Cross-modality Distillation Network for Microvascular Invasion Prediction of Hepatocellular Carcinoma from MRI Images

Zheng Wang[1], **Hong Liu**[2], **Ziyi Wang**[3], **Dongyu Pan**[4], **Xinyu Chen**[5],

[1] 23020231154232, [2] 24520230157442, [3] 23020230157351, [4] 31520231154308, [5] 30920230157369,
Hong Liu and Dongyu Pan from Information Institute class, rest of all from AI class

Abstract

Microvascular invasion (MVI) of hepatocellular carcinoma (HCC) is a crucial histopathologic prognostic factor associated with cancer recurrence after liver transplantation or hepatectomy. Recently, clinicoradiologic characteristics are combined with medical images to enhance the HCC prediction. However, compared to medical imaging data, the clinicoradiologic characteristics is not easy to collect or even unavailable, as it requires more efforts of clinicians and more medical instruments for collecting diverse measurements. This work explores how to transfer the knowledge of a teacher network learned from non-image clinical data and image data to a student network with only image data such that the student network can leverage the transferred clinical information to boost HCC classification with only imaging data as input. Specifically, we present a cross-modality distillation network (CMD-Net) to transform knowledge of non-image clinicoradiologic from the teacher network to the student network. The teacher network integrates non-image clinicoradiologic characteristics with two 3D MRI modality images via MRI-clinical-fusion modules and a cross attention (CA) module, while the student network extracts features from two modality MRI data via two MRI-only modules and then refine these two MRI features via a CA module. Image-level distillation and feature-level distillation are jointly adopt to transfer the clinical information between teacher and student networks.

Introduction

Hepatocellular carcinoma (HCC) is the fifth most common cancer in the world and the third leading cause of cancer-related death. The 5-year overall survival rate of HCC patients after surgery is only 10-20% (Yang et al. 2019). After hepatectomy and liver transplantation, The 5-year recurrence rate can be as high as 50-70% and 35%, respectively(Mazzaferro et al. 2018).

Many literature reports that vascular invasion is one of the important factors that threaten the prognosis of patients (Roayaie et al. 2009; Lee et al. 2017), which limits the implementation of curable treatment strategies for liver resection, liver transplantation, and radiofrequency ablation (Imamura et al. 2003; Okada et al. 1994). According to its detection methods, vascular invasion can usually be classified into Macrovascular invasion (MaVI) and Microvascular invasion (MVI) (Sumie et al. 2014; Roayaie et al. 2009; Lee et al. 2017). MVI refers to the presence of nests of cancer cells in the vascular cavity lined by endothelial cells under the microscope (Yang et al. 2019; Cong et al. 2016; Lei et al. 2016; Rodriguez-Peralvarez et al. 2013), which is present in 15-57.1% of postoperative liver cancer specimens (Lei et al. 2016). MVI is also a risk factor for poor outcomes after liver resection or liver transplantation in patients with liver cancer (Xu et al. 2019; Lee et al. 2017; Shindoh et al. 2020; Iguchi et al. 2015). However, MVI is only visible under the postoperative pathology microscope (Cong et al. 2016; Lei et al. 2016) and requires extensive sampling (Hu et al. 2018). Its relatively lagging gold standard for pathology severely limits the timely and effective adjustment of surgical treatment strategies. Therefore, the accurate stratification of MVI grades before surgery can be used as an important evaluation reference index for the formulation of treatment plans for patients with liver cancer and the follow-up monitoring after surgery. According to the number and distribution of microvessels involved, MVI can be further divided into M0 (no MVI), M1 (MVI $<=$5 and within 1cm of the tumor edge) and M2 (MVI $>$ 5 or $>$ 1cm from the tumor surface) (Cong et al. 2016).

As verified by previous studies (Zhang et al. 2021), clinical radiological characteristics of patients is more relevant to the prediction of MVI. Experiments have also proved that the introduction of clinical radiological characteristics have greatly improved the prediction effect of MVI. However, clinical radiology characteristics are not always available in practice, and there are many experimental indicators that are not fully obtained. So we thought of transferring the knowledge of the teacher network that introduced clinical radiology characteristics to the student network that only had image data.

In this paper, we present a cross-modality distillation network (CMD-Net) for predicting MVI of HCC to distill a teacher network with a combination of imaging and clinical data to a student network with only imaging data. By doing so, the inference stage does not require any clinical data and the network performance with only imaging data in the inference stage can be further improved due to the clinical information transferred from the teacher network in our method. Here, multiple MRI data and 52 clinical items are

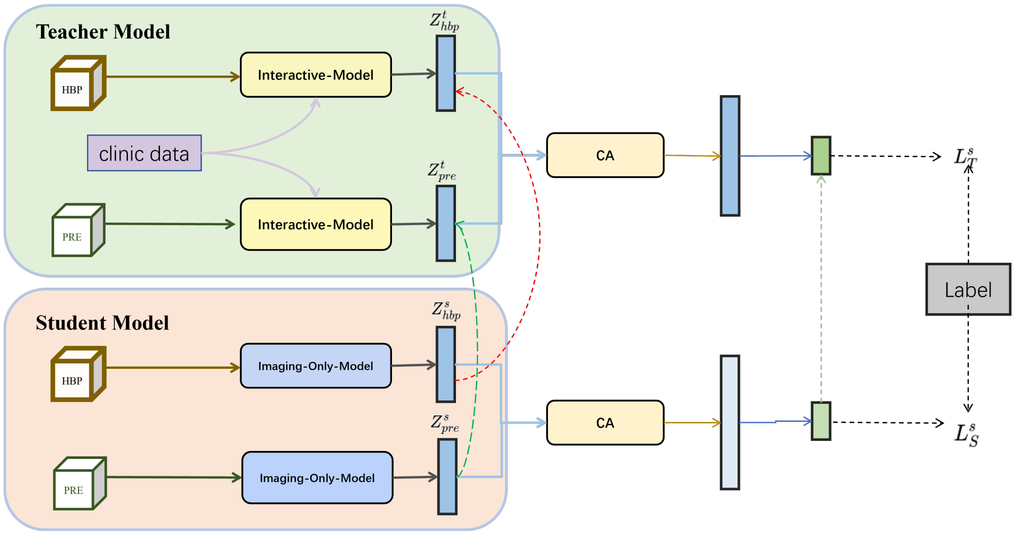

Figure 1: Illustration of our modality-aware distillation network, which transfers knowledge from a teacher network with clinical data to a student network without clinical data. The student network is trained with a supervised learning loss, and a Kullback Leibler (KL) Divergence based mimcry loss to match the probability estimates of the teacher network.

utilized in our work.

Related work

Microvascular invasion of hepatocellular carcinoma

Early MVI prediction works mainly examined radiologic features at the local lesion area of an MR volume (Chong et al. 2021; Xu et al. 2019; Yang et al. 2019; Feng et al. 2019), and these features included non-smooth tumor margin, peritumoral enhancement on arterial phase (AP), peritumoral hypointensity on hepatobiliary phase (HBP), and so on (Feng et al. 2019). However, these hand-crafted features are sensitive to the acquisition methods and reconstruction parameters, thereby suffering from limited capability in handle diverse clinical usage. Motivated by the superior performance of deep features over hand-crafted features in diverse medical image analysis tasks, convolutional neural network (CNNs) have been developed to classify MVI of HCC patients. Jiang et al. (Jiang et al. 2021) utilized eXtreme Gradient Boosting (XGBoost) and deep learning from CT images to predict MVI preoperatively. Zhang et al. (Zhang et al. 2021) developed a 3D CNN prediction model to fuse features from multiple MR sequences. Men et al. (Men et al. 2019) embedded long short-term memory (LSTM) into a CNN to fuse multi-modal MR volumes for predicting MVI of HCC patients. Xiao et al. (Xiao, Zhao, and Li 2022) proposed a task relevance driven adversarial learning framework (TrdAL) for simultaneous HCC detection, size grading, and multi-index quantification using multi-modality MRI. However, only MR images are involved to predict MVI status in those CNN-based methods. To boost the MVI prediction accuracy, our work leverages both imaging modality and clinical modality within a knowledge distillation learning framework. Unlike conventional knowledge distillation, the teacher network in our method takes multiple imaging data and non-image clinical data, while the student network only utilizes imaging data. Hence, our modality-aware distillation network transfers the clinical information from the teacher network and the student network to enhance the classification accuracy of the student network with only imaging data.

Proposed Solution

Figure 1 shows the schematic illustration of the proposed Cross-Modality Distillation Network (CMD-Net) for MVI prediction of HCC. As a distillation network, our CMD-Net consists of a teacher network and a student network, but it distills a teacher network with diverse clinical information to a student network without any clinical data. The student network takes two MRI sequences as the input, pass each MRI image into a MRI-only module to extract MRI features, and then develops a cross attention (CA) module to refine two MRI features for final MVI prediction. On the other hand, the teacher network presents two MRI-clinical-fusion module to first extract integrated features of the MRI image and the clinical data, and then refine these two obtained features via another CA module for generating a MVI classification result. After that, we devise a distillation scheme by considering both class-level distillation and feature-level distillation. The class-level distillation makes the two predictions of the student network and the teacher network to be similar, while the feature-level distillation transfers the clinical-guided features of the teacher network to student's features, which do not consider any clinical data.

Teacher network

Non-image Clinical data has shown its capability of providing complementary information for the classification task with only image data (Duanmu et al. 2020). Motivated by this, we integrate clinical data, the Hepatobiliary phase (HBP) MRI image, and the pre-contrast (PRE) MRI image as the input of the teacher network of our CMD-Net for the prediction of MVI in HCC and training the student network with only image data. Specifically, the teacher network first passes the HBP image and clinical data into a MRI-clinical-fusion CNN to extract a 512-dimensional vector Z^t_{hbp}, and the PRE image and clinical data are then feed into another MRI-clinical-fusion CNN for obtaining another 512-dimensional vector Z^t_{pre}. Then, we concatenate Z^t_{hbp} and Z^t_{pre} to produce Z^t, which is passed into two fully-connected layers to predict a classification result P^t with three elements of the teacher network.

MRI-clinical-fusion module. Similar to (Duanmu et al. 2020), our MRI-clinical-fusion module integrates MRI data and non-imaging clinical data for a HCC prediction. taking a 3D MRI data and a vectorized clinical data as the inputs, the image-clinical fusion module first applies four fully-connected (FC) layers on the input clinical data to obtain four feature maps, which feature channels are 64, 128, 256, and 256. Meanwhile, we utilize four convolutional blocks on the input MRI image to obtain another 3D feature maps, and the feature channels are also set as 64, 128, 256, and 256. Each convolutional block consists of two 3×3 convolutional layers. And then we channel-wisely multiply four feature maps from the clinical data and the corresponding four features from the MRI data for integrating them together. Specifically, let C denote the clinical features (a vector), X denote the MRI image features (3D), and Y denote the output 3D feature map. Note that the number of element of the vector C is the same as the number of channels of 3D feature map X. Then, the channel-wise calculation (between C and X) is summarized as follows: (1) for i-th element of C, we multiply it with the i-th channel of X and take the multiplication results as the i-th channel of Y:

$$Y_i(u, v) = C_i * X_i(u, v) \ , \tag{1}$$

where C_i denote the i-th element of C. X_i and Y_i represent the i-th channel of the 3D feature map X and Y, respectively. (u, v) denotes the pixel coordinates at the X_i and Y_i. (2) Then, we conduct such operation for all elements of the clinical feature vector C and thus can generate the multiplication result Y. After that, we then apply a 3×3 convolutional layer and one fully-connected (FC) layers to output a feature vector with 512 elements.

Student network

Although a fusion of the clinical data and the image data can improve the HCC classification result, the clinical data are not often available when compared to the MRI images for classifying HCC patients in clinical diagnosis. In order to improve the flexibility of our model in clinical application, we devise a modality-aware knowledge distillation network to transfer the knowledge learned by a teacher network with a fusion of a clinical data modality and the image modality to a student network with only the image modality. By doing so, the clinical data knowledge can be distilled from the teacher network to the student network, and thus the classification performance of the student work can be enhanced even though the student network does not get any clinical data in the testing stage.

As shown in Figure 1, our student work takes a 3D HBP MRI image and a 3D PRE MRI image as the input, and then passes the HBP data into a MRI-only module to obtain features Z^s_{hbp} and the PRE data into another MRI-only module to obtain the feature map Z^s_{pre}. Note that Z^s_{hbp} and Z^s_{pre} are two vectors with 512 elements. After that, we concatenate Z^s_{hbp} and Z^s_{pre} to obtain Z^s and feed Z^s into a fully-connected layer to predict a HCC classification result P^s, which has three elements.

Cross Attention (CA) Module

Our CA module is to refine two features from different image modalities by leveraging their complementary information based on self-attention frameworks (Oh et al. 2019; Vaswani et al. 2017; Wang et al. 2018). Specifically, let X and Y to denote the input two feature maps of the SA module. Then, the SA module first applies a linear transformation layers on X to obtain three feature maps, including query Q_x, key K_x, and value V_x. Meanwhile, we apply a linear transformation layers on Y to generate a key feature map K_y and a value feature map V_y. After that, we generate a score map S_x by multiplying Q_x and the transpose of K_x, and another score map S_y by multiplying Q_y and the transpose of K_y. Then, we multiple the obtained score maps S_x with the value feature map V_x, and multiply S_y with V_y to produce two resultant feature maps, which are then added together to generate the output refined feature map $\hat{X}$:

$$\hat{X} = V_x \times (Q_x \times {K_x}^T) + V_y \times (Q_x \times {K_y}^T) \ . \tag{2}$$

Similarly, the CA module applies another transformation layer on Y to obtain a feature map Q_y. Then, two score maps are computed by multiplying Q_y and the transpose of K_y, and multiplying Q_y and the transpose of V_y. After that, the refined feature map $\hat{y}$ is computed by:

$$\hat{Y} = V_x \times (Q_y \times {K_x}^T) + V_y \times (Q_y \times {K_y}^T) \ . \tag{3}$$

Cross-Modality Distillation

We apply the knowledge distillation strategy to transform the clinical information of the teacher network to the student network. Apart from the straightforward classification result-level distillation, we present an auxiliary feature-level distillation loss to distill features fused from clinical data and MRI image of the teacher network to features from only MRI image.

Classification-level distillation. Let $q^s_m(x_i)$ denote the class probabilities for the class of the MRI x_i data produced from the student network, while $q^t_m(x_i)$ represent the class probabilities for the class of the MRI x_i data produced from the teacher network network. Then, the classification-level distillation loss L^d_{class} is simply defined to push make the class probabilities from the teacher network as targets for training the student network. To do so, we utilize the Kullback Leibler (KL) divergence to measure the difference of two distribution:

$$\begin{aligned} L^{d}_{class} &= DKL\left(q^{t}_{m}(x_i)\|q^{s}_{m}(x_i)\right) \\ &= \sum_{i=1}^{N}\sum_{m=1}^{M} p_2^{m}\left(x_i\right)\log\frac{p_2^{m}\left(x_i\right)}{p_1^{m}\left(x_i\right)} , \end{aligned} \tag{4}$$

where N and M denote the number of training sample and the number of total class. $DKL()$ represents the Kullback-Leibler divergence between two probabilities.

Feature-level distillation. Apart from the classification-level knowledge distillation, we also transfer the intermediate features of the teacher network with the clinical information to that of the student network. In this regard, we devise a feature-level distillation strategy. Specifically, we distill the output features of two Interactive Models of the teacher network, since these two features integrate the clinical data and the HBP image and the clinical data and the PRE image respectively. Hence, we compute a feature-level distillation loss $L^{d}_{feature}$ as the combination of the Kullback Leibler (KL) divergence between Z^{t}_{hbp} and Z^{s}_{hbp} and the Kullback Leibler (KL) divergence between Z^{t}_{pre} and Z^{s}_{pre}:

$$\begin{aligned} L^{d}_{feature} &= DKL\left(Z^{t}_{hbp}\|Z^{s}_{hbp}\right) + \beta_1 DKL\left(Z^{t}_{pre}\|Z^{s}_{pre}\right) \\ &= \sum_{i=1}^{N}\sum_{m=1}^{M} p_2^{m}\left(x_i\right)\log\frac{p_2^{m}\left(x_i\right)}{p_1^{m}\left(x_i\right)} , \end{aligned} \tag{5}$$

where β_1 is to weight Kullback-Leibler divergence terms, and the weight β_1=1. $DKL(Z^{t}_{hbp}\|Z^{s}_{hbp})$ denote the Kullback-Leibler divergence between two features Z^{t}_{hbp} and Z^{s}_{hbp}. $DKL(Z^{t}_{pre}\|Z^{s}_{pre})$ represents the Kullback-Leibler divergence between two features Z^{t}_{pre} and Z^{s}_{pre}.

Our loss function. The loss function of our network consists of two supervised losses on the teacher network and the student network, the self-supervised loss for the clinical data prediction, and distillation loss between the student network and the teacher network. The definition of our loss function is given by:

$$L_{total} = L^{s}_{T} + L^{s}_{S} + L^{d}_{class} + L^{d}_{feature} , \tag{6}$$

where L^{s}_{T} and L^{s}_{S} denote the supervised loss of the teacher network prediction and the supervised loss of the student network prediction, respectively. Here, we utilize focal loss (Lin et al. 2017) to compute the prediction loss of L^{s}_{T} and L^{s}_{S}. $L_{clinical}$ represents the self-supervised loss for the clinical data prediction. L^{d}_{class} denotes the classification-level distillation loss of Eq. (4) and $L^{d}_{feature}$ is the feature-level distillation loss of Eq. (5) between the teacher network and the student network. We utilize the loss function of Eq. (6) to train our modality-aware distillation network for MVI prediction.

Experiments

Dataset and Evaluation Metric

Dataset. We collected a inhouse dataset consisting of 270 pathologically confirmed HCC patients with preoperative MRI met the inclusion criteria. The HCC MRI data were taken by a 7-point baseline sample collection protocol (Cong et al. 2016). According to the high-risk factors of adverse outcomes, all 270 patients were classified into M_0 (no MVI), or M_1 (invaded vessels were no more than five and located at the peritumoral region adjacent to the tumor surface within 1 cm), or M_2 (MVI of $>$5 or at $>$1 cm away from the tumor surface), respectively. The collected dataset consists of 128 M_0 patients, 93 M_1 patients, and 49 M_2 patients. Pre phase images (denoted as ”PRE”), hepatobiliary phase images (denoted as ”HBP”), and clinical data are collected for each patient. We do not require any registration operation between HBP and PRE images. Moreover, we utilize a five-fold cross-validation strategy to test our network and state-of-the-art classification methods. Specifically, following the stardard steps of a leave-one-out five-fold cross-validation scheme, we split the whole datasets wih 270 cases (128 M_0 patients, 93 M_1 patients, and 49 M_2 patients) into five folds. In each round of the cross-validation, we take one fold as the testing set and other four folds as the training set. Then, we compute the mean and variance value of five rounds for all evaluation metrics, which are F1-score, AUC, and accuracy, in order to conduct the comparisons between our network and compared methods.

Clinical Data. The clinical data consists of 52 Preoperative laboratory indexes. Non-image clinical data in our work are collected and obtained from the report of blood tests, the patient's medical record report, as well as the MRI hallmarks by the radiologists' reviews. In summary, our clinical data is sufficient since it includes diverse liver detection indicators of the liver in clinical diagnosis.

Evaluation Metrics. We employ three widely-used classification metrics for quantitatively comparing different methods. They are F1-score, Accuracy, and the macro-averaged one-versus-one Area under the curve (AUC). In general, a better HCC's MVI classification result shall have larger values for all three metrics.

Comparisons against State-of-the-art Methods

Compared Methods. We evaluate the effectiveness of our classification network by comparing it against seven state-of-the-art methods, including concatenation-based feature fusion method (Nie et al. 2019) (denoted as ”Concat”), “3DCNN” (Jiang et al. 2021), LSTM-based multi-modality fusion method (Men et al. 2019) (denoted as “LSTM”), M^{2}Net (Zhou et al. 2020), stage wise multi-modality fusion network (Zhou et al. 2019) (denoted as “Concat_2S”), traditional knowledge distillation (Hinton, Vinyals, and Dean 2015) with our module (denoted as “KD_ours”), and similarity-preserving knowledge distillation(Tung and Mori 2019) with our module (denoted as “SP_ours”). For a fair comparison, we obtain the classification results of all competitors by exploiting its public implementations or implementing them by ourselves, and the network parameters of each network are fine-turned to obtain the best classification results for comparisons.

Table 1: Quantitative results (mean ± variance) of our network and state-of-the-art methods on our dataset.

Method	F1-score (%)	Accuracy (%)	AUC (%)	p-value
Concat (Nie et al. 2019)	58.82 ± 1.28	60.75 ± 0.99	67.13 ± 0.73	1.22e-2
3DCNN (Jiang et al. 2021)	43.54 ± 6.20	53.93 ± 2.76	67.00 ± 1.02	9.44e-4
LSTM (Men et al. 2019)	58.20 ± 4.32	61.37 ± 3.24	67.97 ± 1.58	4.29e-3
M^{2}Net (Zhou et al. 2020)	57.45 ± 3.24	59.12 ± 2.87	66.79 ± 1.66	2.57e-3
Concat_2S (Zhou et al. 2019)	58.50 ± 0.69	61.13 ± 0.99	68.84 ± 1.34	1.02e-2
KD_ours (Hinton, Vinyals, and Dean 2015)	61.69 ± 3.08	63.14 ± 2.02	71.01 ± 1.60	2.20e-1
SP_ours (Tung and Mori 2019)	58.55 ± 4.89	61.57 ± 3.65	68.43 ± 2.64	3.43e-3
Our method	$\mathbf{62.16 \pm 3.35}$	$\mathbf{63.38 \pm 3.22}$	$\mathbf{71.46 \pm 1.84}$	

Table 2: Quantitative results of our method and baseline networks of the ablation study. "S" indicate student network, T indicate teacher network.

Method	T/S	F1-score (%)	Accuracy (%)	AUC (%)
S-pre	S	57.67 ± 4.36	60.52 ± 2.65	68.35 ± 1.61
S-hbp	S	59.31 ± 1.73	60.83 ± 1.60	69.35 ± 1.43
S-pre-hbp	S	60.02 ± 0.98	61.95 ± 0.83	69.98 ± 0.96
ours-w/o-hbp	S→T	58.02 ± 3.88	60.57 ± 2.78	69.24 ± 2.23
ours-w/o-pre	S→T	60.41 ± 3.44	62.23 ± 2.26	70.27 ± 1.49
ours-w/o-CA	S→T	61.03 ± 3.74	61.87 ± 3.02	69.71 ± 1.75
Our method	S→T	$\mathbf{62.16 \pm 3.35}$	$\mathbf{63.38 \pm 3.22}$	$\mathbf{71.46 \pm 1.84}$

Quantitative Comparisons. Table 1 reports the mean ± variance results of three metrics for our method and seven compared networks under a five-fold cross-validation experiment on our dataset. From the results, we can find that "KD_ours" has the best performance on three metrics on all compared methods, and they are the F1-score score of 61.69, the Accuracy score of 63.14, and the AUC score of 71.01. More importantly, our method has larger F1-score, Accuracy, and AUC scores than "KD_ours". Specifically, our method has a F1-score improvement of 0.47%, an Accuracy improvement of 0.24%, and an AUC improvement of 0.45%, when compared to KD_ours. Moreover, our method outperforms "KD_ours" and "SP_ours" on all three metrics, which demonstrating the superior performance of our distillation method over "KD_ours" and "SP_ours". We compute p-values between our network and compared methods in terms of the AUC metric, and reported the corresponding p-values in Table 1. Apparently, we can find that all the p-values of our network over compared methods (except KD_ours) are smaller than 0.05. It indicates that our method has a AUC significant improvement between our network and each compared method.

Ablation Study

We also conduct the ablation study experiments to verify the major components in our network design. Here, we construct six baseline networks, and compare the quantitative results of our method and baseline networks.

According to the quantitative results of Table 2, we can find that "S-pre-hbp" has higher F1-score, Accuracy, AUC values than "S-pre" and "S-hbp". It shows that combining the PRE and HBP MRI data together can enhance the MVI classification performance of our student network. Specifically, "S-pre-hbp" further enhances the mean F1-score value from 59.31% to 60.02%, the mean Accuracy value from 60.83% to 61.95%, and the mean AUC value from 69.35% to 69.98%.

According to the quantitative comparisons in Table 2, it can be easily observed that our method has a superior performance of F1-score, Accuracy, and AUC over "ours-w/o-pre", "ours-w/o-hbp" and "ours-w/o-CA". To be specific, the F1-socre, Accuracy, and AUC values of the "ours-w/o-pre" are 58.02%, 60.57%, and 69.24%, while "ours-w/o-hbp" has a F1-score of 60.41%, a Accuracy of 62.23%, and a AUC of 70.27%. Apparently, the HBP MRI data has better MVI prediction results than the PRE MRI data in our method. And the F1-socre, Accuracy, and AUC values of the "ours-w/o-CA" are 61.03%, 61.87%, 69.71. Moreover, compared to the best-performing results of "ours-w/o-pre" and "ours-w/o-CA", our method improves F1-score from 61.03% to 62.16%, Accuracy from 62.23% to 63.38%, and AUC from 70.27% to 71.46%. It demonstrates that removing the PRE MRI data, HBP MRI data or CA module from our network degrades the MVI classification performance of our network.

Conclusion

This work presents a Cross-modality knowledge distillation network (CMD-Net) for a MVI prediction in HCC. Our CMD-Net transfers the teacher network with a non-image clinical modality and a multi-MRI image modality to a student network with only image multi-MRI image by formulating a classification-level distillation and a feature-level distillation. By doing so, with the help of the distilled clinical information, our student network can obtain a superior HCC prediction in the testing stage, which does not have any clinical modality data. In the teacher network, we formulate MRI-clinical-fusion CNNs and a cross attention (CA) module to integrate two groups of the MRI data and the clinical data. Then, we formulate two MRI-only module and a SA module to fuse features from two MRI data in the student network of our MD-Net. Experimental results on our collected dataset and a multi-modal sarcasm detection dataset show that our CMD-Net outperforms state-of-the-art methods in terms of a MVI prediction in HCC.

References

Chong, H.-H.; Yang, L.; Sheng, R.-F.; Yu, Y.-L.; Wu, D.-J.; Rao, S.-X.; Yang, C.; and Zeng, M.-S. 2021. Multi-scale and multi-parametric radiomics of gadoxetate disodium–enhanced MRI predicts microvascular invasion and outcome in patients with solitary hepatocellular carcinoma 5 cm. *European Radiology*, 1–15.

Cong, W.-M.; Bu, H.; Chen, J.; Dong, H.; Zhu, Y.-Y.; Feng, L.-H.; Chen, J.; Committee, G.; et al. 2016. Practice guidelines for the pathological diagnosis of primary liver cancer: 2015 update. *World journal of gastroenterology*, 22(42): 9279.

Duanmu, H.; Huang, P. B.; Brahmavar, S.; Lin, S.; Ren, T.; Kong, J.; Wang, F.; and Duong, T. Q. 2020. Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer Using Deep Learning with Integrative Imaging, Molecular and Demographic Data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 242–252. Springer.

Feng, S.-T.; Jia, Y.; Liao, B.; Huang, B.; Zhou, Q.; Li, X.; Wei, K.; Chen, L.; Li, B.; Wang, W.; et al. 2019. Preoperative prediction of microvascular invasion in hepatocellular cancer: a radiomics model using Gd-EOB-DTPA-enhanced MRI. *European radiology*, 29(9): 4648–4659.

Hinton, G.; Vinyals, O.; and Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Hu, H.-T.; Shen, S.-L.; Wang, Z.; Shan, Q.-Y.; Huang, X.-W.; Zheng, Q.; Xie, X.-Y.; Lu, M.-D.; Wang, W.; and Kuang, M. 2018. Peritumoral tissue on preoperative imaging reveals microvascular invasion in hepatocellular carcinoma: a systematic review and meta-analysis. *Abdominal Radiology*, 43(12): 3324–3330.

Iguchi, T.; Shirabe, K.; Aishima, S.; Wang, H.; Fujita, N.; Ninomiya, M.; Yamashita, Y.-i.; Ikegami, T.; Uchiyama, H.; Yoshizumi, T.; et al. 2015. New pathologic stratification of microvascular invasion in hepatocellular carcinoma: predicting prognosis after living-donor liver transplantation. *Transplantation*, 99(6): 1236–1242.

Imamura, H.; Matsuyama, Y.; Tanaka, E.; Ohkubo, T.; Hasegawa, K.; Miyagawa, S.; Sugawara, Y.; Minagawa, M.; Takayama, T.; Kawasaki, S.; et al. 2003. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. *Journal of hepatology*, 38(2): 200–207.

Jiang, Y.-Q.; Cao, S.-E.; Cao, S.; Chen, J.-N.; Wang, G.-Y.; Shi, W.-Q.; Deng, Y.-N.; Cheng, N.; Ma, K.; Zeng, K.-N.; et al. 2021. Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning. *Journal of Cancer Research and Clinical Oncology*, 147(3): 821–833.

Lee, S.; Kim, S. H.; Lee, J. E.; Sinn, D. H.; and Park, C. K. 2017. Preoperative gadoxetic acid–enhanced MRI for predicting microvascular invasion in patients with single hepatocellular carcinoma. *Journal of hepatology*, 67(3): 526–534.

Lei, Z.; Li, J.; Wu, D.; Xia, Y.; Wang, Q.; Si, A.; Wang, K.; Wan, X.; Lau, W. Y.; Wu, M.; et al. 2016. Nomogram for preoperative estimation of microvascular invasion risk in hepatitis B virus–related hepatocellular carcinoma within the milan criteria. *JAMA surgery*, 151(4): 356–363.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Mazzaferro, V.; Sposito, C.; Zhou, J.; Pinna, A. D.; De Carlis, L.; Fan, J.; Cescon, M.; Di Sandro, S.; Yi-Feng, H.; Lauterio, A.; et al. 2018. Metroticket 2.0 model for analysis of competing risks of death after liver transplantation for hepatocellular carcinoma. *Gastroenterology*, 154(1): 128–139.

Men, S.; Ju, H.; Zhang, L.; and Zhou, W. 2019. Prediction Of Microvascular Invasion Of Hepatocellar Carcinoma With Contrast-Enhanced MR Using 3D CNN And LSTM. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 810–813. IEEE.

Nie, D.; Lu, J.; Zhang, H.; Adeli, E.; Wang, J.; Yu, Z.; Liu, L.; Wang, Q.; Wu, J.; and Shen, D. 2019. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. *Scientific reports*, 9(1): 1–14.

Oh, S. W.; Lee, J.-Y.; Xu, N.; and Kim, S. J. 2019. Video object segmentation using space-time memory networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9226–9235.

Okada, S.; Shimada, K.; Yamamoto, J.; Takayama, T.; Kosuge, T.; Yamasaki, S.; Sakamoto, M.; and Hirohashi, S. 1994. Predictive factors for postoperative recurrence of hepatocellular carcinoma. *Gastroenterology*, 106(6): 1618–1624.

Roayaie, S.; Blume, I. N.; Thung, S. N.; Guido, M.; Fiel, M.-I.; Hiotis, S.; Labow, D. M.; Llovet, J. M.; and Schwartz, M. E. 2009. A system of classifying microvascular invasion to predict outcome after resection in patients with hepatocellular carcinoma. *Gastroenterology*, 137(3): 850–855.

Rodriguez-Peralvarez, M.; Luong, T. V.; Andreana, L.; Meyer, T.; Dhillon, A. P.; and Burroughs, A. K. 2013. A systematic review of microvascular invasion in hepatocellular carcinoma: diagnostic and prognostic variability. *Annals of surgical oncology*, 20(1): 325–339.

Shindoh, J.; Kobayashi, Y.; Kawamura, Y.; Akuta, N.; Kobayashi, M.; Suzuki, Y.; Ikeda, K.; and Hashimoto, M. 2020. Microvascular invasion and a size cutoff value of 2 cm predict long-term oncological outcome in multiple hepatocellular carcinoma: reappraisal of the American Joint Committee on Cancer Staging System and validation using the surveillance, epidemiology, and end-results database. *Liver cancer*, 9(2): 156–166.

Sumie, S.; Nakashima, O.; Okuda, K.; Kuromatsu, R.; Kawaguchi, A.; Nakano, M.; Satani, M.; Yamada, S.; Okamura, S.; Hori, M.; et al. 2014. The significance of classifying microvascular invasion in patients with hepatocellular carcinoma. *Annals of surgical oncology*, 21(3): 1002–1009.

Tung, F.; and Mori, G. 2019. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1365–1374.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *Advances in neural information processing systems*, 5998–6008.

Wang, X.; Girshick, R.; Gupta, A.; and He, K. 2018. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7794–7803.

Xiao, X.; Zhao, J.; and Li, S. 2022. Task relevance driven adversarial learning for simultaneous detection, size grading, and quantification of hepatocellular carcinoma via integrating multi-modality MRI. *Medical Image Analysis*, 81: 102554.

Xu, X.; Zhang, H.-L.; Liu, Q.-P.; Sun, S.-W.; Zhang, J.; Zhu, F.-P.; Yang, G.; Yan, X.; Zhang, Y.-D.; and Liu, X.-S. 2019. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *Journal of hepatology*, 70(6): 1133–1144.

Yang, L.; Gu, D.; Wei, J.; Yang, C.; Rao, S.; Wang, W.; Chen, C.; Ding, Y.; Tian, J.; and Zeng, M. 2019. A radiomics nomogram for preoperative prediction of microvascular invasion in hepatocellular carcinoma. *Liver cancer*, 8(5): 373–386.

Zhang, Y.; Lv, X.; Qiu, J.; Zhang, B.; Zhang, L.; Fang, J.; Li, M.; Chen, L.; Wang, F.; Liu, S.; et al. 2021. Deep Learning With 3D Convolutional Neural Network for Noninvasive Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *Journal of Magnetic Resonance Imaging*.

Zhou, T.; Fu, H.; Zhang, Y.; Zhang, C.; Lu, X.; Shen, J.; and Shao, L. 2020. M2Net: Multi-modal Multi-channel Network for Overall Survival Time Prediction of Brain Tumor Patients. *arXiv preprint arXiv:2006.10135*.

Zhou, T.; Thung, K.-H.; Zhu, X.; and Shen, D. 2019. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human brain mapping*, 40(3): 1001–1016.