A flexible deep learning framework for liver tumor diagnosis using variable multi-phase contrast-enhanced CT scans

Shixin Huang[1,2] · Xixi Nie[3] · Kexue Pu[4] · Xiaoyu Wan[2] · Jiawei Luo[5]

Abstract

Background Liver cancer is a significant cause of cancer-related mortality worldwide and requires tailored treatment strategies for different types. However, preoperative accurate diagnosis of the type presents a challenge. This study aims to develop an automatic diagnostic model based on multi-phase contrast-enhanced CT (CECT) images to distinguish between hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (ICC), and normal individuals.

Methods We designed a Hierarchical Long Short-Term Memory (H-LSTM) model, whose core components consist of a shared image feature extractor across phases, an internal LSTM for each phase, and an external LSTM across phases. The internal LSTM aggregates features from different layers of 2D CECT images, while the external LSTM aggregates features across different phases. H-LSTM can handle incomplete phases and varying numbers of CECT image layers, making it suitable for real-world decision support scenarios. Additionally, we applied phase augmentation techniques to process multiphase CECT images, improving the model's robustness.

Results The H-LSTM model achieved an overall average AUROC of 0.93 (0.90, 1.00) on the test dataset, with AUROC for HCC classification reaching 0.97 (0.93, 1.00) and for ICC classification reaching 0.90 (0.78, 1.00). Comprehensive validation in scenarios with incomplete phases was performed, with the H-LSTM model consistently achieving AUROC values over 0.9.

Conclusion The proposed H-LSTM model can be employed for classification tasks involving incomplete phases of CECT images in real-world scenarios, demonstrating high performance. This highlights the potential of AI-assisted systems in achieving accurate diagnosis and treatment of liver cancer. H-LSTM offers an effective solution for processing multi-phase data and provides practical value for clinical diagnostics.

Keywords Liver tumor · Contrast-enhanced CT scans · Diagnostic model · Deep learning · Feature integration

✉ Xiaoyu Wan
wanxy@cqupt.edu.cn

✉ Jiawei Luo
jiaweiluo@wchscu.cn

1 Department of Scientific Research, The People's Hospital of Yubei District of Chongqing city, Chongqing 401120, China

2 School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

3 School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

4 School of Medical Informatics, Chongqing Medical University, Chongqing 400016, China

5 West China Biomedical Big Data Center, Med-X Center for Informatics, West China Hospital, Sichuan University, Chengdu 610044, China

Introduction

According to global cancer statistics for the year 2020, liver cancer ranked as the sixth most prevalent cancer worldwide and was the third leading cause of cancer-related fatalities (Sung et al. 2021). A significant proportion of malignant liver tumors are primary, including Hepatocellular Carcinoma (HCC) and Intrahepatic Cholangiocarcinoma (ICC) (Siegel et al. 2020; Rumgay et al. 2022). Hepatocellular Carcinoma is the most common primary liver cancer, while Intrahepatic Cholangiocarcinoma accounts for approximately 15–20% of primary liver cancer cases (Mcglynn et al. 2021). Although ICC represented a relatively small percentage of primary liver cancers, it could directly threaten the patient's life when metastasis occurred. Treatment strategies for different liver tumor subtypes varied considerably (Petrowsky et al. 2020), and the use of multi-phase

Contrast-Enhanced Computed Tomography (CECT) had become a primary diagnostic tool for liver tumor evaluation prior to surgery (Ayuso et al. 2018). However, accurately distinguishing malignant liver tumors presented substantial challenges, and preoperative misdiagnoses could lead to inappropriate treatment decisions. There was a growing need to develop an automated diagnostic model capable of assisting physicians in liver tumor diagnosis, reducing interobserver variability, and enhancing diagnostic efficiency.

Liver cancer CECT scans typically consisted of four phases, including the plain phase (P), arterial phase (C1), venous phase (C2), and delayed phase (C3). However, in real medical scenarios, patients often underwent examinations for only a subset of these phases. The main reasons included: (1) Multiple scans could increase a patient's radiation exposure. To ensure patient safety, physicians might choose the most appropriate scan phase based on clinical requirements, thereby reducing unnecessary radiation exposure. (2) Performing scans in multiple contrast-enhanced phases could increase scan duration and cost. In certain situations, for the sake of efficiency and cost-effectiveness, only the most critical contrast-enhanced phases might be conducted. (3) Some patients might not tolerate extended scan times or might experience contrast-related allergic reactions or other complications. Additionally, the decision regarding which contrast-enhanced phases to perform depended on various factors, including the patient's specific clinical needs and the balancing of potential risks and benefits. Physicians selected the most suitable scanning strategy based on these factors. As a result, the choice of CECT phases for each patient was variable. Consequently, when constructing a classification network for liver cancer, it's crucial to account for the variability of the number of CECT layers and the stages of CECT. In this study, we aim to develop a diagnostic model that handles variable numbers of phases and image layers in realworld CECT scans while capturing the 3D spatial structure of the liver to improve diagnostic accuracy. We hypothesize that our hierarchical LSTM network will effectively classify liver cancer with incomplete and variable-phase data, outperforming methods like 3D-ResNet and Transformerbased models.

Image classification models based on deep learning play a pivotal role in Computer-Aided Diagnosis (CAD), with numerous studies demonstrating that deep learning algorithms surpass traditional methods (Jakimovski et al. 2019; Yoo and Baek 2018; Doğantekin et al. 2019). Convolutional Neural Networks (CNNs) have shown immense potential in adaptive feature extraction and image classification tasks (Sarvamangala and Kulkarni 2022; Zhang 2022; Bakrania et al. 2023). Many studies (Chen 2019; Yasaka et al 2018; Ponnoprat et al. 2020; Zhou and Wang 2021; Shanmugapriya et al 2022; Romero 2019) rely on manually selected single-layer CT images for classification, necessitating clinician involvement and not fully leveraging the liver's 3D spatial structure. However, they have demonstrated the exceptional feature extraction capabilities of networks like Inception-V3 and ResNet for liver CT images. Ling et al. (2022) utilized a 3D-ResNet to classify complete quadriphasic CECT images. They first resampled and linearly scaled each phase of the CECT to a uniform size and then treated each phase as a separate channel, combining the phased images. However, this model can only handle a fixed number of phases and is not equipped to deal with variable phase numbers. Gao et al. (2021) introduced Long Short-Term Memory (LSTM) to integrate image data from different phases, utilizing LSTM's capability to process variable sequence data (Lipton 2015; Hochreiter et al. 1997). Yet, for each phase, they still opted for single-layer CT images. Wang et al. (2023) employed a Transformer architecture to integrate images from different phases of CECT. However, the Transformer, relying primarily on position embedding, struggles to capture the sequential structure inherent in 3D CECT images.

In this study, we address the issues of variable layer numbers and phases in CECT, along with the need to capture the 3D spatial structure of the liver, by introducing a hierarchical LSTM network (H-LSTM). The model first extracts features from single-layer CT images using a shared pretrained feature extraction network across different phases. Then, a phase-specific bidirectional LSTM (BiLSTM) integrates image features from various layers, capturing distinct spatial patterns within each phase. To help the model understand the phase information, we use one-hot encoding, embed the phases, and concatenate them with image features. Another BiLSTM network then fuses features across phases to handle variability. Our experiments show that this hierarchical LSTM network effectively classifies liver cancer in complex, variable-phase CECT images. We conducted ablation studies to test performance across different phase combinations, comparing feature extractors and substituting the BiLSTM with a Transformer network. The contributions of this paper are as follows:

- We developed a model capable of handling variablephase CECT images for liver cancer classification, managing irregularities and missing phases in real-world data.
- We introduced a hierarchical BiLSTM structure designed to capture the structural relationships between different layers of CT images as well as the correlations between CT images across various phases.
- We demonstrated that in scenarios involving 3D CECT images with a variable number of layers and phases, the hierarchical BiLSTM network outperforms the direct application of the 3D-ResNet algorithm for the classification of liver cancer.

Methods

Problem formulation

For a given patient i, the diagnostic prediction is based on a combination of images from the four available phases: plain phase $(X_{i,1})$, arterial phase $(X_{i,2})$, venous phase $(X_{i,3})$, and delayed phase $(X_{i,4})$. Since contrast-enhanced CT (CECT) images from the plain phase are always present during the diagnosis of primary liver cancers for comprehensive assessment of liver lesions, the input for diagnostic prediction must include the plain phase and at least one additional phase image. The diagnostic prediction outcome, y_i, can take values from the set $\{0, 1, 2\}$, representing normal individuals, hepatocellular carcinoma, and intrahepatic cholangiocarcinoma, respectively. For a detailed illustration of the diagnostic process and phase combinations, see Fig. 1.

Model structure

As illustrated in Fig. 2, our proposed H-LSTM network comprises an image feature extractor shared by all phases of 3D CECT images, four phase-specific intra-phase BiLSTM networks, one phase index embedding network, an inter-phase BiLSTM network, and an output layer. Each phase's 3D CECT is initially processed by the shared feature extractor for feature extraction. Subsequently, the extracted features of different tiers within each phase are integrated using the phase-specific BiLSTM. These integrated features are then concatenated with the output from the phase index embedding network, followed by the application of the inter-phase BiLSTM to aggregate features across different phases. Finally, the aggregated features are fed into the output layer for classification and prediction.

Image feature extractor

The purpose of the image feature extractor is to extract features from CECT images across different phases. We have selected ResNet (Sarwinda et al. 2021), VGGNet (Simonyan 2014), DenseNet (Huang 2017), InceptronNet (Szegedy 2015) and EfficientNet (Tan 2019) as candidate networks. The rationale behind choosing these well-known networks is the availability of a substantial number of pre-trained model parameters. Through transfer learning (Gao et al. 2020; Aslan et al. 2021), we can finetune these parameters to build a local model that is specific to our task on a limited dataset. Leveraging pre-trained model parameters is advantageous because these models have already learned to extract general image features, which can accelerate the convergence of our model and enhance its performance. Within each phase, we map the images $X_{i,j,k}$ to a
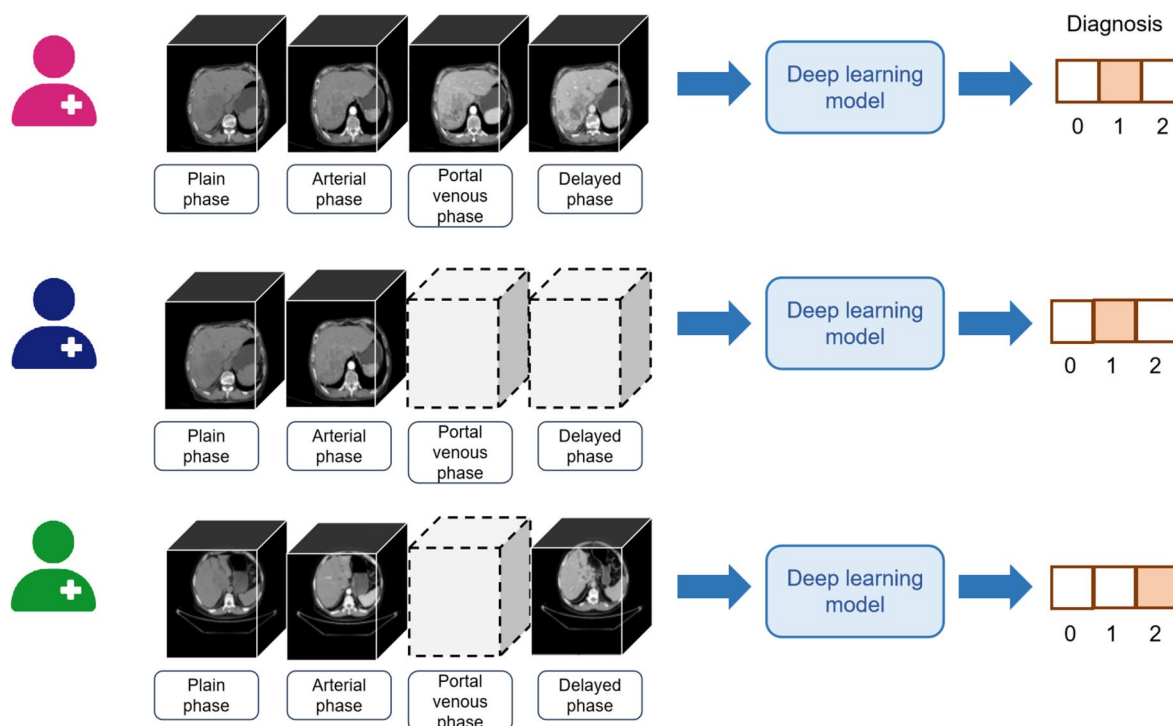

Fig. 1 Diagnosis prediction of patients based on variable multi-phase CECT. Dashed boxes indicate the absence of CECT images for that phase of the patient. 0, 1, and 2 represent individuals with normal conditions, HCC, and ICC, respectively
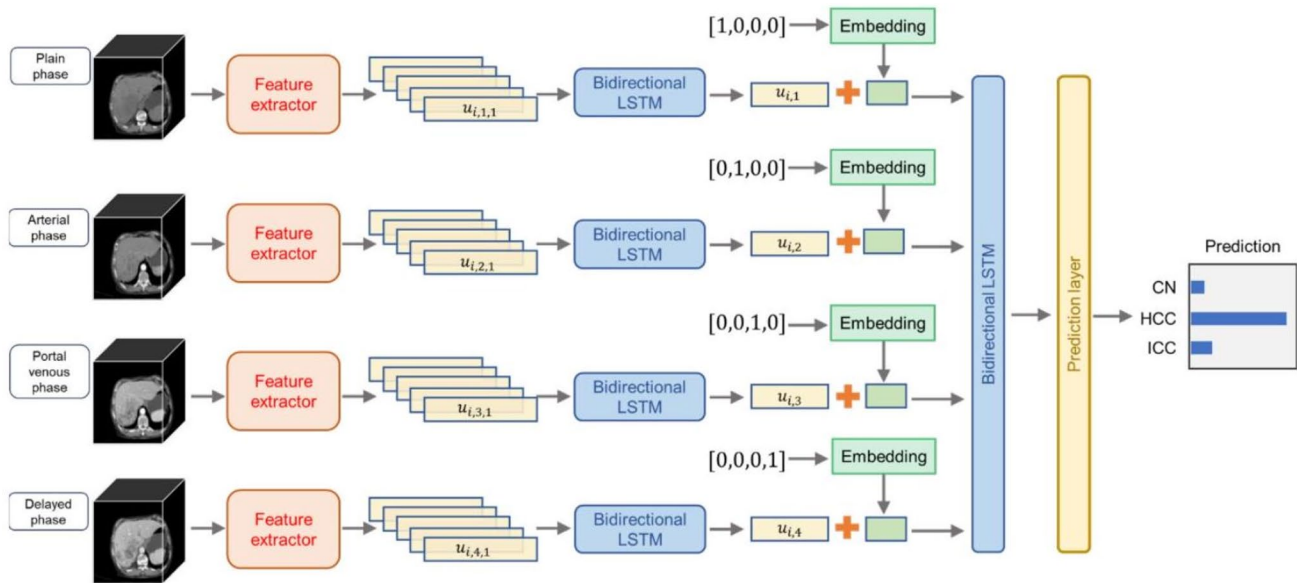
Fig. 2 Model Architecture Diagram. The figure illustrates the diagnostic classification process of the model for an individual with CECT images in all four phases. CN, HCC, and ICC respectively denote normal, Hepatocellular Carcinoma, and Intrahepatic Cholangiocarcinoma

vector $u_{i,j,k} \in \mathbb{R}^{d_x}$ using the feature extractor, as depicted in Eq. (1). f represents the feature extractor, and θ denotes its parameters.

$$u_{i,j,k} = f\left(X_{i,j,k};\theta\right) \tag{1}$$

Internal LSTM

Next, we address the aggregation of CECT images with varying numbers of layers using a BiLSTM. The forward computation process of the BiLSTM is detailed in Eqs. (2, 3). This sequential architecture aids in capturing the structural information of the liver stored within the CECT data.

LSTM-Forward:

$$
\begin{aligned}
f_k^f &= sigmoid\left(W_f^f x_k + U_f^f h_{k-1}^f + b_f^f\right), \\
i_k^f &= sigmoid\left(W_f^i x_k + U_f^i h_{k-1}^f + b_f^i\right) \\
\tilde{c}_k^f &= \tanh\left(W_f^g x_k + U_f^g h_{k-1}^f + b_f^g\right), \\
c_k^f &= f_k^f \circ c_{k-1}^f + i_k^f \circ \tilde{c}_k^f, \\
o_k^f &= sigmoid\left(W_f^o x_k + U_f^o h_{k-1}^f + b_f^o\right), \\
h_k^f &= o_k^f \circ \tanh\left(c_k^f\right)
\end{aligned}
\tag{2}
$$

LSTM-Backward:

$$
\begin{aligned}
f_k^b &= sigmoid\left(W_b^f x_k + U_f^f h_{k-1}^f + b_b^f\right), \\
i_k^b &= sigmoid\left(W_b^i x_k + U_f^i h_{k-1}^f + b_b^i\right) \\
\tilde{c}_k^b &= \tanh\left(W_b^g x_k + U_f^g h_{k-1}^f + b_b^g\right), \\
c_k^b &= f_k^b \circ c_{k-1}^b + i_k^b \circ \tilde{c}_k^b, \\
o_k^b &= sigmoid\left(W_b^o x_k + U_f^o h_{k-1}^f + b_b^o\right), \\
h_k^b &= o_k^b \circ \tanh\left(c_k^b\right)
\end{aligned}
\tag{3}
$$

The superscript b indicates backward propagation, and f indicates forward propagation. The equations above describe a BiLSTM, where there are two directions of hidden state propagation, each with its own set of weight matrices and biases. The input sequence is denoted as x_k, and the hidden state h_k incorporates information from both forward and backward propagation. The hidden states for forward and backward propagation are denoted as h_k^f and h_k^b, respectively, and they are computed based on the input from both directions as well as the hidden state from the previous time step.

We aggregated these vectors using a phase-specific BiLSTM to obtain $u_{i,j}$, allowing the LSTM to capture spatial relationships between different layers in an ordered manner. This BiLSTM is also referred to as an internal LSTM. $u_{i,j}$ represents the feature representation of the

CECT images for the j-th phase, encapsulating the semantic information for that phase. See Eq. (4) for details.

$$
\begin{aligned}
h^f_{i,K_{i,j}} &= LSTM^f\left(\left[u_{i,j,1}, u_{i,j,2}, \ldots, u_{i,j,K_{i,j}}\right], h^f_o\right), \\
h^b_{i,K_{i,j}} &= LSTM^b\left(\left[u_{i,j,1}, u_{i,j,2}, \ldots, u_{i,j,K_{i,j}}\right], h^b_o\right)
\end{aligned}
\tag{4}
$$

Therefore, $u_{i,j}$ is obtained by concatenating h^f_k and h^b_k, as shown in Eq. (5),

$$
u_{i,j} = \left[h^f_{i,K_{i,j}}; h^b_{i,K_{i,j}}\right]
\tag{5}
$$

where $K_{i,j}$ represents the number of layers of a CECT image.

Phase index embedding

Next, to establish the correspondence between images and phases, we encoded the index j of each phase as a one-hot vector and then mapped it to a vector $z_{i,j} \in \mathbb{R}^{d_z}$ using an embedding matrix $W^{embed} \in \mathbb{R}^{4 \times d_z}$ (Eq. (6)). We concatenated $z_{i,j}$ and $u_{i,j}$ to obtain $e_{i,j} = [u_{i,j}; z_{i,j}]$. This augmentation with $e_{i,j}$ incorporates information about the phase, in addition to the features in h_{ij}.

$$
z_{i,j} = onehot(j) \cdot W^{embed}
\tag{6}
$$

External LSTM

Subsequently, we input $[e_{i,j} \mid j \in \{1,2,3,4\}]$ into a BiLSTM to aggregate semantic features from different phases, facilitating the learning of associations between images across various phases. We refer to this BiLSTM as the external LSTM. Apart from the input, the forward computation process of the external LSTM is identical to that of the internal LSTMs in each phase, as illustrated in Eq. (7).

$$
\begin{aligned}
h^f_i &= LSTM^f\left(\left[e_{i,j}\right], h^f_o\right), \\
h^b_i &= LSTM^b\left(\left[e_{i,j}\right], h^b_o\right)
\end{aligned}
\tag{7}
$$

Output layer

Finally, we established a prediction layer, which comprises a single-layer fully connected neural network, to generate a vector of length 3 as the output. We applied the softmax activation function to this output, allowing us to make predictions for the diagnostic outcomes (normal, HCC, or ICC) for patient i, as illustrated in Eq. (8).

$$
\hat{y}_i = softmax\left(W^y\left[h^f_i; h^b_i\right] + b^y\right)
\tag{8}
$$

Loss function

We constructed the cross-entropy loss function based on the real diagnosis of patients, denoted as y_i, and the predicted values, denoted as y'_i, as shown in Eq. (9).

$$
\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{3} y_{i,c} \cdot log\left(y'_{i,c}\right) + \lambda \, ||\Theta||_2^2
\tag{9}
$$

In this context, Θ represents all the model parameters to be learned, and λ is the coefficient for L2 regularization.

Experimental configuration

Dataset

The dataset utilized in this research was gathered from the Chongqing Yubei District People's Hospital, Chongqing Wanzhou Three Gorges Central Hospital, and the Radiology Departments of Southwest Hospital. It encompasses a total of 276 participants, segmented into three distinct groups. The normal group consists of 83 individuals who underwent routine physical examination CECT scans, revealing no liver or bile duct abnormalities. The diseased group is composed of 193 individuals, including 94 cases of Hepatocellular Carcinoma and 99 cases of Intrahepatic Cholangiocarcinoma, all confirmed by pathological examination to be primary liver cancer. This dataset offers two subtypes of primary liver cancer cases, providing a robust basis for a comprehensive analysis and evaluation of the proposed methods. For a detailed account of the patient selection process and the inclusion and exclusion criteria, refer to Fig.A1.

Data preprocessing

Liver cancer CECT scans typically comprise four phases: plain, arterial, venous, and delayed. The CECT images for each phase are three-dimensional with dimensions $n_{i,j} \times 512 \times 512$, where $n_{i,j}$ varies. Here, i denotes the patient's index, and j indicates the phase. We rescale each phase to $n_{i,j} \times 224 \times 224$ using cubic interpolation. The purpose of rescaling to 224×224 is to facilitate the use of most pre-trained mature networks for transfer learning, which is particularly crucial for small-sample medical data. Moreover, to adapt the data for direct modeling with 3D network structures like 3D-ResNet, we uniformly rescale the images to $60 \times 224 \times 224$, where 60 represents the number of layers. We then concatenate the CECT images from all four phases along the channel dimension, resulting in a final dimension of $60 \times 224 \times 224 \times 4$.

We constructed scenarios of incomplete phase data based on the complete phase datasets. Specifically, in real clinical

settings, every patient must have a plain phase CECT scan. Thus, when creating the incomplete phase dataset, we assumed that each patient would have at least the plain phase, ensuring that any given patient i would have at least two CECT phases, one of which is the plain phase. Let n_i represent the total number of CECT phases available for patient i. During model training, we randomly removed 0–2 phases, excluding the plain phase. After removal, each patient i would have 2–4 CECT phases, simulating a scenario with missing phases. This augmentation strategy ensures that each training sample includes at least two CECT phases, enabling us to model more complex scenarios with missing phase data and enhance the model's generalization ability.

Model comparison

For the H-LSTM network, we employed various feature extractors to assess their impact on performance. Additionally, we experimented with replacing the BiLSTM structure with a Transformer architecture, which we refer to as the Hierarchical Transformer (H-Transformer). Furthermore, we utilized a 3D-ResNet to directly classify and predict across multiple phases of images for individual patients.

Implementation details

We employed a stratified randomization approach to partition the dataset, ensuring that the training, testing, and validation sets maintained class balance within the hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (ICC), and normal populations. Specifically, the dataset was divided into training, testing, and validation sets in a 7:2:1 ratio, as shown in Table 1. This approach was essential to maintain class balance across the various datasets.

For the feature extractor, we utilized pre-trained models from the ImageNet dataset, obtained through the torchvision package. The rationale for selecting these models (e.g., ResNet, DenseNet) was their proven performance in medical imaging tasks and the availability of well-established, transferable parameters. We then fine-tuned these parameters in an end-to-end manner while training the H-LSTM model. The Adam optimizer was used due to its efficiency in handling sparse gradients and its ability to adjust learning rates dynamically. Detailed hyperparameter settings are presented in Table 2. The training set was employed for model parameter training, while the validation set was used for hyperparameter

Table 1 Composition of patients from different categories in various datasets

	Train	Test	Valid	All
HCC	65	18	11	94
ICC	69	19	11	99
CN	58	16	9	83

Table 2 List of hyperparameters for model construction

Parameter	Value
Image Feature Extractor	
Input Size	224×224
Output Feature Vector	512
Phase Internal LSTM	
Layers	1
Hidden Size	64
Between Phase LSTM	
Layers	1
Hidden Size	64
Model Training	
Initial Learning Rate	0.001
Regularization Coefficient	0.001
Optimizer	Adam
Batch Size	18
Max Epochs	100
Loss Function	Cross Entropy

tuning and implementing an 'early stopping' strategy. Specifically, we stipulated that training would halt if there was no improvement in performance on the validation set for two consecutive epochs. The model that achieved the highest AUC on the validation set was selected as the final model and subsequently evaluated on the test set.

Evaluation metrics

To assess the performance of our classification model, we employed the following metrics: Accuracy, Recall (Sensitivity), Precision, F1 Score, Area Under the Receiver Operating Characteristic Curve (AUROC), and Area Under the Precision-Recall Curve (AUPRC). For multi-class classification problems, we employed the "macro" averaging method to calculate each metric across all classes. This approach treats all classes with equal importance. Bootstrap resampling was utilized to estimate confidence intervals for each metric, providing a measure of statistical reliability. All results were reported based on the test dataset.

Results

Impact of different feature extractors on performance

As demonstrated in Fig. 3; Table 3, we tested the overall average performance of the H-LSTM model with different feature extractors on the test dataset. We found that the performance of various types of ResNet networks and VggNets was relatively similar. However, the DenseNet,

Fig. 3 Performance of the H-LSTM model with different image feature extractors
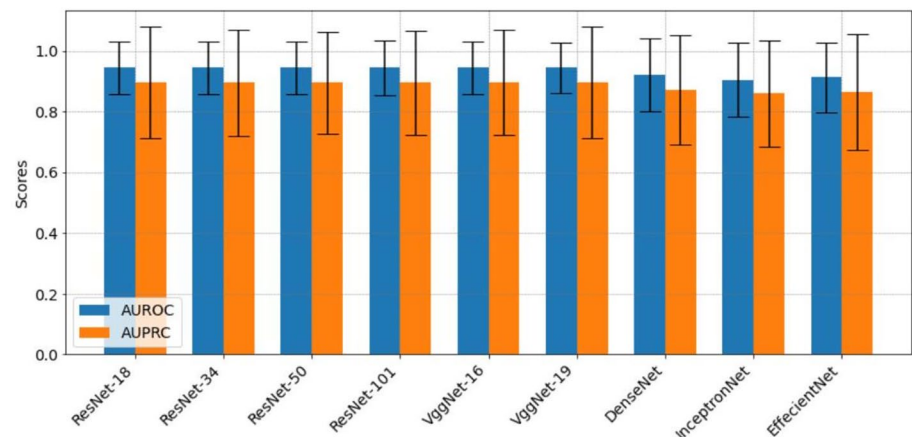
Table 3 Performance of the H-LSTM model with different image feature extractors

Feature Extractor	Accuracy	Recall	Precision	F1	AUROC	AUPRC
ResNet-18	0.91(0.81,0.98)	0.91(0.85,0.98)	0.92(0.85,0.98)	0.91(0.82,0.98)	0.93(0.90,1.00)	0.86(0.86,1.00)
ResNet-34	0.91(0.83,0.96)	0.91(0.85,0.96)	0.92(0.86,0.97)	0.91(0.83,0.96)	0.93(0.90,0.99)	0.86(0.87,0.99)
ResNet-50	0.91(0.83,0.98)	0.91(0.84,0.98)	0.92(0.85,0.98)	0.91(0.82,0.98)	0.93(0.90,1.00)	0.86(0.86,0.99)
ResNet-101	0.91(0.81,0.98)	0.91(0.84,0.97)	0.92(0.85,0.98)	0.91(0.81,0.98)	0.93(0.91,1.00)	0.86(0.87,0.99)
VggNet-16	0.91(0.83,0.98)	0.91(0.85,0.98)	0.92(0.85,0.98)	0.91(0.83,0.98)	0.93(0.90,1.00)	0.86(0.85,1.00)
VggNet-19	0.91(0.83,0.96)	0.91(0.83,0.97)	0.92(0.84,0.98)	0.91(0.81,0.97)	0.93(0.89,0.99)	0.86(0.86,0.99)
DenseNet	0.85(0.74,0.94)	0.85(0.75,0.94)	0.91(0.84,0.97)	0.88(0.79,0.95)	0.90(0.86,0.98)	0.83(0.83,0.98)
InceptronNet	0.83(0.72,0.92)	0.84(0.74,0.92)	0.91(0.84,0.98)	0.86(0.77,0.94)	0.89(0.85,0.97)	0.81(0.83,0.97)
EffecientNet	0.85(0.75,0.94)	0.85(0.77,0.94)	0.92(0.85,0.97)	0.87(0.79,0.95)	0.90(0.86,0.98)	0.82(0.83,0.97)

InceptionNet, and EfficientNet models exhibited somewhat lower performance in comparison. Therefore, considering both performance and the number of parameters, we opted for ResNet-18 as our feature extractor due to its relatively fewer parameters and superior performance.

We utilized ResNet-18 as the feature extractor and then designed a series of ablation experiments to investigate the effects of sharing the feature extractor and the internal LSTM across different phases. Figure 4 presents the results of these ablation studies. Sharing the internal LSTM between different phases resulted in poorer performance, while sharing the feature extractor across different phases led to improved performance.

Performance comparison of different models

In this preliminary study, we developed classification models for liver cancer diagnosis using a relatively small dataset comprising 94 confirmed cases of HCC and 99 cases of ICC. As shown in Fig. 5; Table 4, the proposed H-LSTM network outperformed the H-Transformer and 3D-ResNet models in overall performance, with an average AUROC of 0.93 (0.90, 1.00) and an average AUPRC of 0.86 (0.87, 1.00). For the CN category, the AUROC and AUPRC of the H-LSTM and

H-Transformer models were similar. However, in the HCC and ICC categories, the performance of the H-LSTM model was superior to both the H-Transformer and 3D-ResNet models. The H-LSTM model achieved an average accuracy of 0.91 (0.83, 0.98) in three-class classification, which was significantly higher than that of the H-Transformer and 3D-ResNet models.

The performance of the model in scenarios with incomplete phases

Next, we proceeded to fine-tune the H-LSTM model, which was initially trained on complete CECT image scenarios, using the data with phase augmentation in the context of the incomplete CECT phases we had constructed. Subsequently, we reported the test results for various scenarios involving these incomplete phases on the test dataset. Detailed results are presented in Table 5. Given our assumption that patients possess at least plain phase CECT images, the models in the incomplete phases scenario were fed with a minimum of two phases of CECT images.

From the table, it is evident that in the incomplete phases scenario. For predictions in the "CN" class, the combination of P and C2 phases achieved optimal results
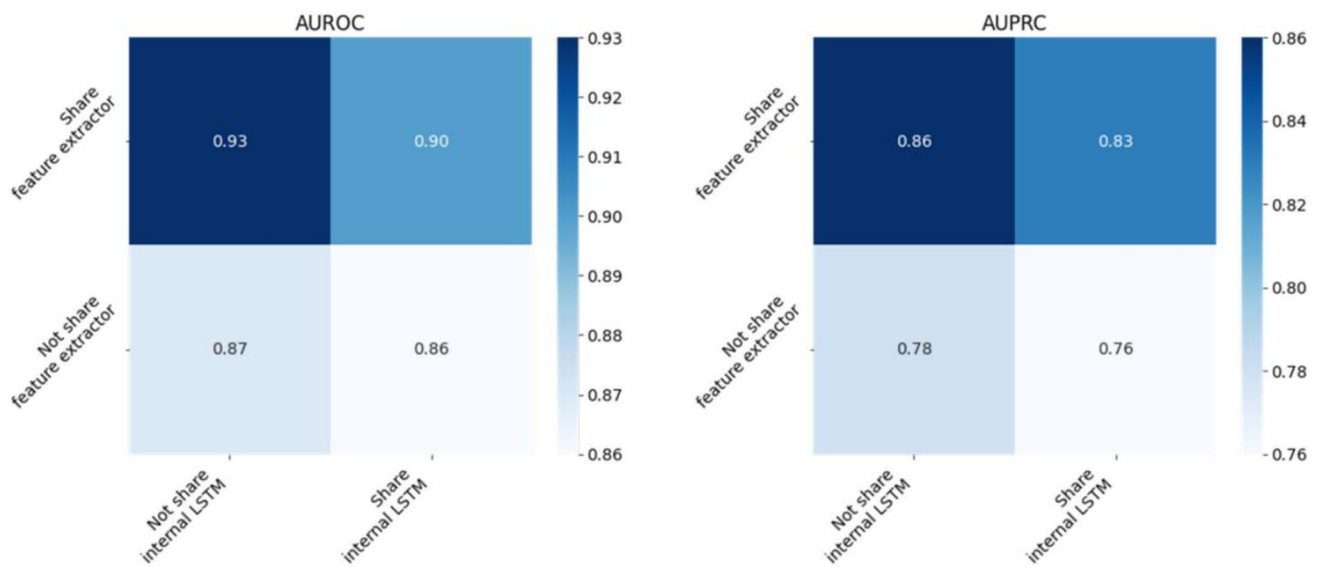
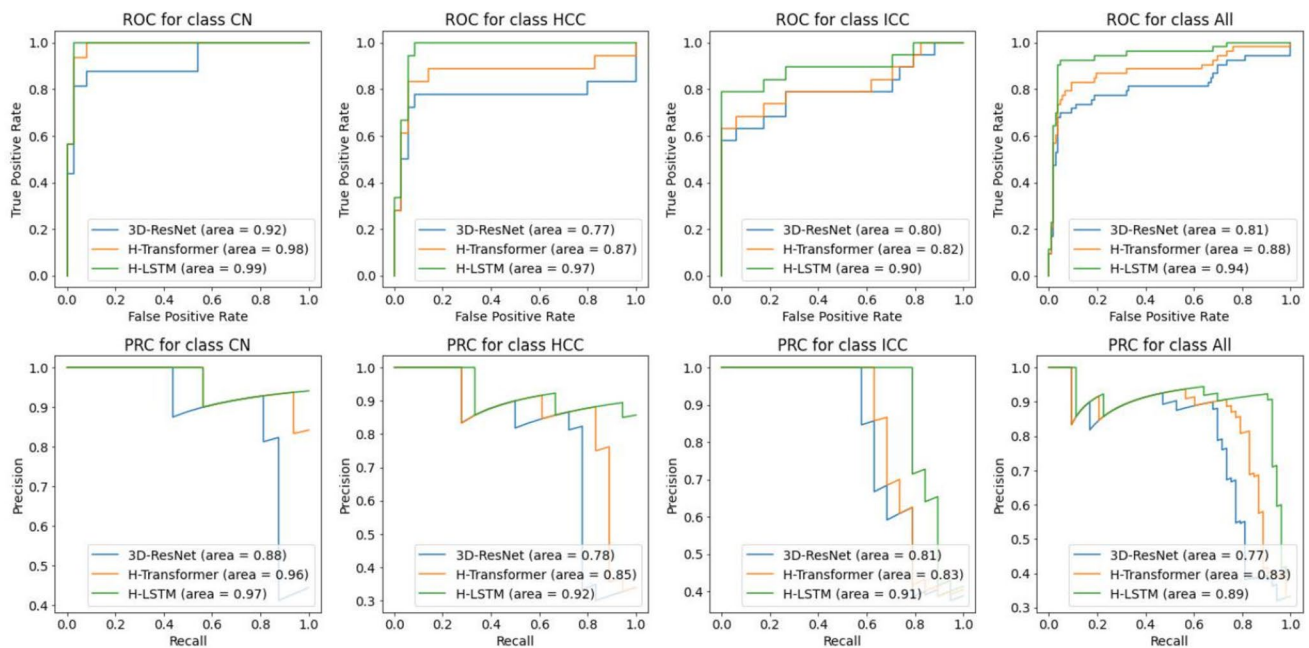**Fig. 4** Results of ablation experiments for different structures of the H-LSTM model

Fig. 5 AUROC and AUPRC of different models for each category and overall

(0.99(0.98,1.00)). For predictions in the "HCC" class, both $P + C1 + C3$ (0.97(0.92,1.00)) and $P + C2 + C3$ (0.97(0.91,1.00)) yielded the best results, suggesting that C3 may be crucial for HCC diagnosis predictions. In the case of predictions for the "ICC" class, the combination of $P + C1 + C3$ achieved the best results (0.90(0.79,0.99)). When evaluating the average predictive performance across all classes, it is evident that combining at least two phases of CECT images results in a significant

performance improvement compared to using single-phase images.

We also evaluated the performance of the H-LSTM model in scenarios where only one phase is present, as depicted in Table B1. Table B2 presents the test results of training the H-LSTM model solely with single-phase CECT images. According to the data, when using only P phase, C1 phase, C2 phase, or C3 phase CECT images, H-LSTM exhibits a reduction in the average AUROC for all categories by 0.031,

Table 4 Performance comparison of different models for each category and overall

Model	Class	Accuracy	Recall	Precision	F1	AUROC	AUPRC
3D-ResNet	CN	0.91(0.81,0.98)	0.75(0.50,0.95)	0.92(0.75,1.00)	0.83(0.61,0.96)	0.92(0.79,0.99)	0.89(0.73,0.99)
	HCC	0.85(0.74,0.92)	0.78(0.59,0.95)	0.78(0.54,0.95)	0.78(0.60,0.89)	0.77(0.58,0.91)	0.78(0.61,0.90)
	ICC	0.83(0.72,0.92)	0.53(0.30,0.77)	1.00(1.00,1.00)	0.69(0.46,0.87)	0.80(0.64,0.95)	0.81(0.66,0.93)
	ALL	0.68(0.55,0.79)	0.68(0.56,0.79)	0.90(0.79,0.97)	0.77(0.64,0.85)	0.82(0.74,0.91)	0.71(0.74,0.92)
H-Transformer	CN	0.92(0.85,0.98)	0.81(0.58,1.00)	0.93(0.78,1.00)	0.87(0.69,0.97)	0.98(0.95,1.00)	0.96(0.88,1.00)
	HCC	0.87(0.77,0.94)	0.83(0.64,1.00)	0.79(0.58,0.94)	0.81(0.64,0.91)	0.87(0.71,0.98)	0.85(0.67,0.96)
	ICC	0.85(0.75,0.94)	0.58(0.35,0.82)	1.00(1.00,1.00)	0.73(0.52,0.90)	0.82(0.67,0.94)	0.83(0.71,0.94)
	ALL	0.74(0.60,0.87)	0.74(0.62,0.86)	0.91(0.83,0.97)	0.80(0.69,0.90)	0.85(0.82,0.96)	0.75(0.81,0.96)
H-LSTM	CN	0.98(0.94,1.00)	1.00(1.00,1.00)	0.94(0.81,1.00)	0.97(0.90,1.00)	0.99(0.96,1.00)	0.97(0.89,1.00)
	HCC	0.92(0.85,0.98)	1.00(1.00,1.00)	0.82(0.68,0.96)	0.90(0.81,0.98)	0.97(0.93,1.00)	0.93(0.83,1.00)
	ICC	0.91(0.81,0.98)	0.74(0.53,0.94)	1.00(1.00,1.00)	0.85(0.69,0.97)	0.90(0.78,1.00)	0.91(0.79,1.00)
	ALL	0.91(0.83,0.98)	0.91(0.84,0.98)	0.92(0.86,0.98)	0.91(0.83,0.98)	0.93(0.90,1.00)	0.86(0.87,1.00)

Table 5 The performance of the H-LSTM model in scenarios with incomplete phases

Phases	Class	Accuracy	Recall	Precision	F1	AUROC	AUPRC
P+C1	CN	0.94(0.87,1.00)	0.88(0.71,1.00)	0.93(0.77,1.00)	0.90(0.77,1.00)	0.99(0.95,1.00)	0.96(0.86,1.00)
	HCC	0.91(0.81,0.96)	1.00(1.00,1.00)	0.78(0.58,0.92)	0.88(0.74,0.96)	0.96(0.89,1.00)	0.90(0.72,0.99)
	ICC	0.91(0.81,0.98)	0.74(0.50,0.93)	1.00(1.00,1.00)	0.85(0.67,0.96)	0.86(0.72,0.97)	0.87(0.69,0.97)
	ALL	0.87(0.75,0.94)	0.87(0.77,0.94)	0.91(0.83,0.97)	0.88(0.77,0.95)	0.91(0.87,0.98)	0.82(0.83,0.98)
P+C2	CN	0.94(0.87,1.00)	0.88(0.69,1.00)	0.93(0.76,1.00)	0.90(0.77,1.00)	0.99(0.98,1.00)	0.99(0.95,1.00)
	HCC	0.91(0.81,0.96)	1.00(1.00,1.00)	0.78(0.60,0.92)	0.88(0.75,0.96)	0.96(0.91,1.00)	0.91(0.77,1.00)
	ICC	0.91(0.83,0.98)	0.74(0.55,0.93)	1.00(1.00,1.00)	0.85(0.71,0.97)	0.87(0.75,0.98)	0.88(0.77,0.98)
	ALL	0.87(0.75,0.94)	0.87(0.77,0.95)	0.91(0.83,0.97)	0.88(0.77,0.95)	0.91(0.88,0.99)	0.82(0.86,0.99)
P+C3	CN	0.94(0.89,1.00)	0.88(0.72,1.00)	0.93(0.78,1.00)	0.90(0.78,1.00)	0.99(0.97,1.00)	0.98(0.93,1.00)
	HCC	0.91(0.81,0.96)	1.00(1.00,1.00)	0.78(0.58,0.92)	0.88(0.74,0.96)	0.97(0.91,1.00)	0.92(0.79,1.00)
	ICC	0.91(0.83,0.98)	0.74(0.55,0.93)	1.00(1.00,1.00)	0.85(0.71,0.96)	0.89(0.76,0.99)	0.90(0.77,0.99)
	ALL	0.87(0.79,0.94)	0.87(0.78,0.95)	0.91(0.85,0.97)	0.88(0.79,0.95)	0.91(0.90,1.00)	0.82(0.86,0.99)
P+C1+C2	CN	0.96(0.91,1.00)	0.94(0.79,1.00)	0.94(0.80,1.00)	0.94(0.84,1.00)	0.99(0.96,1.00)	0.97(0.90,1.00)
	HCC	0.91(0.83,0.98)	1.00(1.00,1.00)	0.78(0.62,0.95)	0.88(0.77,0.98)	0.96(0.90,1.00)	0.91(0.79,1.00)
	ICC	0.91(0.81,0.98)	0.74(0.52,0.91)	1.00(1.00,1.00)	0.85(0.69,0.95)	0.88(0.74,0.99)	0.89(0.78,0.98)
	ALL	0.89(0.79,0.96)	0.89(0.80,0.97)	0.91(0.84,0.97)	0.89(0.79,0.96)	0.92(0.89,0.99)	0.84(0.85,0.99)
P+C1+C3	CN	0.96(0.91,1.00)	0.94(0.82,1.00)	0.94(0.80,1.00)	0.94(0.84,1.00)	0.99(0.95,1.00)	0.97(0.88,1.00)
	HCC	0.91(0.83,0.96)	1.00(1.00,1.00)	0.78(0.61,0.92)	0.88(0.76,0.96)	0.97(0.92,1.00)	0.93(0.83,1.00)
	ICC	0.91(0.81,0.96)	0.74(0.52,0.92)	1.00(1.00,1.00)	0.85(0.69,0.96)	0.90(0.78,0.99)	0.91(0.78,0.99)
	ALL	0.89(0.81,0.96)	0.89(0.82,0.97)	0.91(0.84,0.98)	0.89(0.80,0.97)	0.92(0.89,0.99)	0.84(0.85,0.99)
P+C2+C3	CN	0.96(0.91,1.00)	0.94(0.80,1.00)	0.94(0.77,1.00)	0.94(0.83,1.00)	0.99(0.97,1.00)	0.98(0.93,1.00)
	HCC	0.91(0.83,0.96)	1.00(1.00,1.00)	0.78(0.60,0.93)	0.88(0.75,0.96)	0.97(0.91,1.00)	0.93(0.81,1.00)
	ICC	0.91(0.83,0.98)	0.74(0.53,0.95)	1.00(1.00,1.00)	0.85(0.69,0.97)	0.90(0.79,0.99)	0.91(0.80,0.99)
	ALL	0.89(0.79,0.96)	0.89(0.81,0.96)	0.91(0.83,0.96)	0.89(0.79,0.95)	0.92(0.90,0.99)	0.84(0.87,0.98)

0.093, 0.122, and 0.048, respectively, compared to dedicated single-phase baseline models. This suggests that H-LSTM is less effective in single-phase scenarios than in models specialized for a single phase.

We conducted an external test on the TICA (Translational Imaging in Cancer Alliance) dataset. Within TICA, 89 cases of HCC met the experimental requirements. However, their phase images were incomplete, with each case missing at least one phase of imaging. As a result, we only calculated the accuracy of the H-LSTM model for these 89 HCC patients, which was found to be 0.76 (0.67, 0.84). Fig. C1 displays misclassified cases along with the potential reasons for the misclassification.

Discussion

In real-world scenarios, it can be challenging to effectively integrate variable-phase CECT images when some patients may be missing certain phases. Furthermore, the varying slice counts in CECT images across different phases add complexity to the phase integration task. This research focuses on the classification of patients into three categories: CN, HCC, and ICC, based on their liver CECT scans. To achieve this, we developed a liver diagnostic classification model using ResNet and BiLSTM. This model accommodates variable numbers of CECT images from different phases, and the number of CECT slices in each phase can also vary. Importantly, this process eliminates the need for radiologists to manually select highlighted slices or perform rough annotations of target regions.

Our experimental results demonstrate that our model performs exceptionally well in scenarios involving the integration of variable multi-phase CECT images. The average AUROC exceeds 0.9 for scenarios with 2, 3, and 4 phases, particularly reaching an impressive average AUROC of 0.93(0.90,1.00) and an average AUPRC of 0.86(0.87,1.00) when all 4 phases of CECT images are available. Compared to previous studies that relied primarily on single-phase images (e.g., arterial or venous phase) for HCC and ICC diagnosis (dis Zhao 2022; Xia 2022), our model integrates multi-phase data and demonstrates superior performance across varying combinations of phases. For example, studies such as Gao et al. (2021) and Wang et al. (2023) utilized LSTM and Transformer models but did not fully capture the sequential information from multiple CECT phases. Our H-LSTM model, by contrast, effectively handles incomplete phase data and integrates both intra-phase and inter-phase features, leading to improved classification performance. However, it is noteworthy that in scenarios with only single-phase CECT images, the comprehensive model performs below the single-phase baseline models, indicating that using specialized single-phase models might be a better choice when dealing with single-phase data. Nevertheless, when the number of phases is two or more, we observe a significant improvement in model performance. This suggests that having more phases may provide valuable information, resulting in enhanced classification performance.

Currently, there are significant shortcomings in the research related to HCC and ICC. Firstly, researchers often fail to fully harness the potential value of 3D contrast-enhanced CT images of these two types of liver tumors, limiting their analysis to a single phase and only a portion of 2D contiguous images. For instance, in the case of HCC, studies predominantly rely on arterial-phase contrast-enhanced CT images, while ICC research primarily depends on venous-phase or delayed-phase images (dis Zhao 2022; Xia 2022). This approach fails to comprehensively consider the information available from all four phases of imaging, thereby restricting a comprehensive understanding of the tumors.

Secondly, the data commonly used in existing studies consists of a few consecutive 2D CT images, rather than 3D images. This means that the correlated information between upper and lower layers and spatial information within the images is not fully utilized. This limitation hampers precise tumor localization and in-depth investigations into morphological features.

Furthermore, most researchers concentrate on the classification of a single disease type, such as HCC or ICC, without integrating data from both types of typical primary liver cancers for related classification studies. This segregated approach fails to harness the similarities and differences between these different diseases, limiting progress in the study of overall liver tumors.

Taking into account data completeness from three different aspects, our research offers more comprehensive data and results that are more reliable and accurate. It's worth noting that there is currently a lack of publicly available datasets worldwide that provide 3D images of all four phases of enhanced CT scans for HCC (Jia and Sun 2021)and ICC (Tan 2021). As a result, our research serves as a valuable supplement to the existing data landscape in the field of liver cancer studies. In contrast, widely used liver-related research datasets like LiTS (Liver Tumor Segmentation Challenge) (Bilic et al. 2023) and CHAOS (Combined CT-MR Healthy Abdominal Organ Segmentation) fall short in terms of data volume and data completeness compared to our study.

Our study has some limitations. First, although our sample is derived from multiple centers, the relatively small number of liver cancer patients may limit the generalizability of our model. The H-LSTM model demonstrated a comparatively lower accuracy of 0.76 in HCC patients in the external TCIA dataset due to misclassification resulting from overlapping enhancement patterns, tumor-specific factors, and the TCIA dataset's imaging limitations. To address this shortcoming, we plan to include a larger cohort of patients in the future to enhance the representativeness of our sample. Additionally, we will test the performance of the H-LSTM on more diseases, especially those with incomplete phase scenarios.

Conclusion

In summary, we developed a deep learning model based on ResNet and BiLSTM that effectively addresses the challenge of classifying liver cancer using variable-phase CECT images. The model demonstrated strong performance in

distinguishing normal cases, HCC, and ICC, even in real-world scenarios with incomplete data. This highlights the potential of AI-assisted systems in enhancing the accuracy and efficiency of liver cancer diagnosis and treatment.

Data availability The datasets used and/or analyzed in this study can be made available to the corresponding author upon reasonable request.

Code availability The code is publicly available at https://github.com/ljwa2323/ResNet_BiLSTM.

Declarations

Conflict of interest Shixin Huang, Xixi Nie, Kexue Pu, Xiaoyu Wan, and Jiawei Luo have no conficts of interest to declare that may be relevant to the contents of this article.

Ethical approval Ethics approval was obtained from the Institutional Review Committees at the following institutions: People's Hospital of Yubei District, Chongqing City; Chongqing University Three Gorges Hospital; and The Southwest Hospital of Army Medical University.

References

Aslan MF, Unlersen MF, Sabanci K et al (2021) CNN-based transfer learning–BiLSTM network: a novel approach for COVID-19 infection detection. Appl Soft Comput 98:106912

Ayuso C, Rimola J, Vilana R et al (2018) Diagnosis and staging of hepatocellular carcinoma (HCC): current guidelines. Eur J Radiol 101:72–81

Bakrania A, Joshi N, Zhao X et al (2023) Artificial intelligence in liver cancers: decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases. Pharmacol Res 189:106706

Bilic P, Christ P, Li HB et al (2023) The liver tumor segmentation benchmark (lits). Med Image Anal 84:102680

Chen X Lin L, Hu H (2019) A cascade attention network for liver lesion classification in weakly-labeled multi-phase ct images; proceedings of the Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data: First MICCAI Workshop, DART 2019, and First International Workshop, MIL3ID 2019, Shenzhen, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13 and 17, 2019, Proceedings 1, F, [C]. Springer

dis Zhao Y-Q (2022) TD-Net: a hybrid end-to-End Network for Automatic Liver Tumor Segmentation from CT images. IEEE J Biomedical Health Inf 27(3):1163–1172

Doğantekin A, Özyurt F, Avci E et al (2019) A novel approach for liver image classification: PH-C-ELM. Measurement 137:332–338

Dutta P, Upadhyay P et al (2020) [C] DE M,. Medical image analysis using deep convolutional neural networks: CNN architectures and transfer learning; proceedings of the 2020 International Conference on Inventive Computation Technologies (ICICT), F, IEEE

Gao F, Yoon H, Wu T et al (2020) A feature transfer enabled multi-task deep learning model on medical imaging. Expert Syst Appl 143:112957

Gao R, Zhao S, Aishanjiang K et al (2021) Deep learning for differential diagnosis of malignant hepatic tumors based on multiphase contrast-enhanced CT and clinical data. J Hematol Oncol 14(1):1–7

Hamm CA, Wang CJ, Savicl J et al (2019) Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI. Eur Radiol 29:3338–3347

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Huang G, Liu Z, Van Der Maaten L et al (2017) Densely connected convolutional networks; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, [C]

Jakimovski G, Davcev D (2019) Using double convolution neural network for lung cancer stage detection. Appl Sci 9(3):427

Jia X, Sun Z (2021) A multimodality-contribution-aware tripnet for histologic grading of hepatocellular carcinoma. IEEE/ACM Trans Comput Biol Bioinf 19(4):2003–2016

Kora P, Ooi CP, Faust O et al (2022) Transfer learning techniques for medical image analysis: a review. Biocyber Biomed Eng 42(1):79–107

Kim HE, Cosa-Linan A, Santhanam N et al (2022) Transfer learning for medical image classification: a literature review. BMC Med Imaging 22(1):69

Lakshmipriya B, Pottakkat B (2023) Deep learning techniques in liver tumour diagnosis using CT and MR imaging-A systematic review. Artif Intell Med 141:102557

Lipton ZC, Berkowitz J, Elkan C (2015) A critical review of recurrent neural networks for sequence learning. arXiv preprint, CoRR, arXiv:150600019

Ling Y, Ying S, Xu L et al (2022) Automatic volumetric diagnosis of hepatocellular carcinoma based on four-phase CT scans with minimum extra information. Front Oncol 12:960178

Mcglynn KA, Petrick JL, El-Serag H (2021) B. Epidemiology of hepatocellular carcinoma. Hepatology 73:4–13

Petrowsky H, Fritsch R, Guckenberger M et al (2020) Modern therapeutic approaches for the treatment of malignant liver tumours. Nat Rev Gastroenterol Hepatol 17(12):755–772

Ponnoprat D, Inkeaw P, Chaijaruwanich J et al (2020) Classification of hepatocellular carcinoma and intrahepatic cholangiocarcinoma based on multi-phase CT scans. Med Biol Eng Comput 58:2497–2515

Romero F, Dilera P, Bisson-Gregoire GF et al (2019) End-to-end discriminative deep network for liver lesion classification. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). IEEE, 1243–1246

Rumgay H, Arnold M, Ferlay J et al (2022) Global burden of primary liver cancer in 2020 and predictions to 2040. J Hepatol 77(6):1598–1606

Sarvamangala D, Kulkarni RV (2022) Convolutional neural networks in medical image understanding: a survey. Evol Intel 15(1):1–22

Sarwinda D, Paradisa RH, Bustamam A et al (2021) Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. Procedia Comput Sci 179:423–431

Shanmugapriya S, Pravda J, Rabia N et al (2022) Deep learning for image-based liver analysis—a comprehensive review focusing on malignant lesions. Artif Intell Med 6(120):101016

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556

Siegel RL, Miller KD, Goding Sauer A et al (2020) Colorectal cancer statistics, 2020. Cancer J Clin 70(3):145–164

Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin 71(3):209–249

Szegedy C, Liu W, JiA Y et al (2015) Going deeper with convolutions; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, [C]

Tan M, Efficientnet LEQ (2019) [C] Rethinking model scaling for convolutional neural networks; proceedings of the International conference on machine learning, F, PMLR

Tan JW, Lee K, Lee K et al (2021) [C] Improving the Accuracy of Intrahepatic Cholangiocarcinoma Subtype Classification by Hidden Class Detection via Label Smoothing; proceedings of the 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), F, IEEE

Wang X, Ying H, Xu X et al (2023) [C] TransLiver: A Hybrid Transformer Model for Multi-phase Liver Lesion Classification; proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, F, Springer

Xia W, Liu M, Yang C (2022) Deep Learning Method Based on CT Images to Predict the Pathological Differentiation of Intrahepatic Cholangiocarcinoma; proceedings of the 2022 IEEE 8th International Conference on Computer and, Communications F et al [C] IEEE

Yasaka K, Akai H, Abe O et al (2018) Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. Radiology 286(3):887–896

Yoo Y, Baek J-G (2018) A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. Appl Sci 8(7):1102

ZHANG H, LUO K, DENG R et al (2022) Deep Learning-Based CT Imaging for the Diagnosis of Liver Tumor. Comput Intell Neurosci 1:1

Zhou J, Wang W (2021) Automatic detection and classification of focal liver lesions based on deep convolutional neural networks: a preliminary study. Front Oncol 10:581210