

5. Bitencourt AGV, Gibbs P, Rossi Saccarelli C, Daimiel I, Lo Gullo R, Fox MJ, et al. MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer. *EBioMedicine.* (2020) 61:103042. doi: 10.1016/j.ebiom.2020.103042

6. Zardavas D, Fouad TM, Piccart M. Optimal adjuvant treatment for patients with HER2-positive breast cancer in 2015. *Breast.* (2015) 24:S143–8. doi: 10.1016/j.breast.2015.07.034

7. Choong GM, Cullen GD, O'Sullivan CC. Evolving standards of care and new challenges in the management of HER2-positive breast cancer. *CA Cancer J Clin.* (2020) 70:355–74. doi: 10.3322/caac.21634

8. Cronin KA, Harlan LC, Dodd KW, Abrams JS, Ballard-Barbash R. Population-based estimate of the prevalence of HER-2 positive breast cancer tumors for early stage patients in the US. *Cancer Investig.* (2010) 28:963–8. doi: 10.3109/07357907.2010.496759

9. Boughey JC, McCall LM, Ballman KV, Mittendorf EA, Ahrendt GM, Wilke LG, et al. Tumor biology correlates with rates of breast-conserving surgery and pathologic complete response after neoadjuvant chemotherapy for breast cancer: findings from the ACOSOG Z1071 (Alliance) prospective multicenter clinical trial. *Ann Surg.* (2014) 260:608–14. doi: 10.1097/SLA.0000000000000924

10. Das A, Nair MS, Peter SD. Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast Cancer: a review. *J Digit Imaging.* (2020) 33:1091–121. doi: 10.1007/s10278-019-00295-z

11. Slostad JA, Yun NK, Schad AE, Warrior S, Fogg LF, Rao R. Concordance of breast cancer biomarker testing in core needle biopsy and surgical specimens: a single institution experience. *Cancer Med.* (2022) 11:4954–65. doi: 10.1002/cam4.4843

12. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun*. (2020) 11:1236. doi: 10.1038/s41467-020-15027-z

13. Allott EH, Geradts J, Sun X, Cohen SM, Zirpoli GR, Khoury T, et al. Intratumoral heterogeneity as a source of discordance in breast cancer biomarker classification. *Breast Cancer Res*. (2016) 18:68. doi: 10.1186/s13058-016-0725-1

14. Lu Y, Zhu S, Tong Y, Fei X, Jiang W, Shen K, et al. HER2-low status is not accurate in breast Cancer Core needle biopsy samples: an analysis of 5610 consecutive patients. *Cancers (Basel)*. (2022) 14:6200. doi: 10.3390/cancers14246200

15. Santiago L, Adrada BE, Huang ML, Wei W, Candelaria RP. Breast cancer neoplastic seeding in the setting of image-guided needle biopsies of the breast. *Breast Cancer Res Treat*. (2017) 166:29–39. doi: 10.1007/s10549-017-4401-7

16. Liebens F, Carly B, Cusumano P, van Beveren M, Beier B, Fastrez M, et al. Breast cancer seeding associated with core needle biopsies: a systematic review. *Maturitas*. (2009) 62:113–23. doi: 10.1016/j.maturitas.2008.12.002

17. Stolier A, Skinner J, Levine EA. A prospective study of seeding of the skin after core biopsy of the breast. *Am J Surg*. (2000) 180:104–7. doi: 10.1016/S0002-9610(00)00425-6

18. Xie Y, Chen Y, Wang Q, Li B, Shang H, Jing H. Early prediction of response to neoadjuvant chemotherapy using quantitative parameters on automated breast ultrasound combined with contrast-enhanced ultrasound in breast Cancer. *Ultrasound Med Biol*. (2023) 49:1638–46. doi: 10.1016/j.ultrasmedbio.2023.03.017

19. Boca Bene I, Ciurea AI, Ciortea CA, Ștefan PA, Lisencu LA, Dudea SM, et al. Differentiating breast tumors from background parenchymal enhancement at contrast-enhanced mammography: the role of Radiomics-a pilot reader study. *Diagnostics (Basel)*. (2021) 11:1248. doi: 10.3390/diagnostics11071248

20. Luo J, Tang L, Chen Y, Yang L, Shen R, Cheng Y, et al. A prospective multicenter study on the additive value of contrast-enhanced ultrasound for biopsy decision of ultrasound BI-RADS 4 breast lesions. *Ultrasound Med Biol*. (2024) 50:1224–31. doi: 10.1016/j.ultrasmedbio.2024.04.010

21. Kotopoulis S, Popa M, Mayoral Safont M, Murvold E, Haugse R, Langer A, et al. SonoVue(®) vs. Sonazoid™ vs. Optison™: which bubble is best for low-intensity Sonoporation of pancreatic ductal adenocarcinoma? *Pharmaceutics*. (2022) 14:98. doi: 10.3390/pharmaceutics14010098

22. Alter J, Sennoga CA, Lopes DM, Eckersley RJ, Wells DJ. Microbubble stability is a major determinant of the efficiency of ultrasound and microbubble mediated in vivo gene transfer. *Ultrasound Med Biol*. (2009) 35:976–84. doi: 10.1016/j.ultrasmedbio.2008.12.015

23. Hao Y, Sun Y, Lei Y, Zhao H, Cui L. Percutaneous Sonazoid-enhanced ultrasonography combined with in vitro verification for detection and characterization of sentinel lymph nodes in early breast cancer. *Eur Radiol*. (2021) 31:5894–901. doi: 10.1007/s00330-020-07639-2

24. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting breast Cancer by applying deep learning to linked health records and mammograms. *Radiology*. (2019) 292:331–42. doi: 10.1148/radiol.2019182622

25. Turkki R, Byckhov D, Lundin M, Isola J, Nordling S, Kovanen PE, et al. Breast cancer outcome prediction with tumour tissue images and machine learning. *Breast Cancer Res Treat*. (2019) 177:41–52. doi: 10.1007/s10549-019-05281-1

26. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human epidermal growth factor receptor 2 testing in breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol*. (2018) 36:2105–22. doi: 10.1200/JCO.2018.77.8738

27. Ranstam J, Cook JA. LASSO regression. *Br J Surg*. (2018) 105:1348–8. doi: 10.1002/bjs.10895

28. Iparragirre A, Lumley T, Barrio I, Arostegui I. Variable selection with LASSO regression for complex survey data. *Stat*. (2023) 12:e578. doi: 10.1002/sta4.578

29. He X, Pan J, Jin O, Xu T, Liu B, Xu T, et al. Practical lessons from predicting clicks on ads at Facebook In: Eighth international workshop on data Mining for Online Advertising. New York, NY, USA: Association for Computing Machinery. (2014).

30. Zhou S, Hu C, Wei S, Yan X. Breast Cancer prediction based on multiple machine learning algorithms. *Technol Cancer Res Treat*. (2024) 23:15330338241234791. doi: 10.1177/15330338241234791

31. Maniotis AJ, Folberg R, Hess A, Seftor EA, Gardner LMG, Pe'er J, et al. Vascular channel formation by human melanoma cells in vivo and in vitro: vasculogenic mimicry. *Am J Pathol*. (1999) 155:739–52. doi: 10.1016/S0002-9440(10)65173-5

32. Chiao MT, Yang YC, Cheng WY, Shen CC, Ko JL. CD133+ glioblastoma stem-like cells induce vascular mimicry in vivo. *Curr Neurovasc Res*. (2011) 8:210–9. doi: 10.2174/156720211796558023

33. Wagenblast E, Soto M, Gutiérrez-Ángel S, Hartl CA, Gable AL, Maceli AR, et al. A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. *Nature*. (2015) 520:358–62. doi: 10.1038/nature14403

34. Morales-Guadarrama G, García-Becerra R, Méndez-Pérez EA, García-Quiroz J, Avila E, Díaz L. Vasculogenic mimicry in breast Cancer: clinical relevance and drivers. *Cells*. (2021) 10:1758. doi: 10.3390/cells10071758

35. Zhou YT, Cai WW, Li Y, Jiang X, Feng L, Zhu QY, et al. Correlations between quantitative parameters of contrast-enhanced ultrasound and vasculogenic mimicry in murine tumor model: a novel noninvasive technique for assessment? *Biol Proced Online*. (2019) 21:11. doi: 10.1186/s12575-019-0101-5

36. Liu H, Gao M, Gu J, Wan X, Wang H, Gu Q, et al. VEGFR1-targeted contrast-enhanced ultrasound imaging quantification of Vasculogenic mimicry microcirculation in a mouse model of choroidal melanoma. *Transl Vis Sci Technol*. (2020) 9:4. doi: 10.1167/tvst.9.3.4

37. Rothschild HT, Clelland E, Patterson A, Molina-Vega J, Kaur M, Symmans WF, et al. HER-2 low status in early-stage invasive lobular carcinoma of the breast: associated factors and outcomes in an institutional series. *Breast Cancer Res Treat*. (2023) 199:349–54. doi: 10.1007/s10549-023-06927-x

38. Xue S, Zhao Q, Tai M, Li N, Liu Y. Correlation between breast ultrasound microcalcification and the prognosis of breast Cancer. *J Healthc Eng*. (2021) 2021:6835963. doi: 10.1155/2021/6835963

39. Wang Y, Ikeda DM, Narasimhan B, Longacre TA, Bleicher RJ, Pal S, et al. Estrogen receptor-negative invasive breast cancer: imaging features of tumors with and without human epidermal growth factor receptor type 2 overexpression. *Radiology*. (2008) 246:367–75. doi: 10.1148/radiol.2462070169

40. Eroles P, Bosch A, Alejandro Pérez-Fidalgo J, Lluch A. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev*. (2012) 38:698–707. doi: 10.1016/j.ctrv.2011.11.005

41. Cui H, Sun Y, Zhao D, Zhang X, Kong H, Hu N, et al. Radiogenomic analysis of prediction HER2 status in breast cancer by linking ultrasound radiomic feature module with biological functions. *J Transl Med*. (2023) 21:44. doi: 10.1186/s12967-022-03840-7

42. Xing F, Gao H, Chen G, Sun L, Sun J, Qiao X, et al. CMTM6 overexpression confers trastuzumab resistance in HER2-positive breast cancer. *Mol Cancer*. (2023) 22:6. doi: 10.1186/s12943-023-01716-y

43. Loibl S, Gianni L. HER2-positive breast cancer. *Lancet*. (2017) 389:2415–29. doi: 10.1016/S0140-6736(16)32417-5

44. Mujtaba SS, Ni YB, Tsang JYS, Chan SK, Yamaguchi R, Tanaka M, et al. Fibrotic focus in breast carcinomas: relationship with prognostic parameters and biomarkers. *Ann Surg Oncol*. (2013) 20:2842–9. doi: 10.1245/s10434-013-2955-0

45. Hasebe T, Mukai K, Tsuda H, Ochiai A. New prognostic histological parameter of invasive ductal carcinoma of the breast: clinicopathological significance of fibrotic focus. *Pathol Int*. (2000) 50:263–72. doi: 10.1046/j.1440-1827.2000.01035.x

46. Hasebe T, Tsuda H, Hirohashi S, Shimosato Y, Iwai M, Imoto S, et al. Fibrotic focus in invasive ductal carcinoma: an indicator of high tumor aggressiveness. *Jpn J Cancer Res*. (1996) 87:385–94. doi: 10.1111/j.1349-7006.1996.tb00234.x

47. Wang XY, Hu Q, Fang MY, He Y, Wei HM, Chen XX, et al. The correlation between HER-2 expression and the CEUS and ARFI characteristics of breast cancer. *PLoS One*. (2017) 12:e0178692. doi: 10.1371/journal.pone.0178692

48. Liang X, Li Z, Zhang L, Wang D, Tian J. Application of contrast-enhanced ultrasound in the differential diagnosis of different molecular subtypes of breast Cancer. *Ultrason Imaging*. (2020) 42:261–70. doi: 10.1177/0161734620959780

49. Mie Lee Y, Kim SH, Kim HS, Jin Son M, Nakajima H, Jeong Kwon H, et al. Inhibition of hypoxia-induced angiogenesis by FK228, a specific histone deacetylase inhibitor, via suppression of HIF-1alpha activity. *Biochem Biophys Res Commun*. (2003) 300:241–6. doi: 10.1016/S0006-291X(02)02787-0

50. Jain RK. Normalizing tumor vasculature with anti-angiogenic therapy: a new paradigm for combination therapy. *Nat Med*. (2001) 7:987–9. doi: 10.1038/nm0901-987

51. Huang Z, Zhou PP, Li SS, Li K. Hepatic Angiomyolipoma: clinical features and imaging findings of quantitative contrast-enhanced ultrasound perfusion analysis and magnetic resonance imaging. *J Ultrasound Med*. (2020) 39:2111–22. doi: 10.1002/jum.15316

Frontiers in Medicine

Translating medical research and innovation into improved patient care

A multidisciplinary journal which advances our medical knowledge. It supports the translation of scientific advances into new therapies and diagnostic tools that will improve patient care.

Discover the latest Research Topics

See more →

frontiers | Research Topics