BioMedical Engineering OnLine

RESEARCH

Open Access

Four-phase CT lesion recognition based on multi-phase information fusion framework and spatiotemporal prediction module

Shaohua Qiao[1], Mengfan Xue[2], Yan Zuo[2], Jiannan Zheng[2], Haodong Jiang[2], Xiangai Zeng[2] and Dongliang Peng[2*]

*Correspondence:
dlpeng@hdu.edu.cn

[1] HDU-ITMO Joint Institute, Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China
[2] School of Automation, Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China

Abstract

Multiphase information fusion and spatiotemporal feature modeling play a crucial role in the task of four-phase CT lesion recognition. In this paper, we propose a four-phase CT lesion recognition algorithm based on multiphase information fusion framework and spatiotemporal prediction module. Specifically, the multiphase information fusion framework uses the interactive perception mechanism to realize the channel-spatial information interactive weighting between multiphase features. In the spatiotemporal prediction module, we design a 1D deep residual network to integrate multiphase feature vectors, and use the GRU architecture to model the temporal enhancement information between CT slices. In addition, we employ CT image pseudo-color processing for data augmentation and train the whole network based on a multi-task learning framework. We verify the proposed network on a four-phase CT dataset. The experimental results show that the proposed network can effectively fuse the multi-phase information and model the temporal enhancement information between CT slices, showing excellent performance in lesion recognition.

Keywords: Deep learning, Contrast-enhanced CT, Hepatic malignancy, Classification, Artificial intelligence

Introduction

Liver cancer is one of the most common cancers threatening human health, with a relative survival rate of only 21% over the past 5 years [1]. Computed tomography (CT), as a rapid, efficient, and stable imaging technique, plays an important role in the identification of liver malignancies, particularly in distinguishing between hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC) using multiphase CT. The multiphase CT imaging data typically includes four phases: non-contrast phase (NC), arterial phase (ART), portal venous phase (PV), and delayed phase (DL). Through meticulous assessment of these images, a diagnosis of the patient can be made. However, this diagnostic method is particularly vulnerable to psychological, physiological and other external factors, and exhibits strong subjectivity [2].

Deep learning has become a booming force in the field of computer-aided cancer diagnosis due to its ability to solve extremely complex challenges in cancer diagnosis with high accuracy over time, providing convenience for medical experts in diagnosis and treatment [3]. In contrast to conventional machine learning methods that necessitate manual feature extraction from inputs, deep learning (DL) methods are more advanced strategies in artificial intelligence (AI) that can learn features directly from sample data [4]. In recent years, as a new branch of deep learning, convolutional neural networks (CNNs) have received increasing attention in image pattern recognition and artificial intelligence strategies [5–7]. For example, Yasaka Koichiro et al. [8] proposed a convolutional neural network structure to classify liver tumors in multi-phase CT images. The model has three channels, corresponding to NC, ART and PV phases. Liang et al. [9] proposed a deep learning-based framework for the classification of focal liver lesions in multiphase CT images, which combines a deep residual neural network (ResNet) with global and local paths and bidirectional long short-term memory models to use multiphase information from the feature level. Ling et al. [10] proposed a deep learning model based on CT scans and minimal additional information (MAI), including textual inputs such as patient age and gender. The model is trained using a combination of 3D convolutional neural networks and multilayer perceptron.

However, many current research methods have their own limitations, as follows: (1) Most studies [11–17] only use three-phase CT. Although three-phase CT can provide relatively rich information, it still cannot fully reflect all the details of some complex pathological processes. In contrast, four-phase CT contains more time point data, which enables the deep learning network to accept richer input information and improve the accuracy and performance of the model. (2) Multi-phase CT images reflect the physiological state changes of patients at different time points. This spatio-temporal information correlation is crucial for in-depth understanding of disease characteristics. By fusing the CT image features of different phases, the generalization performance of the deep learning network and the understanding depth of the lesion features can be effectively improved. However, the existing studies [8, 15] pay less attention to the connection between CT images of different phases, which may cause the loss of key information and increase the one-sided understanding of the lesion development trend of the model. (3) Most current studies [8, 18–20] focus on using CT slices for analysis; however, this method has significant limitations in capturing the integrity of spatial information. Although CT slices can provide key cross-sectional information, they lack a comprehensive description of the 3D structure of the tumor, which makes it difficult for deep learning networks to fully learn the distribution and morphological changes of liver tumors in 3D space. To overcome this limitation, some researchers have proposed using 3D CNN to process the entire 3D image, but this method has high demand on computing resources and memory. In the case of limited resources, it will affect the training efficiency of the model and the feasibility of practical applications.

To address the limitations existing in the above methods, we draw inspiration from [21, 22] and propose a four-phase CT lesion recognition framework based on multiphase information fusion and a spatiotemporal prediction module. To achieve higher prediction accuracy and reduce model parameters, we first use ResNet18 to extract features from four-phase CT images. Then, we introduce an interactive perception mechanism to

enhance the feature extraction process and realize the tradeoff of channel-spatial information between different phase CT image features, so as to effectively fuse the multi-phase CT image features. Then, the one-dimensional deep residual network [14] was used to integrate the one-dimensional features of multi-phase, and the GRU architecture [23] was used to solve the problem of incompleteness of the model in capturing spatial information, and the whole network was trained based on the multi-task learning framework. Finally, the average results of the single-phase and the prediction results of the prediction module were weighted to obtain the final result.

Related work

This section mainly describes the related work of the proposed method, mainly from the following aspects, including: (1) the application of CNNs in multi-phase CT images; (2) the application of multi-phase information fusion in multi-phase CT; (3) the application of time sequence method in four-phase CT.

The application of CNNs in multi-phase CT images

CNNs are a deep learning neural network architecture specifically designed to process and analyze data with a network structure, such as images and videos. Convolutional neural networks are also widely utilized in the diagnosis of liver malignant tumors. Previous studies, such as STIC [24], apply VGG16 as a feature extractor to CT images to extract rich spatial features from them. These feature maps are subsequently input into the Temporal-Encoder module to further extract the key information embedded in the time series. In another study [10], 3D ResNet18 was used as the basic network to explore the 3D structural information of four-phase CT for diagnosis. In this study, we choose 2D ResNet18 as the feature extraction network. Compared with VGG16, 2D ResNet18 can not only efficiently extract detailed spatial features from 2D CT slices but also achieve significant optimization in model complexity and the number of parameters, thereby improving training efficiency and inference speed. Meanwhile, compared with 3D ResNet18, although the latter is able to capture 3D spatial information, it may encounter higher computational costs when dealing with large-scale datasets.

Multi-phase information fusion

The diagnosis of liver malignant tumors using multi-phase CT is a challenging task, and the core difficulty is to accurately integrate the information between each phase to achieve accurate diagnostic results. At present, multimodal data fusion methods are mainly divided into two categories: image-level and feature-level. Among them, feature-level fusion has attracted much attention because it can deeply mine the internal relationship between data. Studies [9, 25] have successfully demonstrated how to use multi-stage information at the feature level to classify or segment multi-phase CT images, which significantly improves performance. Wang et al. [26] directly connected brain MRI images as the input of the model to exploit multimodal information. In this paper, a multiphase information fusion framework for four-phase CT is proposed to further improve the diagnostic accuracy of liver malignant tumors. In the feature extraction stage, by introducing an interactive perception mechanism, we can not only extract unique channel-spatial attention feature maps for each phase but also capture the

potential associations between adjacent phases. These feature maps are then effectively weighted into the feature representation of the current phase to achieve effective feature fusion of multiphase CT images. This process not only retains the unique information of each phase but also mines more comprehensive and in-depth features of liver lesions through the interaction and fusion between features, which provides strong support for the subsequent diagnosis of malignant tumors.

Application of time sequence method in four-phase CT

In four-phase CT imaging, due to the abundant temporal enhancement information contained in slices, the temporal modeling method can efficiently integrate and express the global temporal information. The modular deep learning framework proposed by STIC [24] captures the spatial features of CT images through Spatial-Extractor, and subsequently inputs these features into the Temporal-Encoder, and uses the Gated Recurrent Unit (GRU) for temporal modeling. The spatial and temporal features of four-phase CT are effectively extracted. BD-LSTM proposed by Liang et al. [27] performs well in the classification task of focal liver disease by simultaneously utilizing the contextual information of CT sequences. There are other temporal modeling methods that are widely used in the processing of spatio-temporal data. Dai et al. [28] proposed a method to explore and exploit higher-order spatio-temporal dynamics for long-term frame prediction. The proposed method achieves more accurate prediction of video frames by capturing high-order spatio-temporal dependencies. Unlike STIC, this work focuses more on long-term prediction and considers higher-order spatio-temporal dynamics rather than just temporally enhanced information between adjacent slices. UNIMEMnet [29] learns long-term motion and appearance dynamics through a unified memory network for video prediction. The proposed method combines a memory mechanism and a recurrent neural network, which is able to capture long-term dependencies and dynamic changes in videos. Although UNIMEMnet has achieved remarkable results in video prediction, its complex memory mechanism may not be suitable for the task of four-phase CT lesion recognition with a relatively small amount of data. PredRNN [30] is a recurrent neural network for spatio-temporal predictive learning. The proposed method stacks multiple RNN layers and introduces memory units to capture long-term dependencies in spatio-temporal data. Compared with STIC and the proposed method, PredRNN focuses more on the general spatio-temporal prediction task and is not optimized specifically for four-phase CT lesion recognition. In contrast, GRU may be more advantageous in dealing with such tasks due to its simpler structure and fewer parameters. In this paper, we innovatively adopt the GRU architecture to model the temporal enhancement information between CT slices. Different from STIC [24] method, we first use a one-dimensional deep residual network (1D Resent) to extract one-dimensional features from multiphase CT images, which not only contain rich spatial information, but also imply the change trend in time. Subsequently, we feed these 1D features into the GRU architecture to further model the spatio-temporal information in depth. This network design enables the model to capture the spatio-temporal dependencies between different slices more accurately, thus exhibiting higher accuracy and robustness in the liver malignancy classification task.

Experiment

Data collection

A journey of a thousand miles begins with a single step. Data collection is essential for the experiment. The dataset for this study comprises four-phase CT images and corresponding demographic information from the First Affiliated Hospital of Zhejiang University School of Medicine. In total, 398 cases of liver malignant tumors were collected, including 196 cases of intrahepatic cholangiocarcinoma (ICC) and 202 cases of hepatocellular carcinoma (HCC).

We obtained non-contrast, arterial, portal venous, and delayed-phase four-phase CT images using the GE Medical system and Philips CT scanning systems. In addition, the two CT systems used in this study included a total of five CT models, which could provide more valuable images and facilitated the full artificial intelligence process from scanning to reconstruction. Patient characteristic statistics and CT scan acquisition parameters are shown in Table 1.

Experimental detail

Data pre-processing

To maximize the quality of CT images, manual annotation was performed by two annotators using medical image processing software (ITK-SNAP) to create a 3D bounding box, which were then reviewed by a radiologist. These four-phase CT images were stored in DICOM format. We first resampled the voxels to 1 mm size, extracted the lesion and its surrounding 10 mm pixels using a 3D bounding box. Finally, the CT image was adjusted the CT image resolution of $64 \times 128 \times 128$.

Since the CT image slices are single-channel gray scale images, The pre-trained models trained on ImageNet cannot be directly employed. Some studies have resorted to duplicating the single-channel CT slices to create three-channel images. However, this approach does not add genuine color information and may result in redundant data.

In the field of medical images, especially when processing CT images, window width and window level are two important parameters used to adjust the contrast and

Table 1 Patient information and CT scan acquisition parameters

Patient number	398
Age (mean ± std)	59.95 ± 9.83
HCC/ICC	202/196
Male/female	295/103
CT model	Philips Brilliance iCT 256 GE Revolution EVO GE Optima CT540 GE Revolution CT GE Revolution Frontier
Ratio of CT model	298:34:14: 36:16
Convolution kernel	B, STANDARD
Ratio of kernel	298:100
Tube voltage (kV)	120,140,100,80
Tube current (mA)	351.12 ± 93.20
Pixel size (mm)	0.72 ± 0.05
Slice thickness (mm)	3.39 ± 0.97

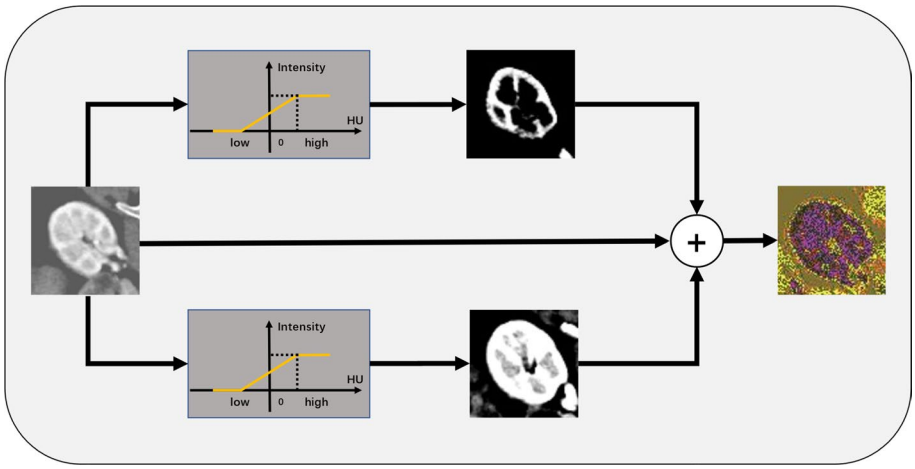

Fig. 1 Channel expansion processing

Table 2 Experimental environment

Name	Metrics
CPU	Intel(R) Core(TM) i9-10900X CPU@3.70GHZ
GPU	Tesla V100-SXM2 32 GB
Operating system	Ubuntu 18.04.6
CUDA	12.0
Programming languages	Python 3.6
Deep Learning Framework	Pytorch

brightness of the image. By utilizing combinations of two different window widths and window levels, we obtain grayscale image slices with varying contrast and brightness. These grayscale slices, processed with different window settings, are then superimposed onto the original data to form pseudo-color CT slices, resulting in three-channel images, this is shown in Fig. 1. Such processing can effectively utilize the fully trained pre-trained models on ImageNet. In this paper, corresponding data enhancement processing was performed on CT slices in different phases to learn CT image features more effectively.

Environment configuration

All experiments were done on the same workstation, whose hardware environment is shown in Table 2. To avoid problems caused by improper partitioning of the dataset and ensure the robustness and generalization ability of the model, we use five-fold cross validation. The dataset was divided into a training set and a test set with a ratio of 4:1. The experiments utilized Stochastic Gradient Descent (SGD) optimizer for parameter optimization, with an initial learning rate of 0.001. The learning rate was decayed by 0.1 every 10 epochs, and the total number of epochs was set to 50. The batch size was set to 2.

Table 3 Performance comparison of other model

Model	ACC (%)	AUC (%)	F1-scor e(%)	NPV (%)
3D ResNet-18	79.25	82.83	78.18	80.38
3D DenseNet-121	79.50	80.21	78.36	79.83
MexPale	78.75	81.59	77.84	80.06
STIC	77.75	78.55	72.18	82.31
TransMed	79.00	80.90	77.64	80.31
TDN	80.25	84.00	80.89	81.25
Ours	**85.5**	**89.73**	**85.44**	**86.35**

Table 4 Parameter comparison

Model	3D ResNet-18	3D DenseNet-121	MexPale	STIC	TransMed	TDN	Ours
Params (M)	33.21	45.01	19.01	132.53	120.55	24.06	**13.29**
FLOPs (G)	172.12	179.99	150.99	2586.09	347.58	755.03	**147.89**

Validation methods and metrics

This section will verify the effectiveness of the proposed method from four aspects: comparing with a 3D model, targeted ablation experiment, verifying the effectiveness of four-phase CT images, and the feasibility of pseudo-color for data enhancement. Evaluation metrics used in the experiments include Receiver Operating Characteristic (ROC) curve, Area Under the Curve (AUC) of the ROC curve, Accuracy (ACC), F1-score, and Negative Predictive Value (NPV).

Comparison with 3D CNN model

To demonstrate the effectiveness of our approach, we conducted comparisons with state-of-the-art methods, including 3D ResNet-18 [19], 3D DenseNet-121 [31], STIC [24], MexPale [10], TransMed [32] and the TDN [14, 33] model for video classification tasks. The comparative results are shown in Table 3, Table 4 compares the parameters of each model and the ROC curve is shown in Fig. 2.

Ablation experiment

Effectiveness of SPM, MIFF

In this section, to evaluate the effectiveness of the proposed multi-phase information fusion framework (MIFF), Spatiotemporal prediction module (SPM) and multi-task learning framework, we conduct a series of ablation experiments for validation. The ROC curve is shown Fig. 3. Including:

(1) Backbone: For each phase of CT images, employ 2D ResNet18 architecture for classification. Finally, achieve results through averaging fusion.

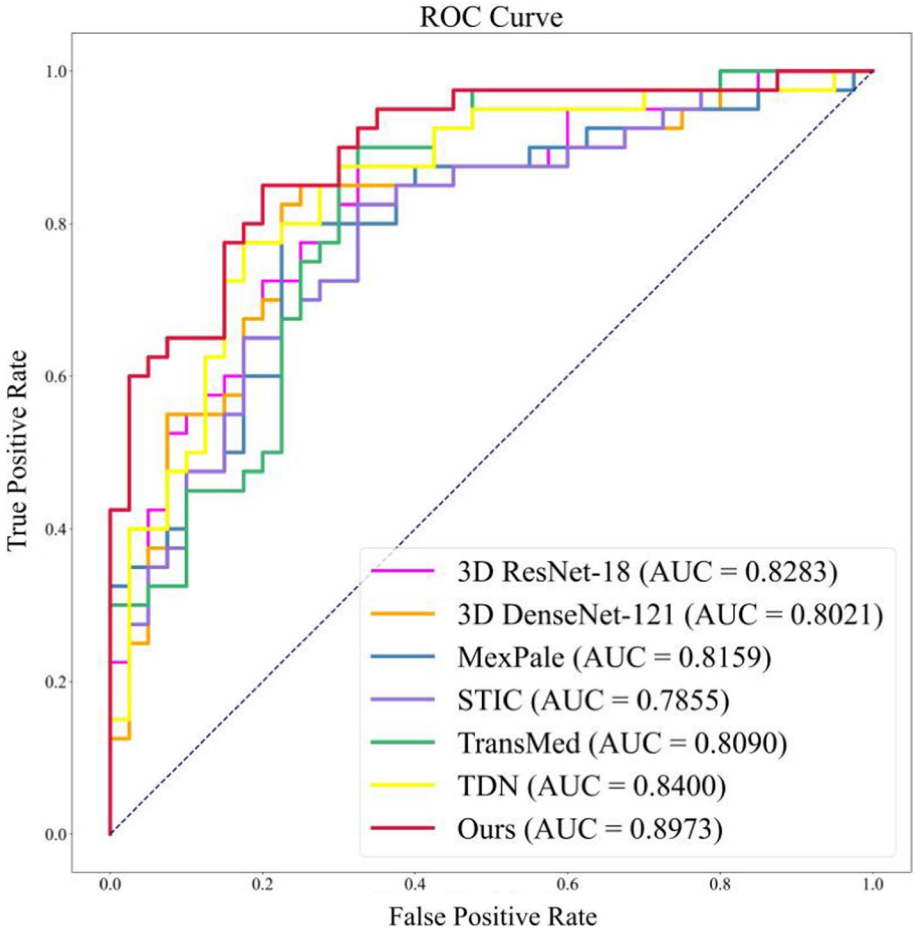

Fig. 2 ROC curves for different 3D CNN model

(2) Backbone + SPM: The 2D ResNet18 network is used for feature extraction, and then the spatiotemporal prediction module (SPM) is used for preliminary prediction to obtain preliminary classification results
(3) Backbone + MIFF: Incorporate an interactive perception mechanism into the 2D ResNet18, and subsequently obtain results through averaging fusion.
(4) Backbone + SPM + MIFF: The spatiotemporal prediction module is used to integrate the one-dimensional features, and finally the GRU classifier is used for classification.
(5) Backbone + SPM + MIFF + Multi-task Learning Framework: Our proposed model.

From Table 5, it can be observed that our proposed multi-phase information fusion framework, along with the spatiotemporal prediction module, significantly enhances the lesion recognition performance of the four-phase CT algorithm based on the Backbone. Compared to using Backbone alone, our approach shows improvements of 4% in accuracy (ACC), 3.61% in area under the curve (AUC), 4.09% in F1-score, and 4.98% in negative predictive value (NPV). This indicates a substantial performance enhancement of our model in the task of lesion recognition.

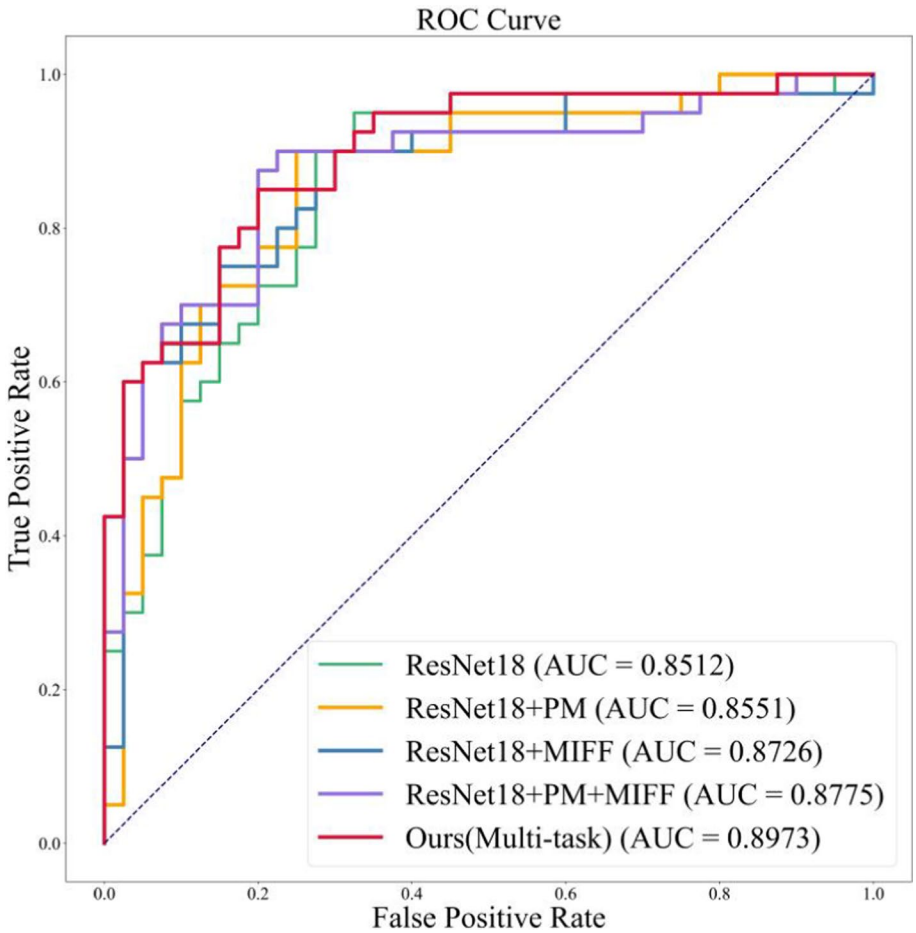

Fig. 3 ROC curves for the model

Table 5 The classification accuracy of a backbone with different modules

Model	ACC (%)	AUC (%)	F1-score (%)	NPV (%)
Backbone	81.50	85.12	81.35	81.73
Backbone + SPM	82.25	85.51	82.94	82.90
Backbone + MIFF	83.00	87.26	82.90	85.16
Backbone + SPM + MIFF	83.75	87.75	83.60	89.13
Ours (Multi-task)	**85.50**	**89.73**	**85.44**	**86.35**

The influence of λ on the diagnosis result

In the framework of multi-task learning, we consider the output of the spatio-temporal prediction module as the primary task and the output of the unimodal network as an auxiliary task. To assess the necessity of employing a multi-task learning framework and justify prioritizing the spatiotemporal prediction module output as the primary task, we conducted a series of comparative experiments. These

Table 6 The influence of λ

	ACC (%)	AUC (%)	F1-score (%)	NPV (%)
None	83.75	86.63	82.68	84.17
λ = 0.5	**85.50**	**89.73**	**85.44**	**86.35**
λ = 0.9	84.00	88.35	83.96	87.32

experiments aim to provide insights into the value of multi-task learning in model training and validate the rationale behind prioritizing prediction module output as our main objective. The results are shown in Table 6.

Effectiveness of the multiphase information fusion framework

The multi-phase information fusion framework successfully facilitates the interaction of information between phases by introducing the interactive perception mechanism. This mechanism calculates the channel and spatial weights of feature maps extracted from two adjacent phases, and then performs cross-modal feature weighting to promote effective information exchange between different modalities. Experimental results show that the proposed mechanism significantly improves the accuracy of diagnosis, and the Accuracy (ACC) and Area Under Curve (AUC) increasing by 3% and 5%, respectively. In addition, the experiment also shows that with the increase of the number of layers applying this mechanism, the diagnosis performance of the network improves further, the results are shown in Table 7.

Feasibility of channel extension

In this section, we primarily assess the feasibility of using pseudo-color for data augmentation. We configured two sets of window width parameters to obtain CT images with varying attenuation scales, the parameters are shown in Table 8. Specifically, our goal was to capture both the essential part of lesions and the information surrounding

Table 7 The effectiveness of the multiphase information fusion framework

Layer1	Layer2	Layer3	Layer4	ACC (%)	AUC (%)
				83.50	87.58
√				84.25	88.81
√	√			84.75	88.96
√	√	√		85.25	89.54
√	√	√	√	**85.50**	**89.73**

Table 8 Different combinations of windows widths and window levels

	CT pattern 1		CT pattern 2	
	WL	WW	WL	WW
NP	65 HU	30 HU	40 HU	20 HU
AP	55 HU	70 HU	50 HU	40 HU
PVP	150 HU	30 HU	80 HU	100 HU
DP	175 HU	50 HU	75 HU	80 HU

Table 9 Comparison of original image and pseudo-color image results

	ACC (%)	AUC (%)	F1-score (%)	NPV (%)
Original data	81.50	85.72	81.38	79.35
Multichannel data (ours)	8**5.50**	**89.73**	**85.44**	**86.35**

Table 10 Comparison of different phase combinations

Model	ACC (%)	AUC (%)	F1-score (%)	NPV (%)
NC + ART + PV	84.00	87.57	83.95	85.68
NC + ART + DL	83.50	88.69	83.48	84.66
Ours	**85.50**	**89.73**	**85.44**	**86.35**

the lesion are. Compared to the original single-channel data, the expanded three-channel images more clearly depict the shape of the lesions, the results are shown in Table 9.

Comparison of three and four-phase CT images

Many studies [8] 11–13 have utilized three-phase CT images, lacking either the PV or DL phases, whereas four-phase CT images provide more comprehensive feature information than three-phase CT images. This is particularly beneficial for diseases involving dynamic temporal changes. Through comparative experiments, we have demonstrated the necessity of using four-phase CT images. We employed two sets of different three-phase data: one comprising NC, ART, PV, and the second comprising NC, ART, DL, the results are shown in Table 10. This design allows us to compare the imaging features of different phases to determine whether four-phase CT images provide more detailed and comprehensive information in specific disease scenarios. Figure 4 illustrates our experimental results.

Discussion

As a powerful machine learning method, deep learning has made significant progress in the field of computer-aided cancer diagnosis. In medical image analysis, deep learning models can identify, extract and learn features at various levels from different medical images, and provide more accurate and efficient diagnostic insights for medical experts.

In this study, we propose a four-phase CT lesion recognition algorithm based on a multi-phase information fusion framework and a spatiotemporal prediction module for the classification task of hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC). Our model demonstrated significant performance on the four-phase CT dataset, with an ACC of 85.5% and an AUC of 89.73 in the test cohort, outperformed some 3D models. These results indicate the high reliability of our algorithm in distinguishing between these two types of liver cancer lesions.

Most previous studies [8, 11] have used manually selected 2D CT slices as input to the network. However, because lesions are 3D structures in CT images, a single CT slice cannot adequately capture the spatial information of the lesion. The use of 3D CNNs

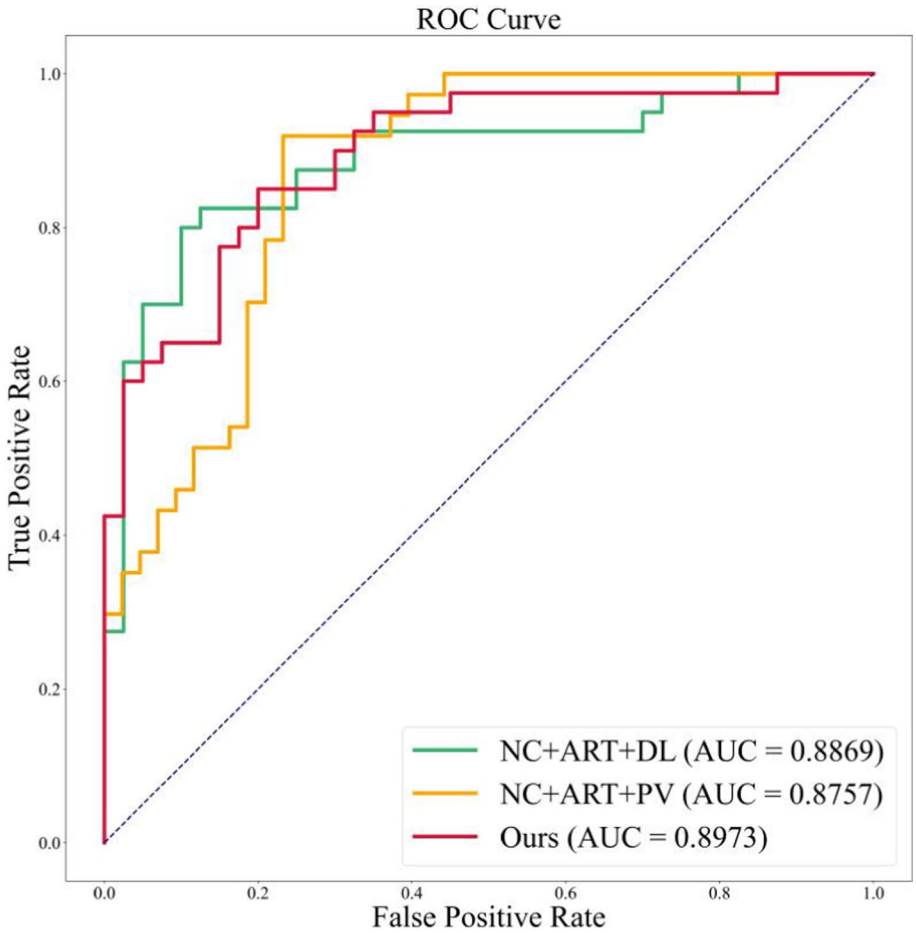

Fig. 4 ROC curves for different combinations of phases

has shown promising results in liver cancer classification [10, 31, 34]. However, the large number of parameters can lead to excessively long training times and increased computational costs. To overcome these challenges, we use 2D ResNet-18 as the base network and enhance the model with the multi-phase information fusion framework and a spatio-temporal prediction module. In our method, since the size and location of tumors may be different in CT images of different phases, we introduce a multi-phase information fusion framework in the feature extraction stage, in which an interactive perception mechanism is adopted. Through this mechanism, we realize the channal-space interaction between phases to ensure the effective fusion of different phase information. It makes it easier for the model to notice the location of the lesion. Given that 2DCNNs struggle to capture temporal enhancement information between CT image slices, we propose a spatio-temporal prediction module, which uses a one-dimensional deep residual network for integrating multi-phase feature vectors and uses a GRU architecture to solve the problem of model incompleteness in capturing spatial information.

The comparison experiment with 3D CNNs proves the superiority of the proposed model. The key difference between this method and 3D CNNs is that the dynamic change information of the lesion across CT slices is incorporated so that the model can learn the characteristics of the lesion more accurately. Zhou et al. [11] proposed that 2D

CNNs ignore the spatial discontinuity between slices, and the spatio-temporal prediction module effectively addresses this issue by using GRU architecture to model the temporal enhancement information between CT slices. ConvLSTM can effectively capture spatial features in sequence through convolution operations, which makes it perform well in spatio-temporal sequence prediction tasks. However, in the case of limited datasets, such as in the four-phase lesion recognition task, GRU is more appropriate for spatio-temporal modeling. GRU has a simple structure and fewer parameters, so it can still maintain good performance with less data and resources. The advantage of this model is that it maintains the efficiency of 2D CNNs while having the same ability to model spatio-temporal features as 3D CNNs.

In addition, we also focused on data augmentation. By using pseudo-color processing techniques on CT images, we not only successfully enriched the dataset, but also utilized model parameters based on transfer learning. Yamashita et al. [19] have validated the superiority of networks trained using transfer learning over those trained from scratch. Furthermore, the two CT systems we used had five CT models, which enhanced the diversity of CT scans. In model training, we employed a multi-task learning framework, deepening the model's understanding across various aspects simultaneously, further optimized the performance of the network, and thus significantly improved the performance in the multi-phase CT image classification task.

Despite our models achieving significant results, there is still room for improvement in several aspects. Firstly, the relatively small size of our dataset might contribute to overfitting, potentially limiting the model's performance, considering more data for training and validation is a direction worth exploring. Second, this study concentrated solely on the classification of HCC and ICC, excluding other disease types such as liver metastases and hepatoblastoma. Introducing more diverse data and encompassing a broader range of disease categories could further enhance the model's robustness and applicability.

Conclusion

In this study, we propose an innovative four-phase CT lesion recognition algorithm for the diagnosis of liver malignancies. Experimental results demonstrate the effectiveness of this method in diagnosing multi-phase liver malignant tumors. This research introduces a novel approach to advancing liver lesion diagnostics, holding significant potential for clinical application in clinical practice. Future directions include further validating the algorithm's robustness, expanding the scope of study subjects, and optimizing the algorithm to adapt to a broader range of clinical scenarios.

Methods

Our proposed model, as shown in Fig. 5, employs a 2D CNN ResNet18 network as the foundational architecture for effectively extracting features from four-phase CT images. We introduce an interactive perception mechanism into each block of ResNet18 to form a multi-phase information fusion framework. This mechanism enables channel-spatial information interactive weighting across multi-phase and multi-dimensional features, thereby enhancing the extraction of CT image features.The prediction module utilizes a 1D deep residual network for initial result prediction, which is well-suitable for the feature extraction task with four-phase CT images. To further improve the model's overall

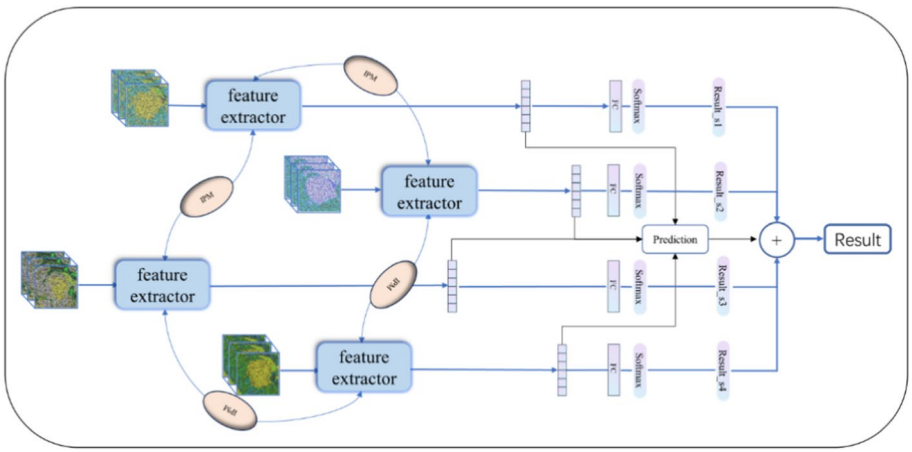

Fig. 5 Overall framework of the model (IPM is the interactive perception mechanism)

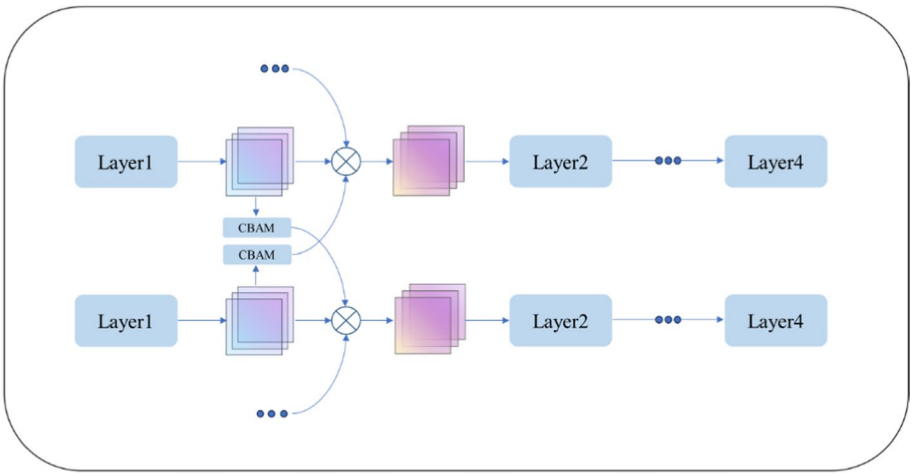

Fig. 6 Multi-phase information fusion framework (MIFF)

efficiency and generalization ability, we employ a multi-task learning framework to train the entire network, enabling it to handle multiple related tasks simultaneously.

It is crucial to note that the input images$X \in R^{B\times T\times C\times H\times W}$,where B represents Batch Size, T signifies the number of CT slices, C indicates the number of channels of CT slices, H and W represents the height and width of CT slices, respectively. As we employ a 2D CNN, certain adjustments are made: multiplying dimensions B and T, transforming the 5D (B, T, C, H, W) into 4D$(B \times T, C, H, W)$. Furthermore, the first convolutional layer of ResNet18 is modified by adjusting the stride to 2. In the single-phase classification task, we introduce a time-series average pooling layer after the fully connected layer of the network, and the classification results of all CT slices in that phase were averaged over the time-series dimension. Finally, the average results of the single-phase and the prediction results from the prediction module are weighted to obtain the final results.

Multi-phase information fusion framework (MIFF)

The features, sizes, and positions of tumors may vary across different phases, making the interaction of information between multi-phase CT images crucial for the task of liver malignant tumor classification. However, many previous studies overlooked this aspect, and some simply add the image features of different phases, with some simply performing a straightforward addition of features from different phases. To address this issue, we introduce a new multi-phase information fusion framework, as shown in Fig. 6, which includes an interactional perception mechanism. The key task of this mechanism is to establish channel-spatial information interactions between different phases to emphasize meaningful features. Deviating from the direct use of a ResNet [10] network for feature extraction, our study employs a convolutional attention module. Taking NC and ART phases as examples, we use CBAM [35] to extract the channel and spatial dependence of two-dimensional features from CT images. Then, the channel-spatial weights of NC and PV phase image features are applied to the ART phase image features. Similarly, we applied the channel-spatial weights of ART and DL phase image features to the NC phase image features, which enabled the model to more intelligently select the information that plays a decisive role in tumor classification. In addition, it is noteworthy that this interactive perception mechanism is not limited to a single layer of the ResNet18 network, weight calculation and channel-spatial information interaction weighting are performed on the output features of different layers to more fully exploit the hidden information from different layers. This contributes to enhancing the model's classification performance, enabling more accurate identification of malignant liver tumors. This process can be expressed by the following formula (using NC and ART phases as examples):

$$\begin{cases} \Theta_O^{\mathrm{NC}} = \Theta_I^{\mathrm{NC}} \otimes W_C^{\mathrm{ART}} + \Theta_I^{\mathrm{NC}} \otimes W_S^{\mathrm{ART}} + \Theta_I^{\mathrm{NC}} \otimes W_C^{\mathrm{DL}} + \Theta_I^{\mathrm{NC}} \otimes W_S^{\mathrm{DL}} \\ \Theta_O^{\mathrm{ART}} = \Theta_I^{\mathrm{ART}} \otimes W_C^{\mathrm{NC}} + \Theta_I^{\mathrm{ART}} \otimes W_S^{\mathrm{NC}} + \Theta_I^{\mathrm{ART}} \otimes W_C^{\mathrm{PV}} + \Theta_I^{\mathrm{ART}} \otimes W_S^{\mathrm{PV}} \end{cases}, \quad (1)$$

Θ_I^{NC} and Θ_I^{ART} represent 2D features extracted by each layer of ResNet18 for the NC and ART phases. W_C^{ART} and W_S^{ART} are the channel attention weights and spatial attention weights for the ART phase. Similarly, W_C^{NC} and W_S^{NC} are the channel attention weights and spatio-temporal attention weights of the NC phase, as are W_C^{DL} and W_S^{DL} and W_C^{PV} and W_S^{PV}, as shown in Eq. 1. This interactive perception mechanism is not only applied to the two phases of NC and ART but involves four phases, in which there is an interactive weighting of channel-spatial information between each of the two phases.

In summary, the introduction of an interactive perception mechanism within the multi-phase information fusion framework enables a more effective handling of variations in the features, sizes, and positions of liver tumors across different phases. This significantly improves the classification performance of the model, which holds great significance for early diagnosis and treatment planning of liver malignancies.

Spatiotemporal prediction module (SPM)

Utilizing the temporal enhancement information between the four-phase CT slices can significantly improve the classification of malignant liver tumors. Simply averaging the results of the four phases and using it as the output for classification does not fully exploit the temporal enhancement information between slices. In this study, by

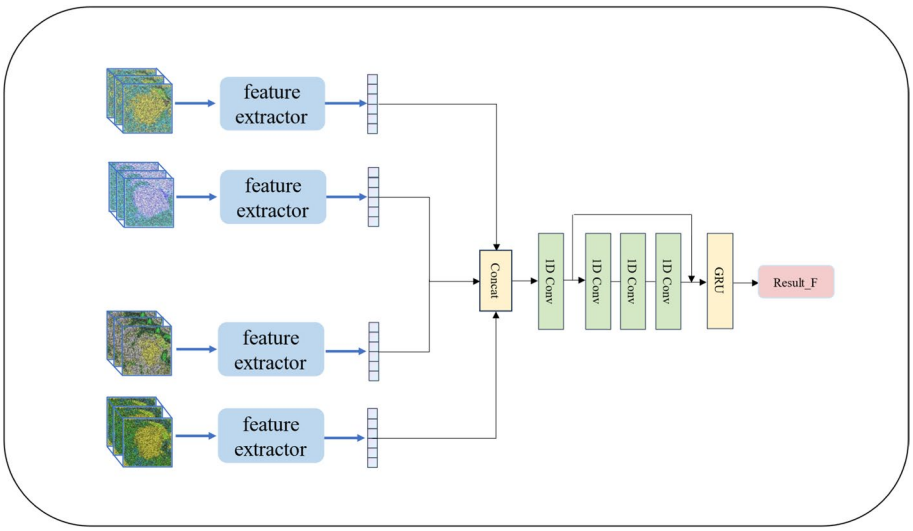

Fig. 7 Spatio-temporal prediction module (SPM)

introducing one-dimensional feature fusion and a GRU architecture, which helps synthesize the spatiotemporal information of different phases, and effectively improve the classification accuracy of liver malignant tumors. To achieve this, we propose a spatiotemporal prediction module that can effectively integrate the 1D features of four-phase CT images and effectively model the dynamic changes between CT slices using the GRU architecture.

As shown in Fig. 7, we initially concatenate the one-dimensional features of the four-phase CT images along the feature dimension, thereby establishing a more enriched and comprehensive feature representation. We then use the GRU architecture to model the temporal enhancement information between CT slices, assisting the model in better leveraging the dynamic changes in the sequence of images to obtain preliminary predictions.

Multi-task learning framework

To ensure the accurate classification of four-phase CT images, we trained the network using a multi-task learning framework. In this multi-task learning framework, the primary task involves making preliminary predictions on the four-phase CT images using the prediction module, while the auxiliary task involves the output from the single-phase networks. The loss function of the multi-task learning framework can be expressed as follows:

$$L = E\left\{(1-\lambda)X_f + \frac{\lambda\left(X_n + X_a + X_p + X_d\right)}{4}, y\right\}, \tag{2}$$

where $E(x, y)$ is the cross-entropy loss, λ is the balance factor of multi-task learning, X_f is the output of the prediction module, and X_n, X_a, X_p, X_d, represent the output of the single-mode network in four phases, respectively.

Author contributions
Q and X wrote the main manuscript and Zuo, Zheng, Jiang and Zeng provided assistance.

Funding
This research received no external funding.

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 9 May 2024 Accepted: 2 October 2024
Published online: 21 October 2024

References

1. Ganesan P, Kulik LM. Hepatocellular carcinoma: new developments. Clin Liver Dis. 2023;27(1):85–102.
2. Blachar A, Federle MP, Ferris JV, Lacomis JM, Waltz JS, Armfield DR, et al. Radiologists' performance in the diagnosis of liver tumors with central scars by using specific CT criteria. Radiology. 2002;223(2):532–9.
3. Lakshmipriya B, Pottakkat B, Ramkumar G. Deep learning techniques in liver tumour diagnosis using CT and MR imaging—a systematic review. Artif Intell Med. 2023. https://doi.org/10.1016/j.artmed.2023.102557.
4. Shi W, Kuang S, Cao S, Hu B, Xie S, Chen S, et al. Deep learning assisted differentiation of hepatocellular carcinoma from focal liver lesions: choice of four-phase and three-phase CT imaging protocol. Abdominal Radiol. 2020;45:2688–97.
5. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44.
6. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst. 2021;33(12):6999–7019.
7. Reyes M, Meier R, Pereira S, Silva CA, Dahlweid FM, Tengg-Kobligk HV, et al. On the interpretability of artificial intelligence in radiology: challenges and opportunities. Radiol Artif Intell. 2020;2(3): e190043.
8. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. Radiology. 2018;286(3):887–96.
9. Liang D, Lin L, Hu H, Zhang Q, Chen Q, Iwamoto Y, et al. Residual convolutional neural networks with global and local pathways for classification of focal liver lesions. In PRICAI 2018: Trends in Artificial Intelligence: 15th Pacific Rim International Conference on Artificial Intelligence, Nanjing, China, August 28–31, 2018, Proceedings, Part I 15. Springer International Publishing; 2018. pp. 617–628.
10. Ling Y, Ying S, Xu L, Peng Z, Mao X, Chen Z, et al. Automatic volumetric diagnosis of hepatocellular carcinoma based on four-phase CT scans with minimum extra information. Front Oncol. 2022;12: 960178.
11. Zhou J, Wang W, Lei B, Ge W, Huang Y, Zhang L, et al. Automatic detection and classification of focal liver lesions based on deep convolutional neural networks: a preliminary study. Front Oncol. 2021;10: 581210.
12. Todoroki Y, Iwamoto Y, Lin L, Hu H, Chen YW. Automatic detection of focal liver lesions in multi-phase CT images using a multi-channel & multi-scale CNN. In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2019. pp. 872–875.
13. Kim DW, Lee G, Kim SY, Ahn G, Lee JG, Lee SS, et al. Deep learning–based algorithm to detect primary hepatic malignancy in multiphase CT of patients at high risk for HCC. Eur Radiol. 2021;31:7047–57.
14. Kaiming H, **angyu Z, Shaoqing R, Jian S. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. Vol. 34, pp. 770–778.
15. Xue M, Jiang H, Zheng J, Wu Y, Xu Y, Pan J, Zhu W. Spatiotemporal excitation module-based CNN for diagnosis of hepatic malignancy in four-phase CT images. In 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2023. pp. 1–5.
16. Wang W, Chen Q, Iwamoto Y, Aonpong P, Lin L, Hu H, et al. Deep fusion models of multi-phase CT and selected clinical data for preoperative prediction of early recurrence in hepatocellular carcinoma. IEEE Access. 2020;8:139212–20.
17. Xu X, Zhu Q, Ying H, Li J, Cai X, Li S, et al. A knowledge-guided framework for fine-grained classification of liver lesions based on multi-phase CT images. IEEE J Biomed Health Inform. 2022;27(1):386–96.
18. Hamm CA, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. Eur Radiol. 2019;29:3338–47.
19. Turnbull R. Using a 3D ResNet for detecting the presence and severity of COVID-19 from CT Scans. In European Conference on Computer Vision. Cham: Springer Nature Switzerland; 2022. pp. 663–676.
20. Yamashita R, Mittendorf A, Zhu Z, Fowler KJ, Santillan CS, Sirlin CB, et al. Deep convolutional neural network applied to the liver imaging reporting and data system (LI-RADS) version 2014 category classification: a pilot study. Abdominal Radiol. 2020;45:24–35.

21. Xue M, Xu Z, Qiao S, Zheng J, Li T, Wang Y, Peng D. Driver intention prediction based on multi-dimensional cross-modality information interaction. Multimed Syst. 2024;30(2):83.
22. Xue M, Zheng J, Li T, Peng D. CLS-Net: an action recognition algorithm based on channel-temporal information modeling. Int J Pattern Recogn Artif Intell. 2023;37(08):2356011.
23. Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. 2014. arXiv preprint arXiv:1412.3555.
24. Gao R, Zhao S, Aishanjiang K, Cai H, Wei T, Zhang Y, et al. Deep learning for differential diagnosis of malignant hepatic tumors based on multi-phase contrast-enhanced CT and clinical data. J Hematol Oncol. 2021;14:1–7.
25. Zhang Y, Yang J, Tian J, Shi Z, Zhong C, Zhang Y, He Z. Modality-aware mutual learning for multi-modal medical image segmentation. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. Springer International Publishing; 2021. pp. 589–599.
26. Wenxuan W, Chen C, Meng D, Hong Y, Sen Z, Jiangyun L. Transbts: multimodal brain tumor segmentation using transformer. In International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer; 2021. pp. 109–119.
27. Liang D, Lin L, Hu H, Zhang Q, Chen Q, Iwamoto Y, et al. Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11. Springer International Publishing; pp. 666–675.
28. Dai K, Li X, Ye Y, Wang Y, Feng S, Xian D. Exploring and exploiting high-order spatial-temporal dynamics for long-term frame prediction. IEEE Trans Circ Syst Video Technol. 2023;34(3):1841–1856
29. Dai K, Li X, Luo C, Chen W, Ye Y, Feng S. UNIMEMnet: learning long-term motion and appearance dynamics for video prediction with a unified memory network. Neural Netw. 2023;168:256–71.
30. Wang Y, Wu H, Zhang J, Gao Z, Wang J, Philip SY, Long M. Predrnn: a recurrent neural network for spatiotemporal predictive learning. IEEE Trans Pattern Anal Mach Intell. 2022;45(2):2208–25.
31. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 4700–4708.
32. Dai Y, Gao Y, Liu F. Transmed: transformers advance multi-modal medical image classification. Diagnostics. 2021;11(8):1384.
33. Wang L, Tong Z, Ji B, Wu G. Tdn: temporal difference networks for efficient action recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. pp. 1895–1904.
34. Tanaka T, Huang Y, Marukawa Y, Tsuboi Y, Masaoka Y, Kojima K, et al. Differentiation of small ($\leq$ 4 cm) renal masses on multiphase contrast-enhanced CT by deep learning. Am J Roentgenol. 2020;214(3):605–12.
35. Woo S, Park J, Lee JY, Kweon IS. Cbam: convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pp. 3–19.

Publisher's Note