Biomedical Signal Processing and Control 110 (2025) 108201

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

DR-CapsNet: deep residual capsule network with dynamic routing for automated identification of hepatocellular carcinoma and cirrhosis in CT images

Biao Qu [a], Wangfeng He [b], Xiaopeng Yao [a], Dongjing Shan [a,*], Jian Shu [c,**]

[a] School of Medical Information and Engineering, Southwest Medical University, Luzhou, China
[b] Fujian Star-net Communication Co., Ltd., Fuzhou, China
[c] Department of Radiology, The Affiliated Hospital of Southwest Medical University, Luzhou, China

ARTICLE INFO

Keywords:
Hepatocellular carcinoma
Deep learning
Dynamic routing
Automated identification

ABSTRACT

Hepatocellular carcinoma (HCC) diagnosis in CT images is challenging due to the overlapping imaging features with cirrhosis, particularly in early-stage small nodules, where rapid differentiation requires both morphological sensitivity and computational efficiency. To address this, we propose DR-CapsNet, a deep residual capsule network that combines lightweight residual blocks with dynamic routing mechanisms. The residual module alleviates gradient degradation through skip connections while enhancing the extraction of high-level image features, whereas the capsule framework leverages vector neurons and dynamic routing to model hierarchical part-whole relationships between cirrhotic nodules and HCC lesions. The dynamic routing mechanism iteratively refines coupling coefficients to establish affine-invariant spatial correlations across multi-scale capsule layers, allowing for precise differentiation of subtle morphological variations. Experimental results indicate that DR-CapsNet outperforms existing state-of-the-art methods in both accuracy and inference speed. Moreover, it exhibits exceptional robustness, even under limited training conditions (400 samples) and class imbalance (1:4 ratio) challenges. Overall, DR-CapsNet presents an accurate, efficient, and robust solution for the diagnosis of hepatocellular carcinoma (HCC), particularly in clinical settings with constrained resources.

1. Introduction

Hepatocellular carcinoma (HCC) is the most common primary liver cancer [1,2], accounting for 75 %–85 % of cases [3], and ranks as the third leading cause of cancer-related deaths worldwide [4,5]. Its high incidence and mortality rates have drawn significant attention in clinical and research communities [6,7]. Early and accurate identification of HCC is critical for enhancing patient outcomes and survival rates [8,9]. However, distinguishing early-stage HCC from liver cirrhosis, particularly in cases involving small nodules with diameters $\leq$3 cm, remains a considerable challenge due to their overlapping CT imaging characteristics, such as nodular features, portal vein abnormalities, and heterogeneous vascular distribution. Traditional diagnostic approaches primarily rely on the meticulous analysis of vascular dynamic enhancement patterns at different phases to differentiate benign from malignant lesions [10]. Nevertheless, these methods are often limited by subjectivity in visual assessment, susceptibility to interobserver variability, and relatively low diagnostic efficiency [8,11]. Consequently, there is an urgent need to develop efficient and accurate diagnostic tools for the automated identify of HCC and liver cirrhosis.

Recent advancements in deep learning have made significant progress [12–15]. Various neural networks have been designed for the intelligent diagnosis of liver cancer. Among them, Convolutional neural networks (CNNs), known for their ability to extract relevant features from imaging data, have been widely applied in liver cancer diagnosis, including the grading of hepatocellular carcinoma, and automatic classification of liver fibrosis [16], as well as multimodal fusion in liver cancer classification [17]. However, CNNs are highly resource-

* Corresponding author at: School of Medical Information and Engineering, Southwest Medical University, No 1, Section1, Xianglin Road, Longmatan District, Sichuan 646000, China.
** Corresponding author at: Department of Radiology, The Affiliated Hospital of Southwest Medical University, No 25 Taiping St, Jiangyang District, Sichuan 646000, China
E-mail addresses: shandongjing@swmu.edu.cn (D. Shan), shujiannc@163.com (J. Shu).

https://doi.org/10.1016/j.bspc.2025.108201
Received 3 March 2025; Received in revised form 23 May 2025; Accepted 9 June 2025
Available online 18 June 2025

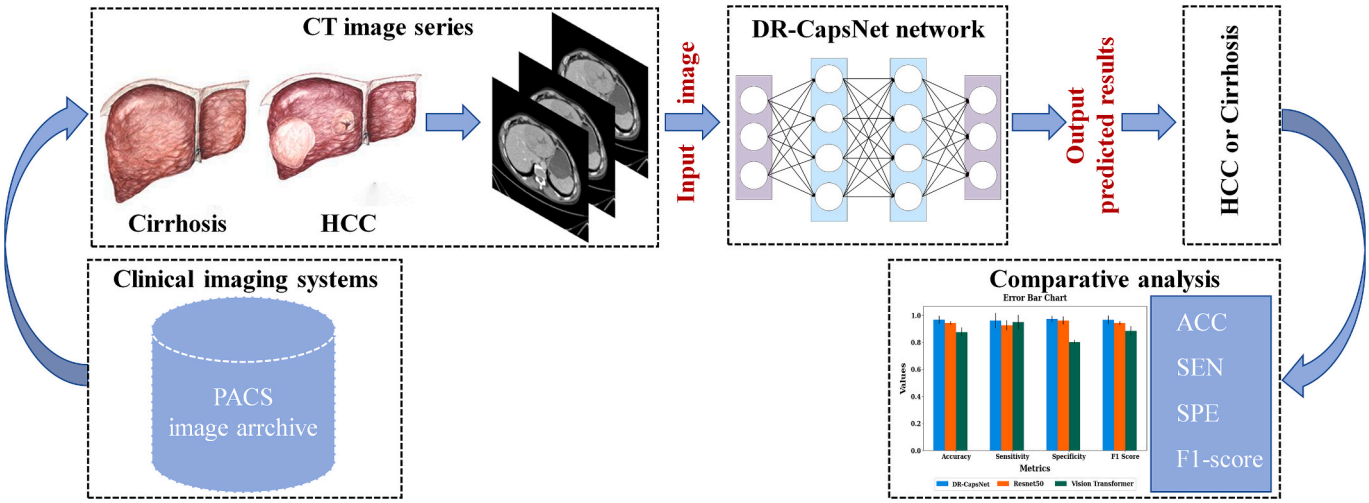

Fig. 1. Workflow of the study. PACS: picture archiving and communication system.

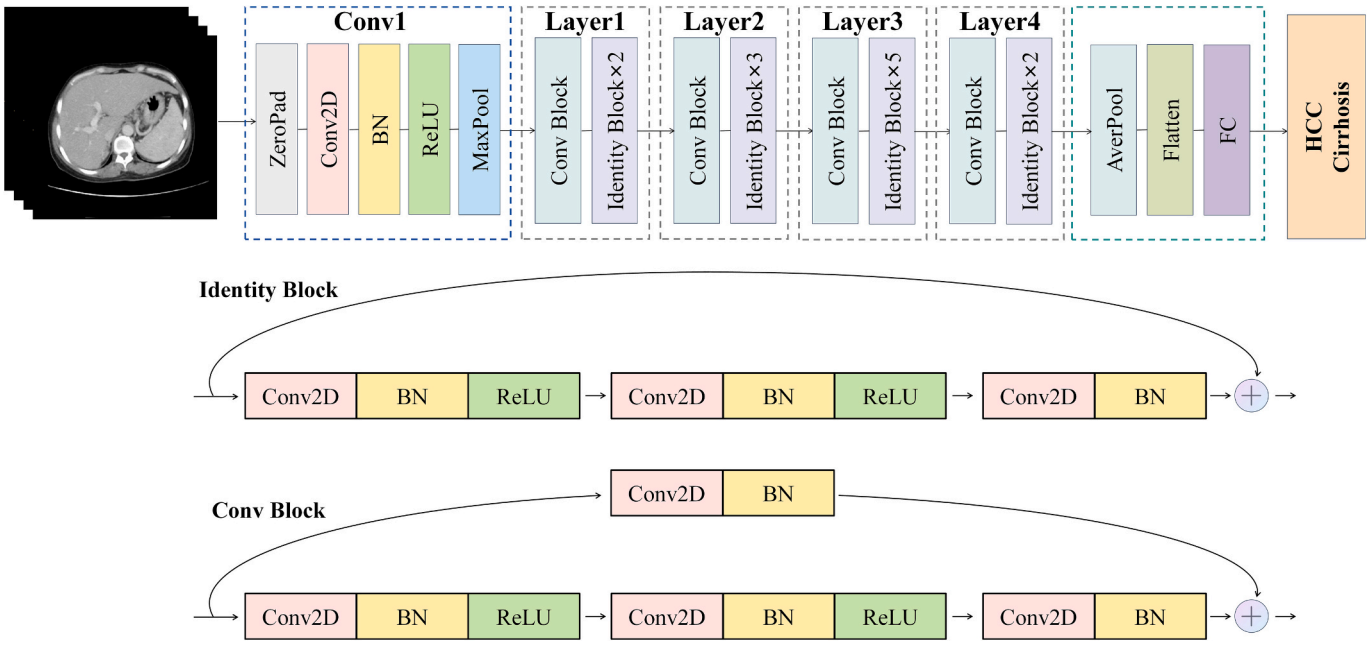

Fig. 2. The network structure of ResNet50. Conv2D: 2D Convolution; BN: Batch normalization. FC: Fully connected layer.

dependent, and as the network depth increases, issues such as gradient vanishing or explosion may become more pronounced, hindering the training process. Residual networks (ResNet) address these limitations by introducing residual learning and skip connections, which allow input information to bypass layers and be combined with convolutional outputs [18]. This design enhances network flexibility, preserves feature integrity, and significantly improves diagnostic accuracy [18]. ResNet has been successfully applied to metastatic tumor image classification [19], distinguishing normal from liver cancer samples [20], and classifying liver tumor CT images [21]. Vision Transformer (ViT), an advanced network architecture, offers a new perspective for automatic disease diagnosis [22]. ViT-based networks have been applied to classify benign and malignant tumors in breast ultrasound images [23], COVID-19 in chest X-ray images [24], and differentiate between benign and malignant lung cancer in CT data [25]. These studies highlight the effectiveness and accuracy of ViT in automatic disease diagnosis. However, ViT models are computationally expensive and require large datasets for training, which limits their practical application in clinical settings.

Capsule networks (CapsNet) [26], another advanced neural network architecture, effectively addresses the issue of information loss in the pooling layers of CNNs [27]. By capturing the spatial hierarchies of objects [28], CapsNet demonstrates higher accuracy and efficiency in classification tasks compared to traditional CNNs. For instance, CapsNet with enhanced amplifiers has significantly improved the classification accuracy of gastrointestinal disease images [29], while convolutional capsule network models have achieved excellent performance in similar applications [30]. However, to the best of our knowledge, deeper CapsNet architectures have not yet been developed and applied to the diagnosis and classification of hepatocellular carcinoma (HCC). Furthermore, the performance of deep CapsNet models under conditions of limited training data and class imbalance remains unexplored.

This study proposes a novel, efficient, and robust deep residual capsule network (DR-CapsNet) designed for the automatic differential of HCC and cirrhosis using enhanced CT images. By incorporating a dual-layer primary capsule network with a deep residual learning

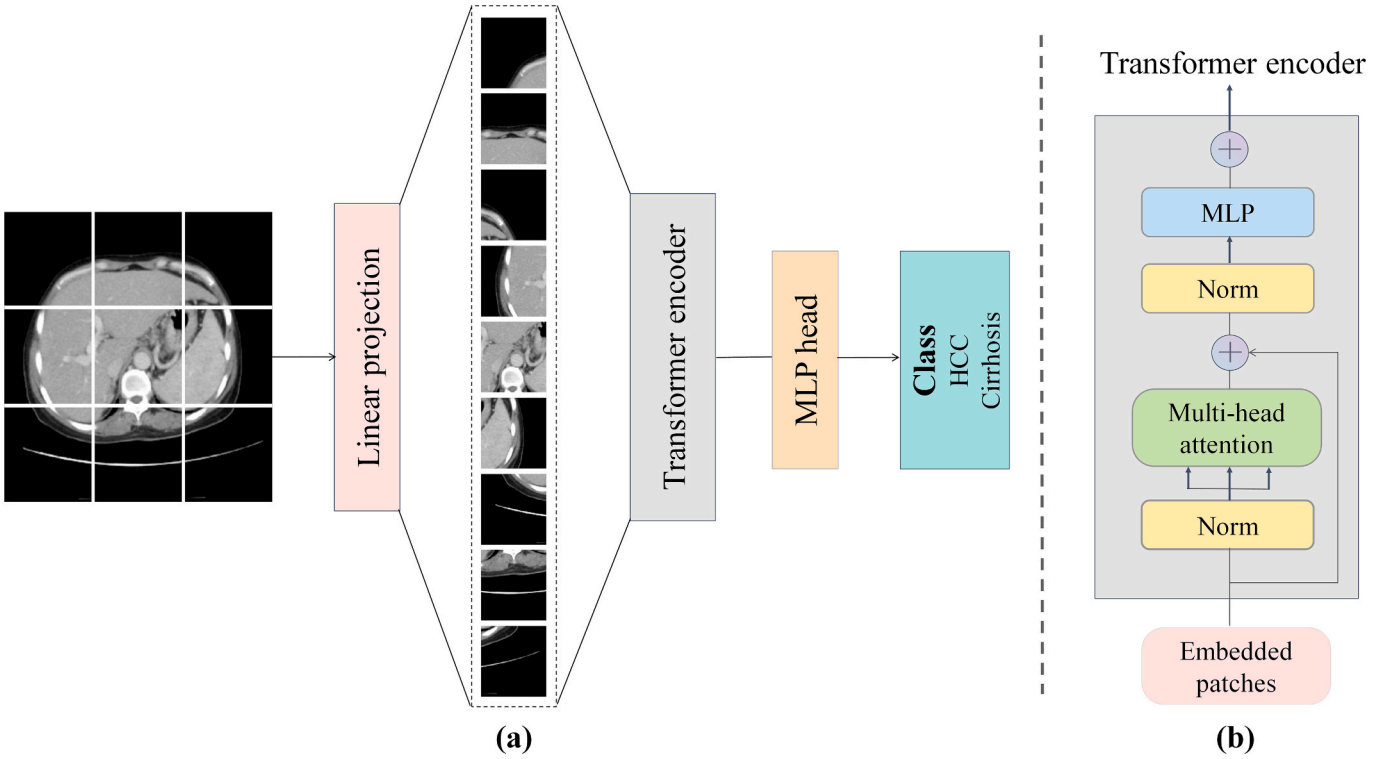

Fig. 3. The network structure of the vision transformer. (a) The network structure for classifying HCC and cirrhosis using the vision transformer. (b) The structure of the transformer encoder within the vision transformer. MLP: Multi-layer perceptron; Norm: Normalization.

mechanism, DR-CapsNet effectively captures subtle lesion features, thereby improving the accuracy and efficiency of HCC diagnosis. Compared to state-of-the-art methods, this model demonstrates superior robustness under conditions of limited training samples and class imbalance. The workflow of this study is illustrated in Fig. 1.

2. Related works

2.1. ResNet50

As a representative deep learning architecture, ResNet enables input signals to be directly passed to subsequent layers, allowing the network to learn residuals rather than global features. This design effectively mitigates the gradient vanishing problem often encountered when training deep networks. Among the ResNet series, ResNet34, ResNet50, and ResNet101 are widely used. To balance computational efficiency and accuracy, ResNet50 was chosen as the comparison method for this study, with its network architecture depicted in Fig. 2**.** ResNet50 consists of a Conv1 block (Zero padding = 3; Conv2D: kernel size = 7 × 7, stride = 2; Maxpool: kernel size = 3 × 3, stride = 2, zero padding = 1), followed by 4 convolutional blocks, 4 identity blocks, an adaptive average pooling layer, a flattening layer, and a fully connected layer.

2.2. Vision transformer

The classic ViT network was used as a comparison method (network structure shown in Fig. 3). In the experiment, the input image $\boldsymbol{D} \in \mathbb{R}^{H\times W\times C}$ is divided into a sequence of flattened 1D patches, which is described as:

$$\boldsymbol{d} \in \mathbb{R}^{N\times(Y^2C)} \tag{1}$$

here, $Y\times Y$ represents the size of each patch, and $N = HW/Y^2$ denotes the number of patches. By adjusting the parameter Y, ViT can control the model's focus on features of different scales within the image.

The ViT network comprises several key components (Fig. 3): Multi-head attention, which enables the model to learn interactions between patches in parallel across different representational subspaces; Multi-layer perceptron (MLP), which further performs non-linear processing of interactions from the multi-head attention mechanism; Layer normalization, applied at the input of each sublayer, which helps stabilize the training process; and residual connections, which allow gradients to flow directly to subsequent layers, thereby mitigating the gradient vanishing problem in deep models.

2.3. Other networks

Densely connected convolutional networks (DenseNet) [31] establishes dense connections between network layers, linking each layer to all previous layers to effectively utilize the feature maps of all layers. This architecture enhances the efficiency of information and gradient propagation, thereby effectively alleviating the gradient vanishing problem. ConvNext [32] combines the advantages of CNNs with modern design principles, improving traditional CNNs for easier initialization and training while retaining their strengths in image feature extraction. Swin Transformer (Swin ViT) [33] as a variant of the transformer, employs a hierarchical representation that enhances the network's ability to capture high-level features. These networks demonstrate varying degrees of advantages in image classification tasks.

3. Methods

3.1. Network architecture

The novel fully automated end-to-end pipeline of DR-CapsNet is illustrated in Fig. 4(a). The network takes enhanced CT images as input and outputs the automatic identification results (either HCC or cirrhosis). DR-CapsNet is carefully designed with two modules, which innovatively integrate the deep residual learning framework and the dynamic routing learning framework [26]. This fusion enhances the

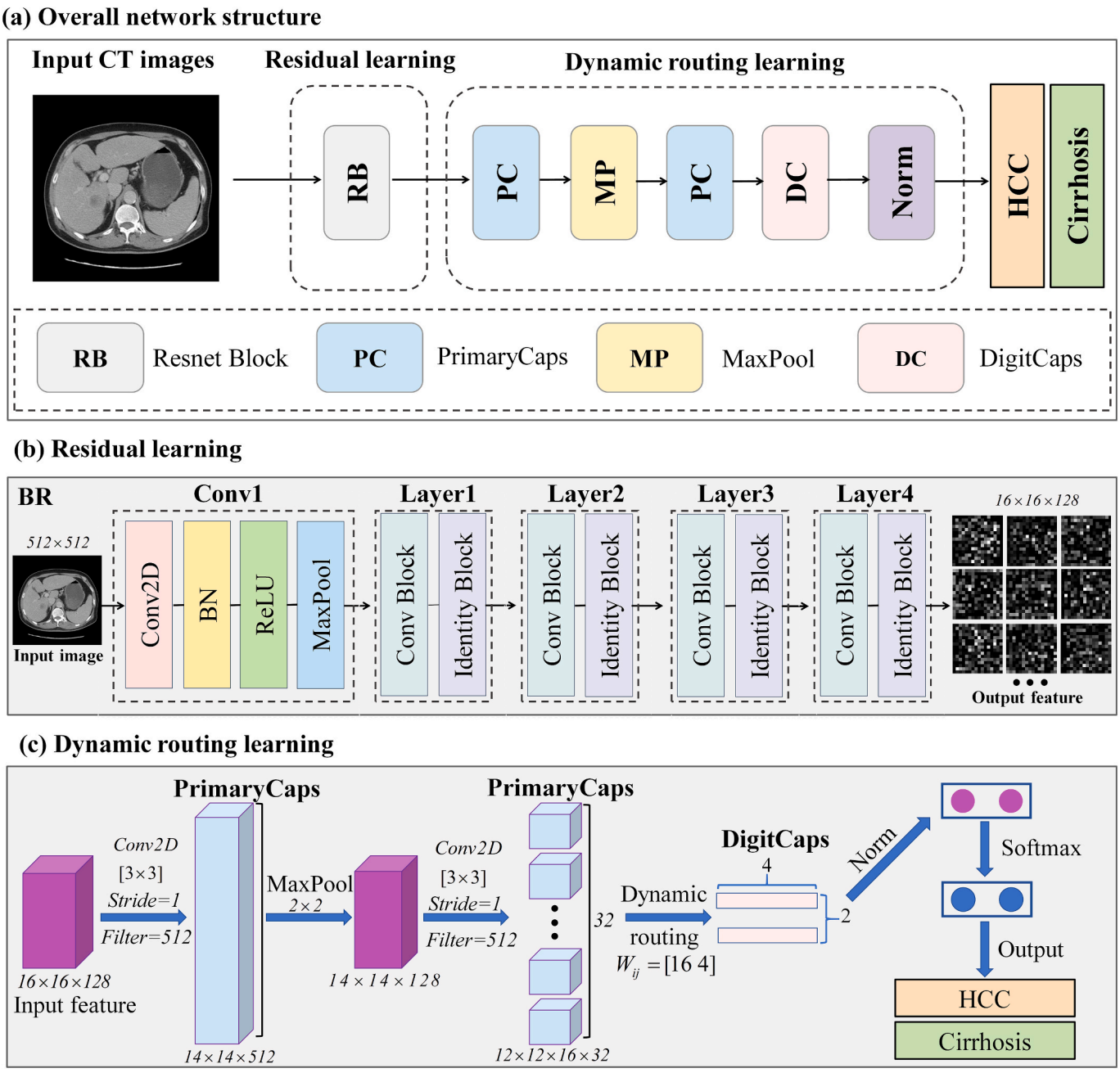

Fig. 4. Network structure of the proposed DR-CapsNet. (a) The main workflow consists of two key modules: a residual learning module and a dynamic routing learning module. (b) The residual learning module comprises a convolutional block and four learning layers. (c) The dynamic routing learning module consists of two PrimaryCaps layers and one DigitCaps layer.

Table 1
The optimal learning rate for network search.

Learning rate	ACC	SEN	SPE	F1-score
0.001	0.887	1.000	0.770	0.899
0.005	0.785	0.771	0.799	0.784
0.0002	0.943	0.888	0.934	0.941
0.0001	**0.972**	**1.000**	**0.943**	**0.973**
0.00005	0.953	0.949	0.957	0.953

Table 2
Performance comparison for different networks.

Methods	ACC	SEN	SPE	F1-score
ResNet50	0.934	0.944	0.923	0.935
DenseNet	0.856	0.850	0.861	0.856
ConvNext	0.863	0.916	0.809	0.871
ViT	0.901	0.986	0.813	0.910
Swin ViT	0.905	1.000	0.809	0.915
DR-CapsNet	**0.972**	**1.000**	**0.943**	**0.973**

ACC: Accuracy; SEN: Sensitivity; SPE: Specificity; The best performance in each metric is identified in bold font.

algorithm's ability to represent high-level image features while better capturing the spatial relationships in the images, ultimately improving the accuracy and robustness of the model's identification performance.

The residual learning framework consists of a convolutional block

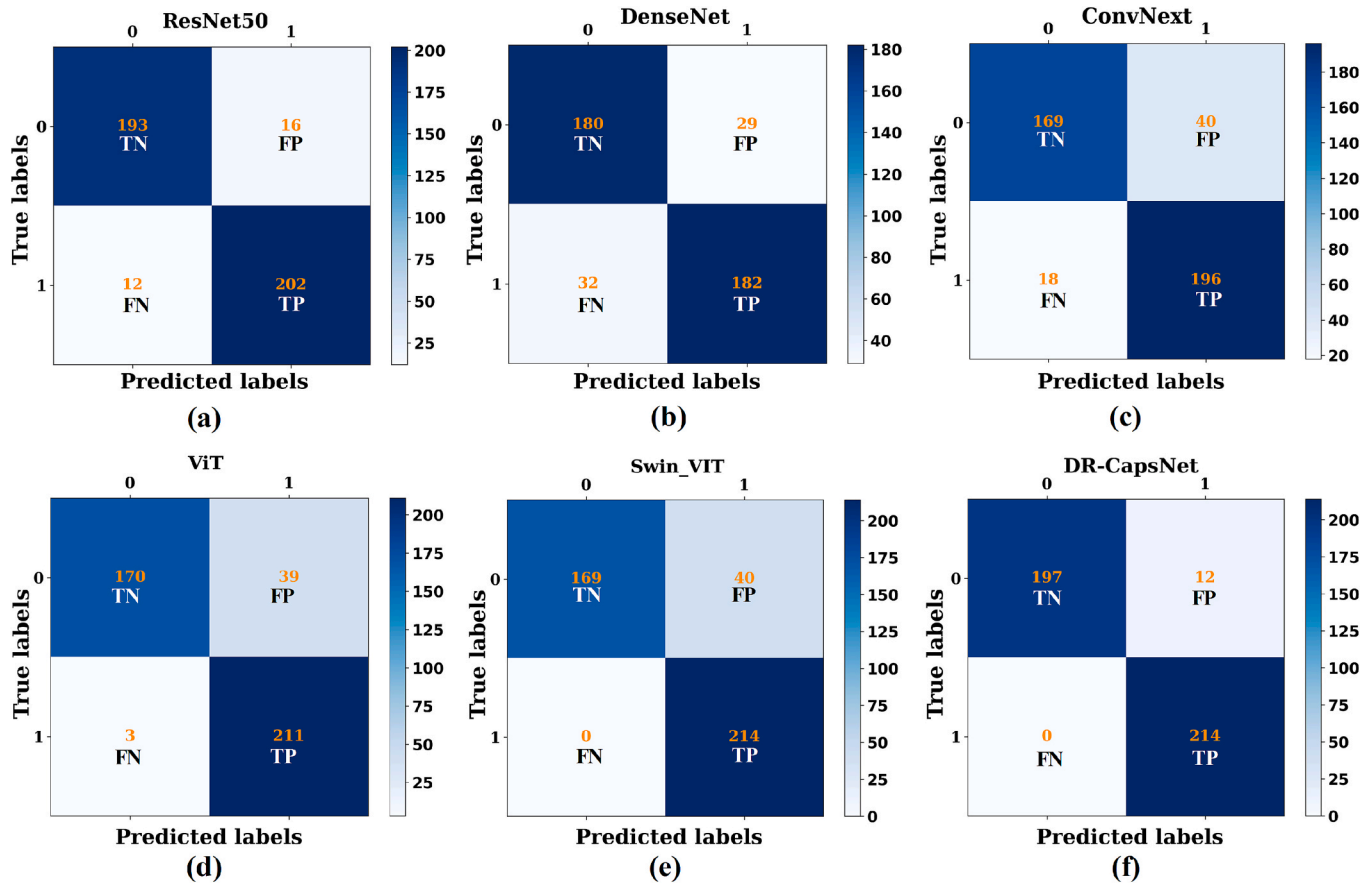

Fig. 5. Prediction confusion of different methods. (a-f) are prediction confusion of ResNet50, DenseNet, ConvNext, ViT, Swin ViT and DR-CapsNet methods respectively. 1: positive class; 0: negative class; TP: true positive; FP: false positive; TN: true negative; FN: false negative.

(including a 2D convolution with a kernel size of 7×7, stride $= 2$, batch normalization, ReLU activation, and a max-pooling layer with a kernel size of 3×3 and stride $= 2$) and four residual learning layers (each comprising a Conv-Block and an Identity Block) (Fig. 4(b)). The Conv-Block includes four convolutional layers, four batch normalization layers, and three ReLU activation functions. The Identity-Block contains three convolutional layers, three batch normalization layers, and three ReLU activation functions. Each block learns the residuals between the input and output, thus enhancing the network's training efficiency and accuracy. The residual learning framework captures more high-level image features as the network depth increases.

The dynamic routing learning framework mainly consists of the PrimaryCaps layer, a max-pooling layer, and the DigitCaps layer (Fig. 4 (c)). The PrimaryCaps layer performs 1D convolutions on the features outputted by the residual learning module and converts the feature maps into a set of capsules, with each capsule representing different properties of the input features. The dynamic routing module employs two layers of PrimaryCaps operations. The first captures the primary features related to spatial information, while the subsequent max-pooling layer reduces redundancy in the features. To enhance the model's ability to capture spatial hierarchical relationships within mages effectively a Primary-Caps layer is carefully designed between the MP layer and the DigitCaps layer, facilitating information transfer between the PrimaryCaps and DigitCaps layers via an iterative dynamic routing mechanism (Fig. 4). Finally, the identification is performed through the DigitCaps layer.

Dynamic routing optimizes the network by iteratively updating coupling coefficients, effectively capturing the relationships between capsules at different levels, thereby enhancing identification accuracy and improving model robustness. Unlike CNNs, which use scalar input activation functions such as ReLU and Sigmoid, DR-CapsNet enables information transfer between low-level and high-level capsules through an iterative dynamic routing algorithm [26] between the PrimaryCaps and DigitCaps layers. The coupling coefficient c_{ij} is updated based on similarity. For each capsule i at layer l and capsule j at layer $l + 1$, the coupling coefficient c_{ij} precisely routes capsule i to capsule j, and the dynamic routing algorithm determines the final routing.

Initially, the logarithmic prior probability b_{ij} between the i^{th} lower-level capsule and the j^{th} higher-level capsule are set to zero. The b_{ij} indicate the similarity of the lower-level capsule i to the higher-level capsule j.

$$b_{ij} = 0 \text{ for all } i, j \tag{2}$$

The coupling coefficient c_{ij} is obtained by performing softmax on b_{ij} and is expressed as:

$$c_{ij} = \frac{e^{b_{ij}}}{\sum_m e^{b_{im}}} \tag{3}$$

Each lower-level capsule i outputs a vector $\boldsymbol{u}_i^{(l)}$, which is transformed by a weight matrix $W_{ij}^{(l)}$ to generate the predicted vector $\widehat{\boldsymbol{u}}_{ji}^{(l)}$ for the

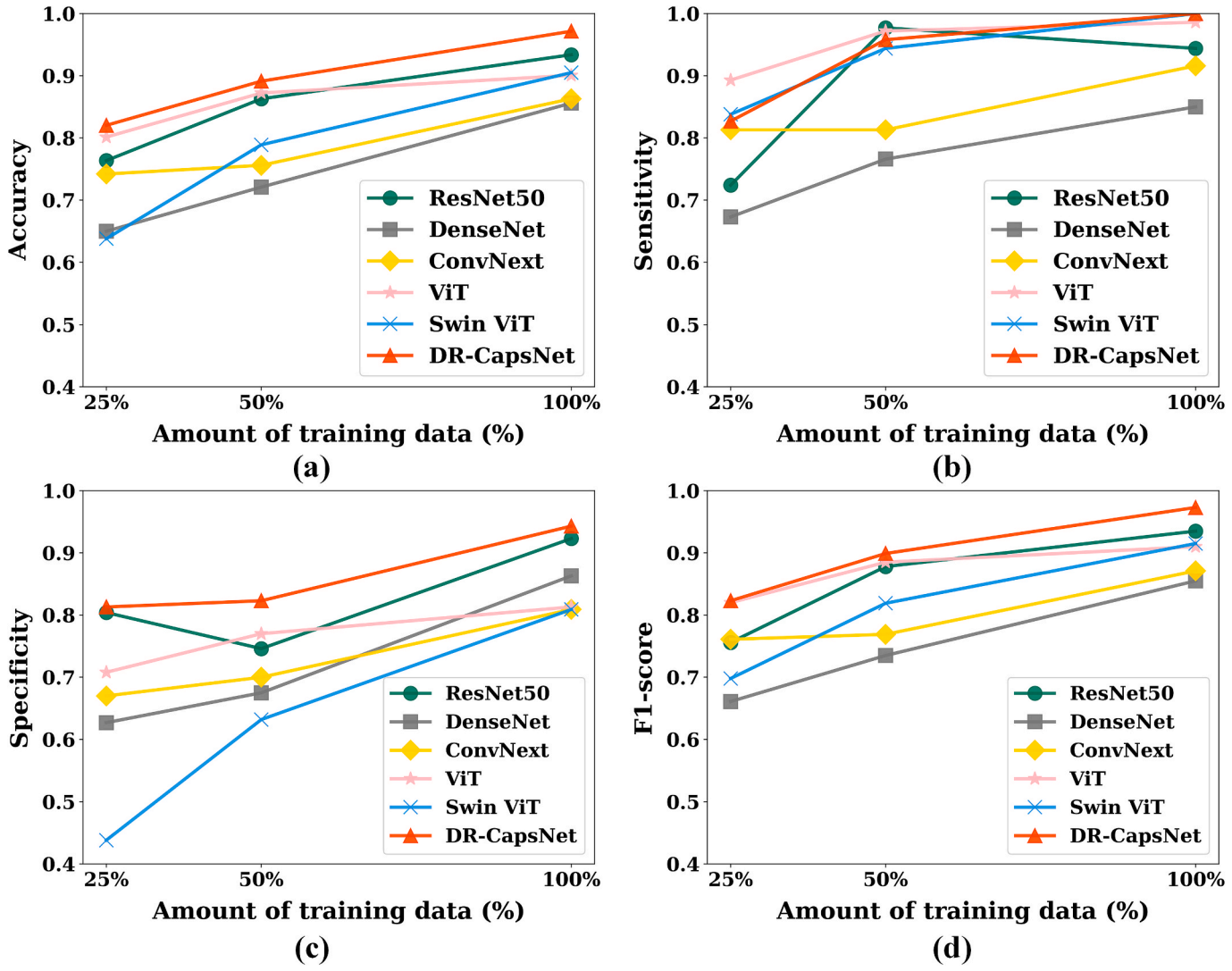

Fig. 6. Comparison of the performance of different methods across three training data sizes. (a-d) represent the accuracy, sensitivity, specificity and F1-score of different methods, respectively.

higher-level capsule j:

$$\widehat{\boldsymbol{u}}_{ji}^{(l)} = W_{ij}^{(l)} \cdot \boldsymbol{u}_i^{(l)} \tag{4}$$

where $\boldsymbol{u}_i^{(l)}$ is the output of the i^{th} capsule at layer l, and $W_{ij}^{(l)}$ is the weight matrix. The input to the j^{th} capsule at layer $l+1$, denoted as $\boldsymbol{s}_j^{(l+1)}$, is then computed as the weighted sum of the predicted vectors:

$$\boldsymbol{s}_j^{(l+1)} = \sum_i c_{ij} \cdot \widehat{\boldsymbol{u}}_{ji}^{(l)} \tag{5}$$

The capsule vectors are compressed using the nonlinear vector activation function Squash, which is defined as:

$$\boldsymbol{v}_j^{(l+1)} = \frac{\left\|\boldsymbol{s}_j^{(l+1)}\right\|^2}{1+\left\|\boldsymbol{s}_j^{(l+1)}\right\|^2} \frac{\boldsymbol{s}_j^{(l+1)}}{\left\|\boldsymbol{s}_j^{(l+1)}\right\|} \tag{6}$$

This function limits the length of the capsule vector $\boldsymbol{v}_j$ to the interval [0, 1), ensuring that the vector's magnitude stays within this range.

Update b_{ij} according to the following formula:

$$b_{ij} = b_{ij} + \widehat{\boldsymbol{u}}_{ji}^{(l)} \cdot \boldsymbol{v}_j^{(l+1)} \tag{7}$$

The dot product $\widehat{\boldsymbol{u}}_{ji}^{(l)} \cdot \boldsymbol{v}_j^{(l+1)}$ measures the similarity between the predicted and output vectors (the complete calculation process is shown in Algorithm 1).

In this work, the objective function of DR-CapsNet is defined as:

$$\mathscr{L} = \sum_{j=1}^{J} \left[T_j \max\left(0, m^+ - \left\|\boldsymbol{v}_j^{(l+1)}\right\|\right)^2 + \lambda(1 - T_j)\max\left(0, \left\|\boldsymbol{v}_j^{(l+1)}\right\| - m^-\right)^2 \right] \tag{8}$$

where J is the number of classes. T_j is the binary indicator of whether the class label is correct, $\left\|\boldsymbol{v}_j^{(l+1)}\right\|$ denotes the length of the capsule vector for the j^{th} class, while m^+ and m^- are the margin thresholds, and λ is a weighting factor. The margin loss ensures that the predicted vector length is less than the target length for incorrect classes, with zero loss for vectors exceeding the target length. The parameters are set as $m^+ = 0.9$, $m^- = 0.1$, and $\max(0, \cdot)^2$, ensuring the loss function emphasizes correct class predictions while penalizing incorrect ones based on the margin.

Table 3
The performance of the methods under three training data sizes.

Training data	Methods	ACC	SEN	SPE	F1-score
100 %	ResNet50	0.934	0.944	0.923	0.935
	DenseNet	0.856	0.850	0.861	0.856
	ConvNext	0.863	0.916	0.809	0.871
	ViT	0.901	0.986	0.813	0.910
	Swin ViT	0.905	1.000	0.809	0.915
	DR-CapsNet	**0.972**	**1.000**	**0.943**	**0.973**
50 %	ResNet50	0.863	**0.977**	0.746	0.878
	DenseNet	0.721	0.766	0.675	0.735
	ConvNext	0.756	0.813	0.700	0.769
	ViT	0.872	0.972	0.770	0.885
	Swin ViT	0.789	0.944	0.632	0.819
	DR-CapsNet	**0.891**	0.958	**0.823**	**0.899**
25 %	ResNet50	0.764	0.724	0.804	0.756
	DenseNet	0.650	0.673	0.627	0.661
	ConvNext	0.742	0.813	0.670	0.761
	ViT	0.801	**0.893**	0.708	0.820
	Swin ViT	0.638	0.838	0.438	0.698
	DR-CapsNet	**0.820**	0.827	**0.813**	**0.823**

ACC: Accuracy; SEN: Sensitivity; SPE: Specificity. 100%, 50%, and 25% represent the three different percentages of training data used, respectively. The best performance in each metric is identified in bold font.

Table 4
Performance comparison of six methods under imbalanced datasets.

IR	Methods	ACC	SEN	SPE	F1-score
1:1	ResNet50	0.934	0.944	0.923	0.935
	DenseNet	0.856	0.850	0.861	0.856
	ConvNext	0.863	0.916	0.809	0.871
	ViT	0.901	0.986	0.813	0.910
	Swin ViT	0.905	1.000	0.809	0.915
	DR-CapsNet	**0.972**	**1.000**	**0.943**	**0.973**
1:2	ResNet50	0.898	0.846	0.952	0.894
	DenseNet	0.747	0.673	0.823	0.729
	ConvNext	0.844	0.813	0.876	0.841
	ViT	0.764	0.738	0.790	0.760
	Swin ViT	0.837	0.841	0.833	0.839
	DR-CapsNet	**0.965**	**0.953**	**0.976**	**0.965**
1:4	ResNet50	0.813	0.664	0.967	0.782
	DenseNet	0.745	0.607	0.885	0.707
	ConvNext	0.806	0.664	0.952	0.776
	ViT	0.629	0.318	0.947	0.464
	Swin ViT	0.704	0.467	0.947	0.615
	DR-CapsNet	**0.905**	**0.841**	**0.971**	**0.900**

IR: Imbalance ratio; ACC: Accuracy; SEN: Sensitivity; SPE: Specificity. 1:1, 1:2 and 1:4 represent the three imbalance ratios of HCC to cirrhosis in the training datasets, respectively. The best performance in each metric is identified in bold font.

Algorithm 1: Dynamic routing algorithm

1: **Input:** for all capsule i in layer l and capsule j in layer $l+1$, logarithmic prior probability $b_{ij} \leftarrow 0$, prediction vector $\widehat{\boldsymbol{u}}_{j|i}$
2: **for** r **iterations do**
3: for all capsule i in layer l: $c_i \leftarrow softmax(b_i)$
4: for all capsule j in layer $l+1$: $s_j \leftarrow \sum_i c_{ij}\widehat{\boldsymbol{u}}_{j|i}$
5: for all capsule j in layer $l+1$: $\boldsymbol{v}_j \leftarrow squash(\boldsymbol{s}_j)$
6: for all capsule i in layer l and capsule j in layer $l+1$: $b_{ij} \leftarrow b_{ij} + \widehat{\boldsymbol{u}}_{j|i} \cdot \boldsymbol{v}_j$
7: **return** $\boldsymbol{v}_j$

3.2. Hyperparameter selection

The Adam optimizer was utilized for DR-CapsNet, and the optimal learning rate was determined through a search process involving pre-training of the network. The network was pre-trained using various learning rates (Table 1), and their predictive performances were evaluated. Results indicated that a learning rate of 0.0001 achieved the highest values for ACC, SEN, SPE, and F1-score. Consequently, a learning rate of 0.0001 was chosen for training the network.

4. Materials

4.1. Study patients

This study was approved by the Medical Ethics Committee of the Affiliated Hospital of Southwest Medical University, and the requirement for written informed consent was waived. A retrospective analysis was conducted using portal venous phase contrast-enhanced CT images from 1009 patients diagnosed with liver cirrhosis and 98 patients with HCC, all of whom were enrolled between March 2016 and December 2020.

Inclusion criteria for liver cirrhosis patients: 1) A diagnosis of liver cirrhosis confirmed through clinical evaluation, biochemical testing, and imaging examinations; 2) The patient had undergone abdominal enhanced CT imaging. Exclusion criteria for liver cirrhosis patients: 1) Patients with liver failure; 2) The presence of tumor lesions in the liver, or more than three hepatic cysts, calcifications, or stones with a diameter $\geq$3 cm.; 3) Significant dilation of the intrahepatic bile ducts, portal venous thrombosis, or a history of hepatic surgery; 4) Patients with concurrent malignancies at other sites.

Inclusion criteria for HCC patients: 1) Patients who underwent abdominal enhanced CT imaging; 2) A pathological diagnosis of HCC, with surrounding hepatic tissue showing cirrhotic changes. Exclusion criteria for HCC patients: 1) The presence of portal venous tumor thrombus or thrombus formation; 2) Concurrent malignancies at other sites.

4.2. Enhanced CT images acquisition

CT images were acquired using both a LightSpeed VCT 64-slice scanner (GE Healthcare) and a Brilliance iCT 256-slice scanner (Philips Healthcare), with field of views (FOVs) of 300 × 300 mm and 400 × 400 mm respectively, matrix size = 512 × 512, slice thickness = 5 mm.

An iodinated contrast agent (2 mL/kg; Ulrich CT Plus 150) was administered via the antecubital vein at an injection rate of 3 mL/s. Arterial phase scanning was initiated when the CT attenuation value of the abdominal aorta reached 150 Hounsfield units (HU). Following the completion of the arterial phase, portal venous phase scanning commenced 30 s later.

CT images were carefully selected by a radiologist with three years of experience. In total, 1014 images of hepatocellular carcinoma (HCC) and 1009 images of liver cirrhosis were included. The selection criteria for the HCC images required the presence of lesion areas. Approximately 80 % of the images were allocated to the training set (800 HCC images and 800 liver cirrhosis images), while the remaining 20 % formed the test set (214 HCC images and 209 liver cirrhosis images).

5. Experimental results

5.1. Experimental environment

All experiments were conducted on a Linux operating system (64-bit version). The hardware configuration included an Intel(R) Xeon(R) Silver 4210 CPU, 64 GB of RAM, and an ASPEED Graphics Family GPU. CUDA version 10.2 was utilized for computations.

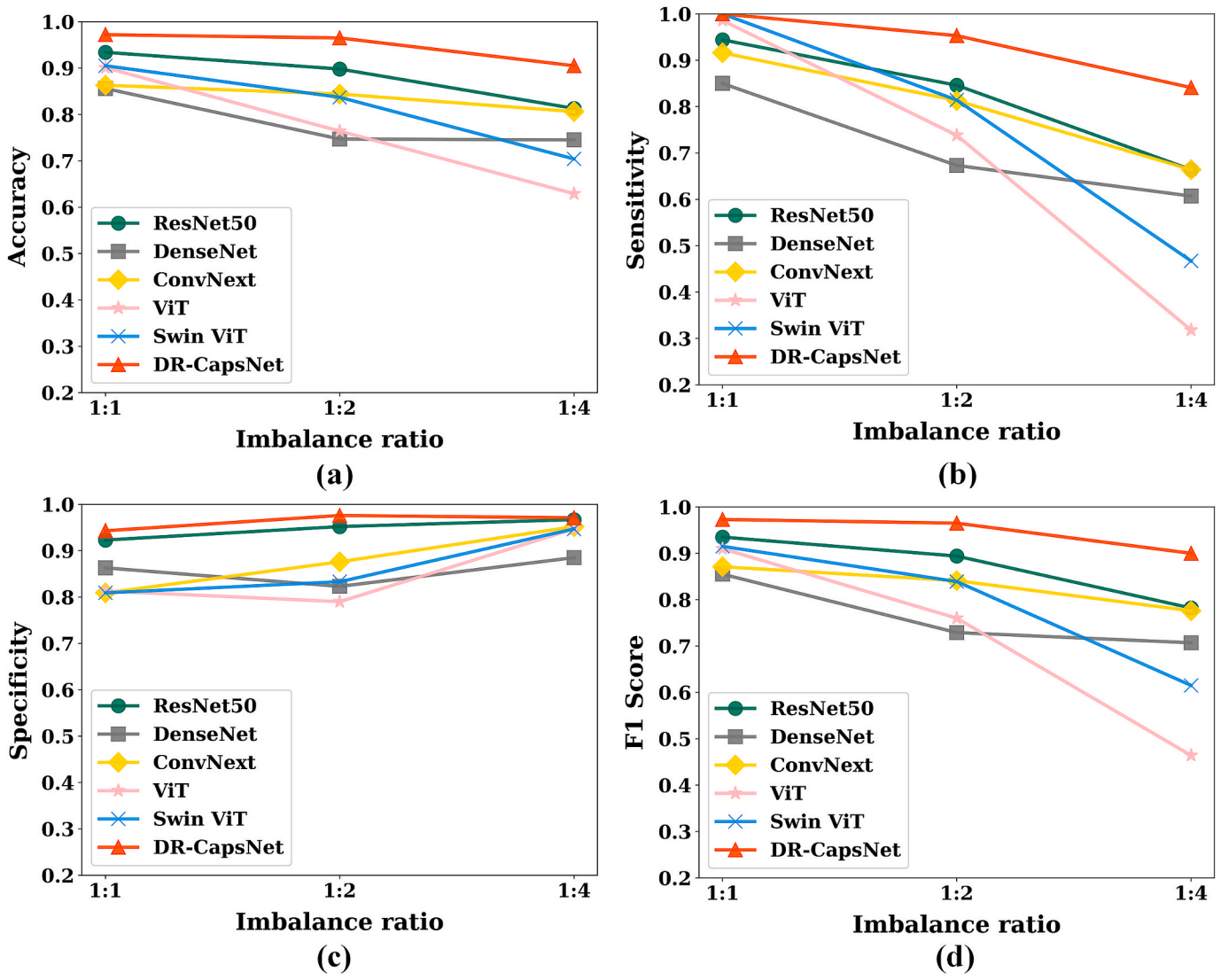

Fig. 7. Comparison of network performance under different imbalance ratios. (a) Accuracy; (b) Sensitivity; (c) Specificity; (d) F1 Score.

5.2. Evaluation criterion

The identification performance of the model was assessed using accuracy (ACC), sensitivity (SEN), specificity (SPE), and F1-score as evaluation metrics. Higher values of these metrics indicate better identification performance. The formulas for these metrics are as follows:

$$ACC = \frac{TP + TN}{TP + TN + FN + TN} \tag{9}$$

$$SEN = \frac{TP}{TP + FN} \tag{10}$$

$$SPE = \frac{TN}{FP + TN} \tag{11}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{12}$$

where True positive (*TP*) refers to the number of positive instances correctly classified by the model; True negative (*TN*) refers to the number of negative instances correctly classified; False positive (*FP*) refers to the number of instances incorrectly classified as positive; and False negative (*FN*) refers to the number of instances incorrectly classified as negative. The F1-score is a comprehensive evaluation metric for identification tasks evaluation indicator that evaluates the performance of the model under unbalanced data sets.

5.3. Comparison with state-of-the-art methods

5.3.1. Performance comparison for different methods

To evaluate the performance of the proposed method, Table 2 and Fig. 5 present the identification results using the full training dataset. Four metrics-ACC, SEN, SPE, and F1-score were employed to assess the identification performance of the different methods. Unless otherwise specified, all metrics are reported based on the test set.

Table 2 summarizes the comparison between the proposed DR-CapsNet method and state-of-the-art models, ResNet50, DenseNet, ConvNext, ViT and Swin ViT, in terms of identification performance. Specifically, the identification accuracies for ResNet50, DenseNet, ConvNext, ViT and Swin ViT, and DR-CapsNet were 0.934, 0.856, 0.863, 0.901, 0.905 and 0.972, respectively (third column). For sensitivity, the corresponding values were 0.944, 0.850, 0.916, 0.986, 1.000, and 1.000 (fourth column). In terms of specificity, ResNet50, ViT, and DR-CapsNet achieved values of 0.923, 0.861, 0.809, 0.813, 0.809, and 0.943, respectively (fifth column). Furthermore, the F1-scores for these methods were 0.935, 0.856, 0.871, 0.910, 0.915 and 0.973, respectively (sixth column). These results show that DR-CapsNet outperforms the

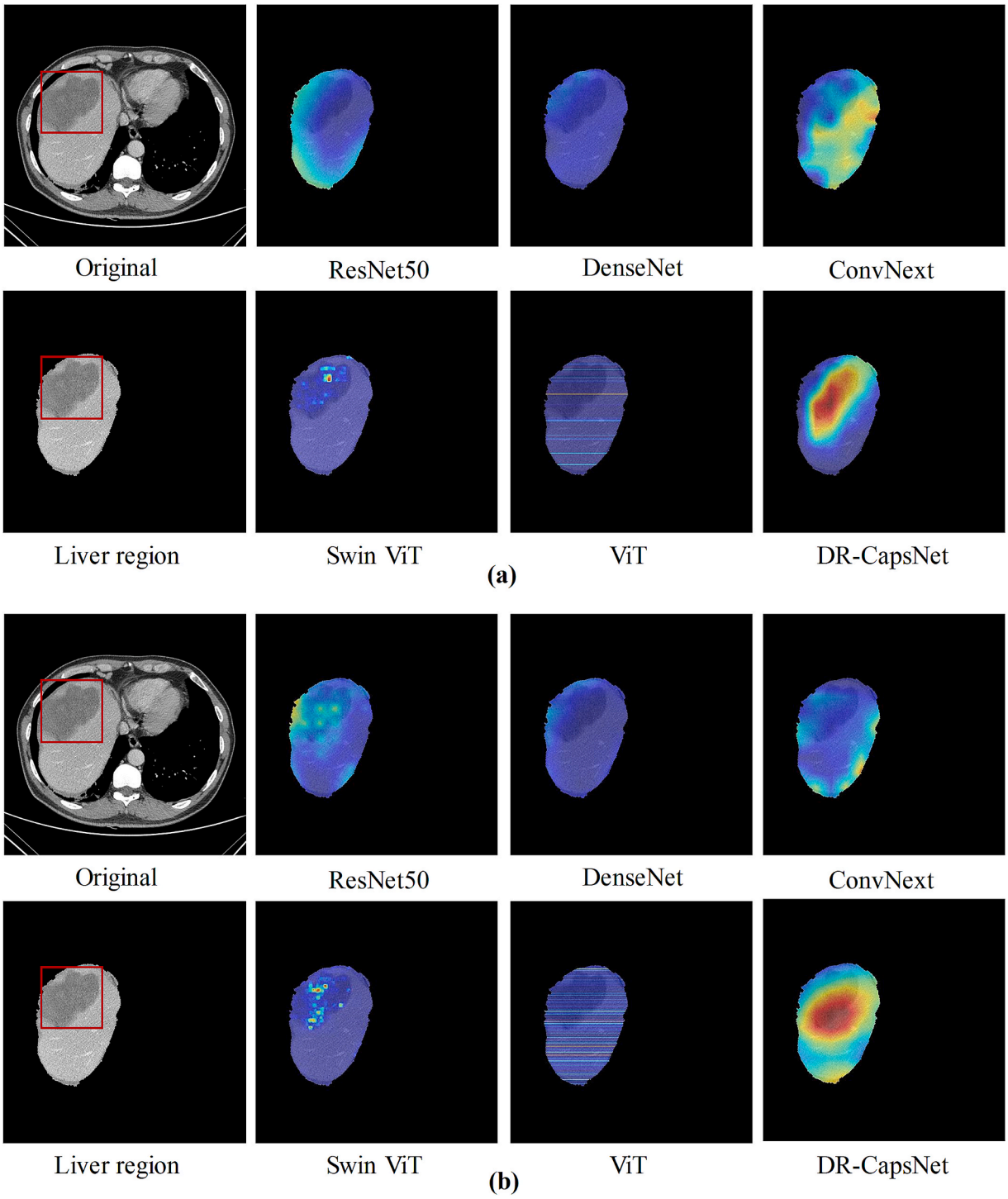

Fig. 8. Visualization of the network. (a) Small sample training using 25% of the training dataset; (b) Class imbalance training with an imbalance ratio of 1:4. Note: The red box indicates the liver cancer lesion area; the more intense the color in the mapping indicates that the region contributes more significantly to the model's prediction.

other methods across all four metrics, achieving the highest ACC, SEN, SPE, and F1-score. Compared to ResNet50, DenseNet, ConvNext, ViT, and Swin ViT, DR-CapsNet improved accuracy by 4.1 %, 13.6 %, 12.6 %, 7.9 %, and 7.4 %, respectively. This indicates that DR-CapsNet excels in automatically identifying HCC and cirrhosis.

Fig. 5 presents the prediction confusion matrices for each method. In the test set, which included 214 HCC samples (positive class) and 209 cirrhosis samples (negative class), DR-CapsNet achieved the highest number of true positives and true negatives, while minimizing both false positives and false negatives. Notably, DR-CapsNet recorded zero false negatives, highlighting its superiority in the automatic identification of HCC and cirrhosis.

5.3.2. Performance comparison under limited training dataset

To assess the identification performance of different methods under limited training dataset, six methods were trained using 100 % (800 HCC images and 800 cirrhosis images), 50 % (400 HCC images and 400 cirrhosis images), and 25 % (200 HCC images and 200 cirrhosis images)

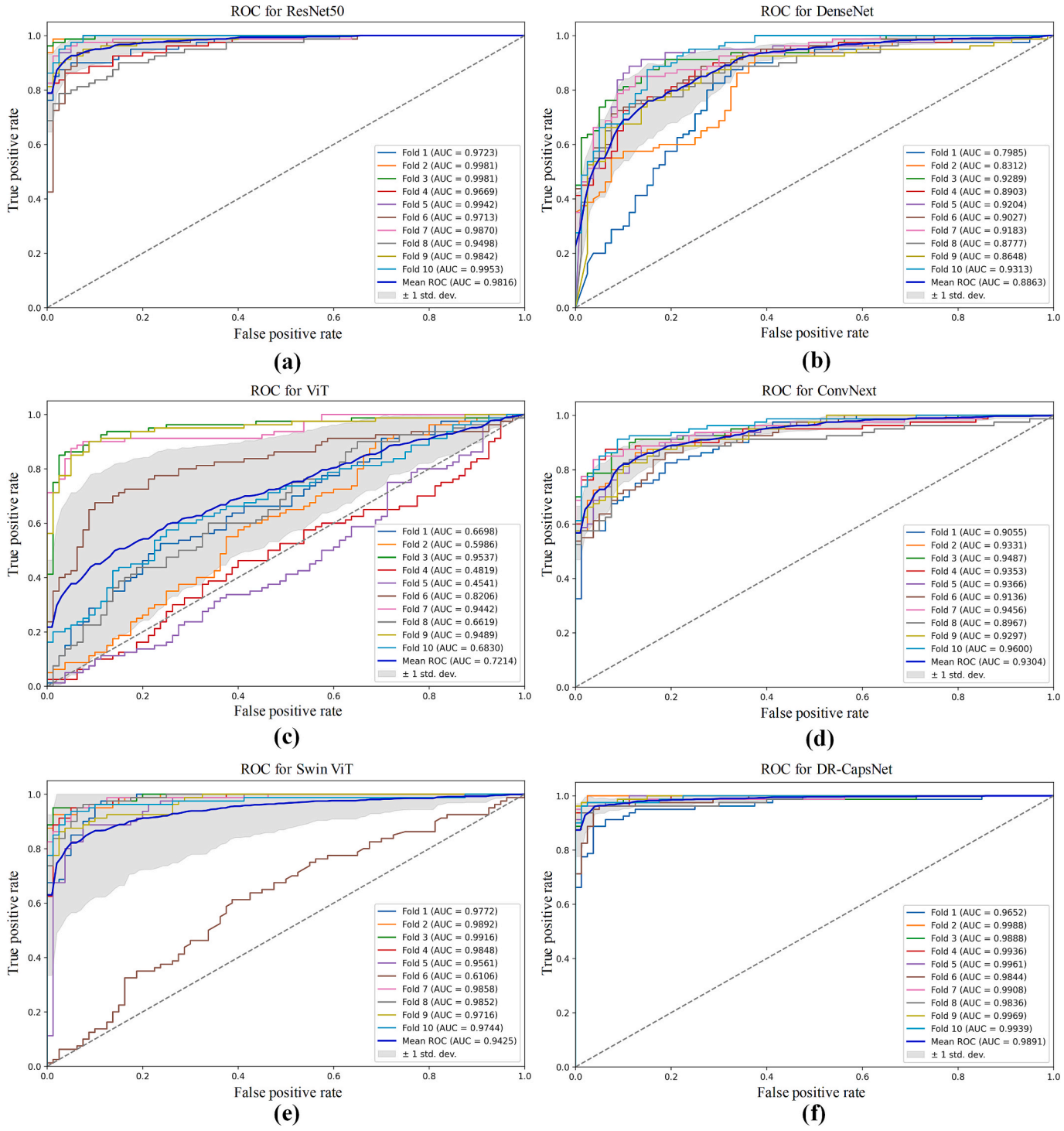

Fig. 9. Performance comparison via 10-fold cross-validated ROC curves.

of the training dataset, respectively.

Results are presented in Fig. 6. As the size of the training dataset decreases, the predictive performance of all methods declined. However, DR-CapsNet consistently outperformed the other methods, achieving the highest ACC, SPE, and F1-score, and maintaining superior overall predictive performance, except for a slight reduction in sensitivity when using 25 % and 50 % of the training data (Table 3). Notably, even with just 25 % of the training data, DR-CapsNet demonstrated the highest identification ACC, SPE, and F1-score. Specifically, with 25 % of the training data, compared to ResNet50, DenseNet, ConvNext, ViT, and Swin ViT, DR-CapsNet improved accuracy by 7.3 %, 26.2 %, 10.5 %, 2.4 %, and 28.5 %, respectively, DR-CapsNet improved F1-score by 8.9 %, 24.5 %, 8.1 %, 0.4 %, and 17.9 %, respectively. These results highlight that the proposed DR-CapsNet method is less reliant on the size of the training dataset compared to other methods. DR-CapsNet exhibits better predictive performance and robustness, even under conditions of limited training data, further demonstrating its potential for clinics applications.

5.3.3. Performance comparison of the method on imbalanced training datasets

To examine the impact of training data imbalance on method performance, the ratio of HCC to liver cirrhosis data in the training set,

Table 5
Performance comparison of different methods with 10-fold cross-validation (Mean ± Standard Deviation).

Methods	ACC	SEN	SPE	F1-score
ResNet50	0.937 ± 0.021	0.911 ± 0.034	0.964 ± 0.027	0.933 ± 0.022
DenseNet	0.833 ± 0.033	0.884 ± 0.068	0.788 ± 0.070	0.839 ± 0.028
ConvNext	0.875 ± 0.026	0.877 ± 0.044	0.865 ± 0.046	0.873 ± 0.029
ViT	0.750 ± 0.095	0.759 ± 0.094	0.665 ± 0.165	0.753 ± 0.090
Swin ViT	0.957 ± 0.038	0.965 ± 0.025	0.918 ± 0.067	0.963 ± 0.035
DR-CapsNet	**0.991 ± 0.005**	**0.987 ± 0.015**	**0.989 ± 0.005**	**0.990 ± 0.006**

Table 6
Ablation study.

Model	ACC	SEN	SPE	F1-score
DR-CapsNet without RL	0.948	0.967	0.928	0.950
DR-CapsNet without DRL	0.953	0.986	0.919	0.955
DR-CapsNet	**0.972**	**1.000**	**0.943**	**0.973**

Note: RL represents the residual learning module; DRL represents the dynamic routing learning module.

referred to as the imbalance ratio, was set to 1:1 (800 HCC images:800 cirrhosis images), 1:2 (400 HCC images:800 cirrhosis images) and 1:4 (200 HCC images:800 cirrhosis images) for network training.

Table 4 reports the performance comparison of different methods. The proposed DR-CapsNet method outperformed others in terms of ACC, SEN, SPE, and F1-score across all three imbalance ratios. As the imbalance ratio decreased from 1:1 to 1:4 (i.e., as the class imbalance increased), DR-CapsNet's performance improvements relative to ResNet50 were notable: ACC rose from 4.0 % to 11.3 %, SEN from 5.9 % to 26.7 %, and F1-score from 4.1 % to 15.1 %. Compared to ViT, the improvements were even more substantial, with ACC increasing from 7.9 % to 43.9 %, SEN from 1.4 % to 62.2 %, and F1-score from 6.9 % to 94.0 %.

Fig. 7 illustrates the identification performance of different methods at three imbalance ratios. As the data imbalance ratio decreased, DR-CapsNet exhibited a growing advantage in ACC, SEN, and F1-score when compared to other methods. These results indicate that the proposed method exhibits better robustness on imbalanced training datasets.

5.3.4. Network interpretation

To enhance the interpretability of the network, Class Activation Mapping (CAM) [34] is employed to visualize the regions of interest. In these maps, darker colors are associated with greater contributions to the network's predictions. When compared to the visual results from other advanced networks, superior focus on liver cancer lesion areas is demonstrated by the proposed DR-CapsNet under both small sample size conditions (Fig. 8(a)) and class imbalance (Fig. 8(b)). This enhanced focus may be regarded as contributing to improved classification performance when training data is limited.

5.3.5. Performance comparison of the method with cross-validation

To prevent overfitting, the 10-fold cross-validation was conducted, with the evaluation metrics averaged from the 10-folds to assess model performance. Experimental results indicated that the proposed model, DR-CapsNet, achieved the highest mean AUC (mean AUC = 0.9891) during 10-fold cross-validation (Fig. 9). Additionally, DR-CapsNet also excelled in ACC, SEN, SPE, and F1-score metrics (Table 5). These results further support the superior classification performance of the proposed method.

5.3.6. Ablation experiments

To verify the rationality of the model design, ablation experiments were conducted focusing on the residual learning module and the dynamic routing learning module. The performance changes of the network were evaluated by removing these two modules separately and assessing the resulting metrics.

As shown in Table 6 and Fig. 10, noticeable declines in the network's performance were observed when either the residual learning module or the dynamic routing learning module was removed, indicated by decreased ACC, SEN, SPE, and F1-score. Conversely, significant improvements in the performance metrics of DR-CapsNet were achieved when both modules were employed concurrently. This enhancement validates the effectiveness of the model design and highlights the contributions of each component to the overall performance.

Table 7
Performance comparison of methods using the PLC-CECT dataset.

Methods	ACC	SEN	SPE	F1-score
ResNet50	0.926	1.000	0.852	0.931
DenseNet	0.894	0.950	0.838	0.899
ConvNext	0.786	0.980	0.592	0.821
ViT	0.868	0.912	0.824	0.874
Swin ViT	0.932	1.000	0.864	0.936
DR-CapsNet	**0.998**	**1.000**	**0.996**	**0.998**

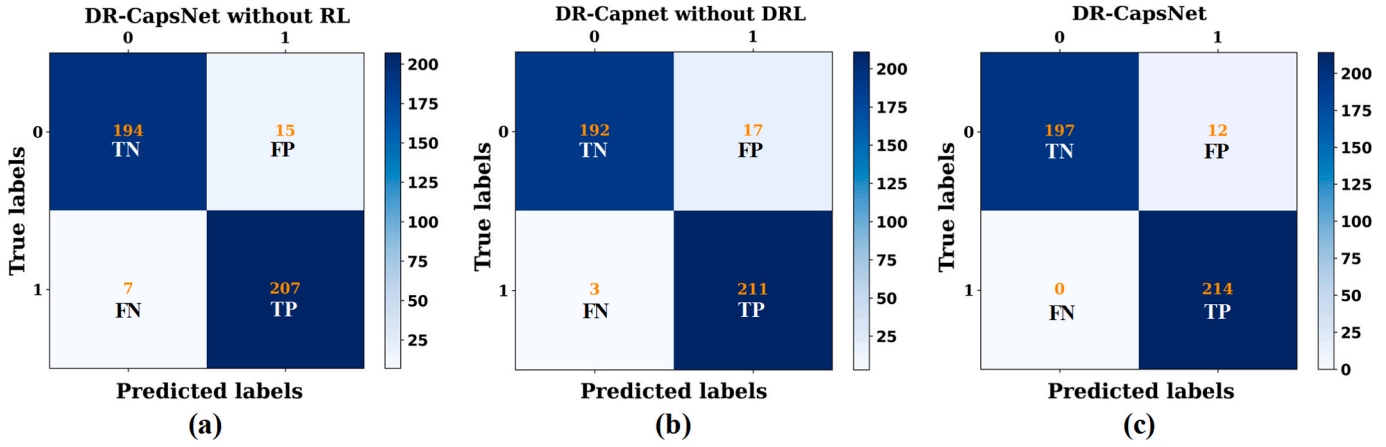

Fig. 10. The confusion matrix predicted by the network under different components. Note: RL represents the residual learning module; DRL represents the dynamic routing learning module.

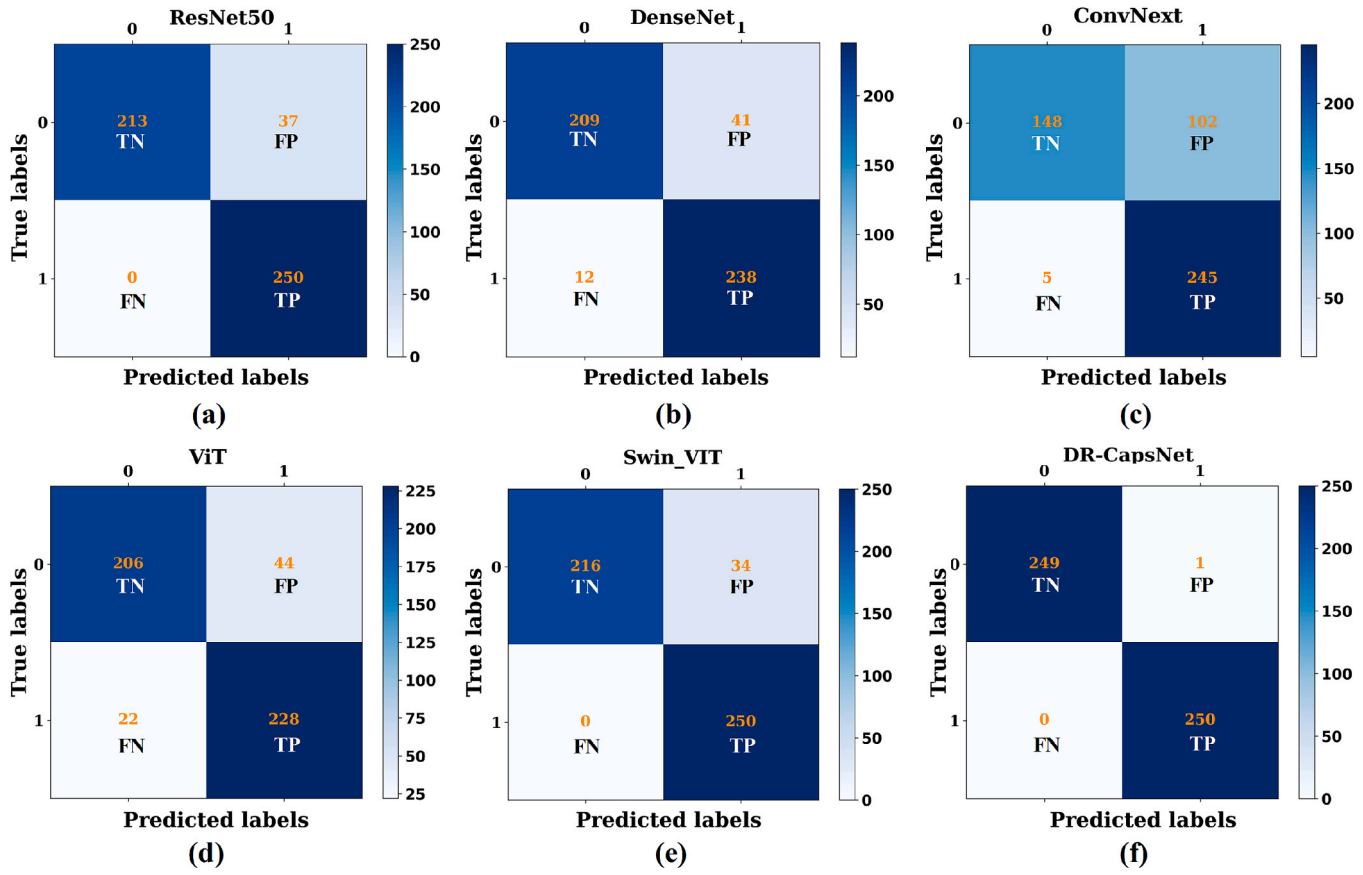

Fig. 11. Prediction confusion of different methods using PLC-CECT dataset. (a-f) are prediction confusion of ResNet50, DenseNet, ConvNext, ViT, Swin ViT and DR-CapsNet methods respectively. 1: positive class; 0: negative class; TP: true positive; FP: false positive; TN: true negative; FN: false negative.

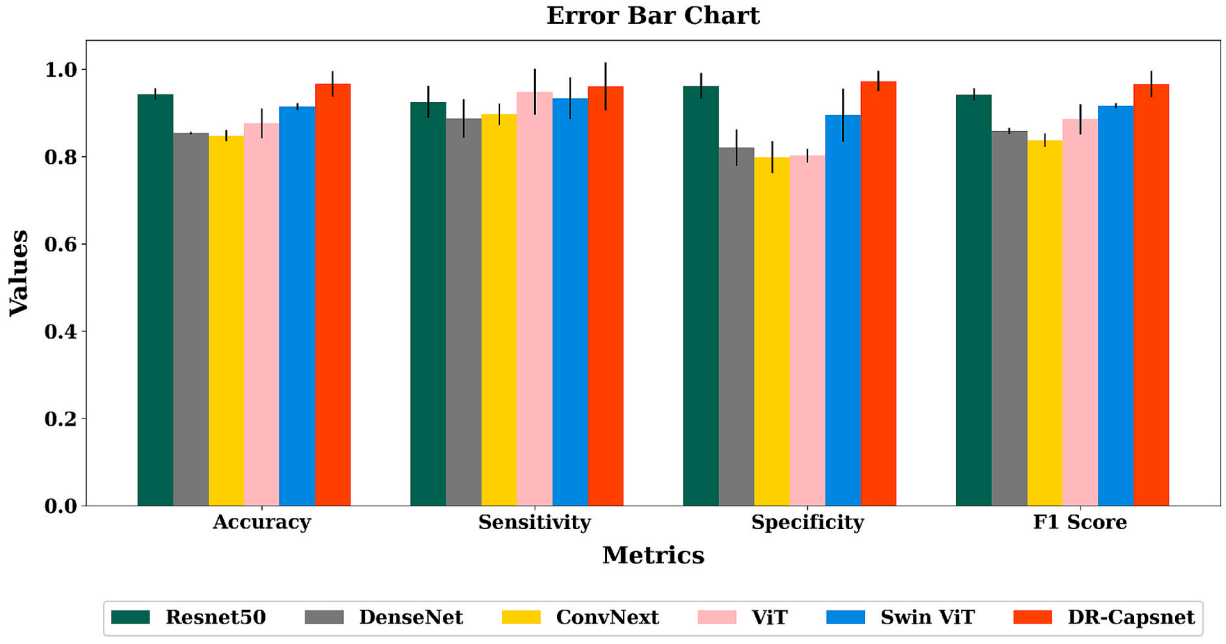

Fig. 12. Performance comparison of the method under three separate experiments.

5.4. Performance comparison on public datasets

To further validate the proposed algorithm, the public Primary Liver Cancer Contrast-Enhanced CT Imaging Dataset (PLC-CECT) [35] was used for the automatic identification of hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC). The training set involved data from 20 HCC patients (1000 images) and 21 ICC patients (1000 images), while the testing set included data from 6 HCC patients (250

Table 8
Computation time (Unit: seconds).

Model	Computer language	Mean time
ResNet50	Python	9.4
DenseNet	Python	6.2
ConvNext	Python	7.8
ViT	Python	14.9
Swin ViT	Python	8.6
DR-CapsNet	Python	**1.6**

Notes: This experiment employed 423 images for testing. The average computation time per image was calculated by dividing the total processing time by the number of images. The best performance in metric is identified in bold font.

images) and 6 ICC patients (250 images). The DR-CapsNet model achieved the highest accuracy, sensitivity, specificity, and F1-score (Table 7 and Fig. 11). These observations are consistent with our previous findings regarding the automatic identification of HCC and cirrhosis, further validating the superiority of DR-CapsNet in identification performance.

6. Discussion

6.1. Analysis of stability across re-split datasets

To evaluate the stability of the proposed method, the original dataset was re-split into three distinct training and testing sets. The splitting strategy involved using the first 80 %, middle 80 %, and last 80 % of the original data as training sets for three separate experiments, with the remaining 20 % of the data used as test sets. To ensure fairness, the same model architecture and parameter configurations were applied across all experiments.

When comparing the five methods, DR-CapsNet consistently achieved the highest average values for ACC, SEN, SPE, and F1-score across all three datasets (Fig. 12), confirming its better identification performance. These results are consistent with the findings reported in Section 5.2.1. Additionally, the error bar analysis in Fig. 12 shows that the variance in the four metrics across the different datasets is small for all six methods, with variance ranges between 0.02 and 0.1. This demonstrates that the performance of all methods is consistent, reliable, and stable across different training datasets.

6.2. Computation time

To further assess the computational efficiency of the proposed method, Table 8 provides a comparison of computation times across six methods. DR-CapsNet achieved the fastest inference speed, with an average identification time of only 1.6 s per image. This highlights the high efficiency of the proposed method for the automatic identification of hepatocellular carcinoma and cirrhosis, which is expected to enhance diagnostic efficiency for clinicians.

6.3. Limitations

Despite the promising results, our study has several limitations. First, the liver dataset was collected from a single hospital in southwestern China, potentially limiting the model's generalizability to more diverse populations. Second, relying solely on CT imaging data as a single modality may have resulted in suboptimal performance compared to multimodal approaches. Third, the relatively small sample size used in this study could increase the risk of overfitting, although we implemented cross-validation techniques to mitigate this concern. Future research should explore the integration of multi-center, multi-regional datasets to enhance model robustness, incorporate multi-modal data, and validate performance on larger cohorts. These enhancements would advance the development of more reliable and clinically applicable intelligent diagnostic tools for hepatocellular carcinoma.

7. Conclusion

In this work, an efficient and robust DR-CapsNet model is proposed for the identification of HCC and liver cirrhosis. By designing a lightweight deep residual module and incorporating the dynamic routing mechanism, the proposed algorithm enhances stability and robustness while significantly reducing computational time. Experimental results demonstrate that DR-CapsNet outperforms state-of-the-art methods in identification performance, particularly on small sample datasets and under imbalanced data conditions. This work offers a promising approach for the rapid and precise identification of HCC and cirrhosis, contributing to enhanced efficiency in clinical auxiliary diagnosis.

Ethics statement

This study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Medical Ethics Committee of the Affiliated Hospital of Southwest Medical University (Approval No: KY2022350), with a waiver of informed consent.

CRediT authorship contribution statement

Biao Qu: Writing – review & editing, Writing – original draft, Visualization, Methodology, Funding acquisition, Data curation. **Wangfeng He:** Visualization, Methodology, Data curation. **Xiaopeng Yao:** Validation, Investigation, Formal analysis. **Dongjing Shan:** Writing – review & editing, Validation, Supervision, Project administration. **Jian Shu:** Writing – review & editing, Supervision, Project administration, Data curation, Conceptualization.

Funding

This work was financially supported by the Luzhou Municipal People's Government-Southwest Medical University Science and Technology Cooperation Project (NO. 2024LZXNYDJ083), Southwest Medical University Technology Program (NO. 2024ZKY052), Southwest Medical University Technology Program (NO. 2021ZKMS048).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Wangfeng He is an employee of Fujian Star-net Communication Co., Ltd. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Mengping Huang for assistance with data collection.

Data availability

The authors do not have permission to share data, the code may be made available upon evaluation of researchers' requests.

References

[1] A. Forner, M. Reig, J. Bruix, Hepatocellular carcinoma, Lancet 391 (10127) (2018) 1301–1314.
[2] P. Maria Jesi, V. Antony Asir Daniel, Differential CNN and KELM integration for accurate liver cancer detection, Biomed. Signal Process. Control 95 (2024) 106419.
[3] J. Calderaro, T.P. Seraphin, T. Luedde, T.G. Simon, Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma, J. Hepatol. 76 (6) (2022) 1348–1361.

[4] H. Rumgay, M. Arnold, J. Ferlay, O. Lesi, C.J. Cabasag, J. Vignat, M. Laversanne, K. A. McGlynn, I. Soerjomataram, Global burden of primary liver cancer in 2020 and predictions to 2040, J. Hepatol. 77 (6) (2022) 1598–1606.
[5] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA-Cancer J. Clin. 71 (3) (2021) 209–249.
[6] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R.L. Siegel, I. Soerjomataram, A. Jemal, Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA-Cancer J. Clin. 74 (3) (2024) 229–263.
[7] R. Archana, L. Anand, Residual u-net with self-attention based deep convolutional adaptive capsule network for liver cancer segmentation and classification, Biomed. Signal Process. Control 105 (2025) 107665.
[8] S. Huang, X. Nie, K. Pu, X. Wan, J. Luo, A flexible deep learning framework for liver tumor diagnosis using variable multi-phase contrast-enhanced CT scans, J. Cancer Res. Clin. Oncol. 150 (10) (2024) 443.
[9] P.V. Nayantara, S. Kamath, K. Manjunath, K. Rajagopal, Computer-aided diagnosis of liver lesions using CT images: a systematic review, Comput. Biol. Med. 127 (2020) 04035.
[10] D.G. Mitchell, J. Bruix, M. Sherman, C.B. Sirlin, LI-RADS (liver imaging reporting and data system): summary, discussion, and consensus of the LI-RADS management working group and future directions, Hepatology 61 (3) (2015) 1056–1065.
[11] A. Kiani, B. Uyumazturk, P. Rajpurkar, et al., Impact of a deep learning assistant on the histopathologic classification of liver cancer, npj Digit. Med. 3 (1) (2020) 23.
[12] D.V. Phan, C.L. Chan, A.A. Li, T.Y. Chien, V.C. Nguyen, Liver cancer prediction in a viral hepatitis cohort: a deep learning approach, Int. J. Cancer 147 (10) (2020) 2871–2878.
[13] Z. Bo, J. Song, Q. He, B. Chen, Z. Chen, X. Xie, D. Shu, K. Chen, Y. Wang, G. Chen, Application of artificial intelligence radiomics in the diagnosis, treatment, and prognosis of hepatocellular carcinoma, Comput. Biol. Med. 173 (2024) 108337.
[14] K. Mohit, R. Gupta, B. Kumar, Contrastive learned self-supervised technique for fatty liver and chronic liver identification, Biomed. Signal Process. Control 100 (2025) 106950.
[15] J. Li, F. Cao, H. Zhang, ALRIGMR: adaptive logistic regression via integrating gene mutation and RNA-seq for liver cancer diagnosis, Biomed. Signal Process. Control 91 (2024) 106025.
[16] J.H. Lee, I. Joo, T.W. Kang, Y.H. Paik, D.H. Sinn, S.Y. Ha, K. Kim, C. Choi, G. Lee, J. Yi, W.C. Bang, Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network, Eur. Radiol. 30 (2) (2020) 1264–1273.
[17] R.A. Khan, M. Fu, B. Burbridge, Y. Luo, F.X. Wu, A multi-modal deep neural network for multi-class liver cancer diagnosis, Neural Netw. 165 (2023) 553–561.
[18] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[19] M. Wang, X. Gong, Metastatic cancer image binary classification based on resnet model, in: IEEE 20th International Conference on Communication Technology, 2020, pp. 1356–1359.
[20] L. Lu, B.J. Daigle Jr, Prognostic analysis of histopathological images using pre-trained convolutional neural networks: application to hepatocellular carcinoma, PeerJ 8 (2020) e8668.
[21] F.P. Romero, A. Diler, G. Bisson-Gregoire, S. Turcotte, R. Lapointe, F. Vandenbroucke-Menu, A. Tang, S. Kadoury, End-to-end discriminative deep network for liver lesion classification, in: Ieee 16th International Symposium on Biomedical Imaging, 2019, pp. 1243–1246.
[22] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold,S. Gelly, An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:201011929 2020.
[23] B. Gheflati, H. Rivaz, Vision transformers for classification of breast ultrasound images, in: Annual International Conference of the IEEE Engineering in Medicine & Biology Society, 2022, pp. 480–483.
[24] D. Shome, T. Kar, S.N. Mohanty, P. Tiwari, K. Muhammad, A. AlTameem, Y. Zhang, A.K.J. Saudagar, Covid-transformer: interpretable covid-19 detection using vision transformer for healthcare, Int. J. Environ. Res. Public Health 18 (21) (2021) 11086.
[25] L. Gai, W. Chen, R. Gao, Y.-W. Chen, X. Qiao, Using vision transformers in 3-D medical image classifications, IEEE Int. Conf. Image Proc. (2022) 696–700.
[26] S. Sabour, N. Frosst, G.E. Hinton, Dynamic routing between capsules, advances in neural information processing systems, in: Proceedings of the Annual Conference on Neural Information Processing Systems, 2017, pp. 3856–3866.
[27] J. Li, Q. Zhao, N. Li, L. Ma, X. Xia, X. Zhang, N. Ding, N. Li, A survey on capsule networks: evolution, application and future development, in: International Conference on High Performance Big Data and Intelligent Systems, 2021, pp. 177–185.
[28] R. Shi, L. Niu, A brief survey on capsule network, in: IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, 2020, pp. 682–686.
[29] Y. Afriyie, B.A. Weyori, A.A. Opoku, Exploring optimised capsule network on complex images for medical diagnosis, in: IEEE 8th International Conference on Adaptive Science and Technology, 2021, pp. 1–5.
[30] H. Zhang, Z. Li, H. Zhao, Z. Li, Y. Zhang, Attentive octave convolutional capsule network for medical image classification, Appl. Sci. 12 (5) (2022) 2634.
[31] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[32] Z. Liu, H. Mao, C. Wu, C. Feichtenhofer, T. Darrell, S. Xie, A convnet for the 2020s, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976–11986.
[33] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 10012–10022.
[34] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, Int. J. Comput. Vis. 128 (2020) 336–359.
[35] J. Luo, X. Wan, J. Du, L. Liu, L. Zhao, X. Peng, M. Wu, S. Huang, X. Nie, Comprehensive multi-phase 3D contrast-enhanced CT imaging for primary liver cancer, Sci. Data 12 (1) (2025) 768.