Prior Knowledge-Aware Fusion Network for Prediction of Macrovascular Invasion in Hepatocellular Carcinoma

Haoran Lai, Sirui Fu, Jie Zhang, Jianyun Cao, Qianjin Feng, *Member, IEEE*, Ligong Lu, and Meiyan Huang

***Abstract*—Macrovascular invasion (MaVI) is a major threat to survival in hepatocellular carcinoma (HCC), which should be treated as early as possible to ensure safety and efficacy. In this aspect, MaVI prediction can be helpful. However, MaVI prediction is difficult because of the inter-class similarity and intra-class variation of HCC in computed tomography (CT) images. Moreover, existing methods fail to include clinical priori knowledge associated with HCC, leading to incomprehensive information extraction. In this paper, we proposed a prior knowledge-aware fusion network (PKAFnet) to accurately achieve MaVI prediction in CT images. First, a perception module was presented to extract features related to tumor marginal heterogeneity in the graph domain, which contributed to rotation invariance and captured intensity variations of tumor margin. Second, a tumor segmentation network was built to obtain global information of a 3D tumor image and information associated with tumor internal heterogeneity in the image domain. Finally, multi-domain features associated with the tumor margin and tumor region were combined by using a multi-domain attentional feature fusion module. Thus, by incorporating MaVI-related prior knowledge, our PKAFnet can alleviate overfitting, which can improve the discriminative ability. The proposed PKAFnet was validated on a multi-center dataset, and remarkable performance was achieved in an independent testing set. Moreover, the interpretability of perception module and segmentation network were presented in our paper, which illustrated the effectiveness and credibility of PKAFnet. Therefore, the proposed method showed great application potential for MaVI prediction.**

***Index Terms*—Hepatocellular carcinoma, macrovascular invasion prediction, prior knowledge, multi-domain fusion, rotation invariance.**

I. Introduction

LIVER cancer has high malignancy, and it has high incidence and mortality worldwide [1]. Pathologically, hepatocellular carcinoma (HCC) counts for 70% to 85% of liver cancer [1]. For HCC, macrovascular invasion (MaVI) is a major threat to survival, which can cause rapid deterioration and preclude further treatments [2]. Thus, it should be treated as early as possible [3]. Unfortunately, although current methods can easily identify existing MaVI, predicting future MaVI is still a clinical challenge. The lack of precise prediction of future MaVI may cause two problems: first, we may fail to preform closer follow-ups for the high-risk population, which may delay treatments; second, we may be unable to explore whether early combination with immunotherapy can prevent future MaVI, which is suggested by researchers [4]. Thus, besides identifying existing MaVI, predicting future MaVI is necessary to ensure early intervention.

Given its wide availability and short acquisition time, computed tomography (CT) is reliable for MaVI diagnosis [5]. However, MaVI prediction using conventional visual interpretation of CT remains a challenge primarily because of the following two reasons: first, HCCs with similar appearance in CT images have different outcomes (i.e., inter-class similarity, Fig. 1). Figs. 1 (a) and (c) show relatively scattered and small necrosis areas and relatively smooth tumor margins, whereas Figs. 1 (b) and (d) show evident necrosis areas, which are large and relatively concentrated. Second, HCCs with different visual interpretations have similar outcomes (i.e., intra-class variation). Compared with Fig. 1 (b),

Manuscript received 31 December 2021; revised 14 February 2022; accepted 12 April 2022. Date of publication 18 April 2022; date of current version 30 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 81601562, Grant 81974275, and Grant U1501256; in part by the Science and Technology Planning Project of Guangzhou under Grant 201904010417; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515012011; in part by the National Key Research and Development Program of China under Grant 2017YFA0205200; and in part by the Nurture Program of Zhuhai People's Hospital under Grant 2020XSYC-09. *(Haoran Lai and Sirui Fu contributed equally to this work.) (Corresponding authors: Qianjin Feng; Ligong Lu; Meiyan Huang.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures was granted by the Ethics Review Committee of the Zhuhai People's Hospital.

Haoran Lai, Jianyun Cao, Qianjin Feng, and Meiyan Huang are with the School of Biomedical Engineering, the Guangdong Provincial Key Laboratory of Medical Image Processing, and the Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology, Southern Medical University, Guangzhou 510515, China (e-mail: haoranlai@163.com; jianyun_cao2021@163.com; fengqj99@smu.edu.cn; huangmeiyan16@163.com).

Sirui Fu and Ligong Lu are with the Zhuhai Interventional Medical Centre, Zhuhai People's Hospital (Zhuhai Hospital Affiliated with Jinan University), Zhuhai 519000, China (e-mail: bicia870428@sina.com; llg0902@sina.com).

Jie Zhang is with the Department of Radiology, Zhuhai People's Hospital (Zhuhai Hospital Affiliated with Jinan University), Zhuhai 519000, China (e-mail: zhangjie201806@sina.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TMI.2022.3167788, provided by the authors.

Digital Object Identifier 10.1109/TMI.2022.3167788

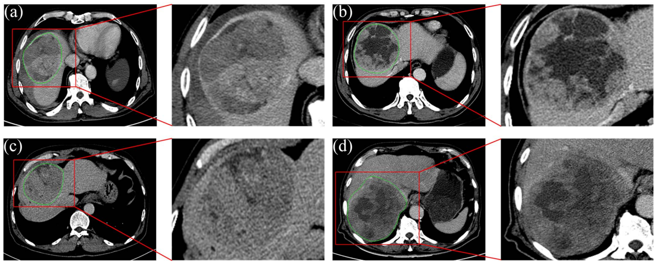

Fig. 1. All the four patients have no MaVIs at diagnosis, where green curves represent tumor contours. (a) and (b) show patients with subsequent MaVIs during follow-ups, whereas (c) and (d) display HCC patients without subsequent MaVI. (a) vs. (c) and (b) vs. (d) present high inter-class similarity, whereas (a) vs. (b) and (c) vs. (d) show high intra-class variation.

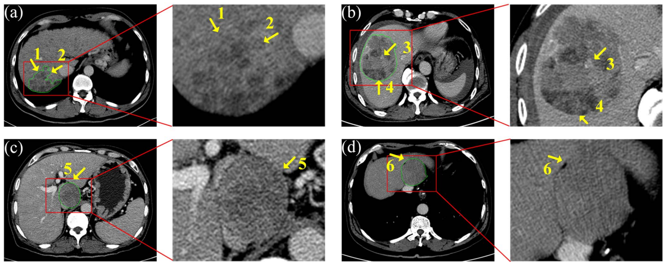

Fig. 2. All the four patients have no MaVIs at diagnosis, where green curves represent tumor contours. (a) and (b) show patients with subsequent MaVIs during follow-ups, whereas (c) and (d) display HCC patients without subsequent MaVI. (a) Corona enhancement and invasive shape are shown, which are indicated by yellow arrows with 1 and 2, respectively. (b) Mosaic architecture is presented within the tumor region, and nodule-in-nodule architecture (indicated by a yellow arrow with 3) and evident necrosis (indicated by a yellow arrow with 4) are shown. (c) HCC (indicated by a yellow arrow with 5) with a smooth margin and complete capsule. (d) HCC (indicated by a yellow arrow with 6) without necrosis, mosaic architecture, and nodule-in-nodule architecture.

Fig. 1 (a) shows a more incomplete capsule, more typical mosaic architecture, and less necrosis areas. Compared with Fig. 1 (c), Fig. 1 (d) presents great difference in tumor size, and it has evident necrosis areas. In addressing these challenges, some existing methods have been proposed to explore the abstract CT imaging features and predict MaVI. For example, Wei *et al.* [6] developed and validated a radiomic method with CT images to preoperatively predict MaVI, which showed that high-dimensional features extracted from CT images were informative. Recently, Wei *et al.* [7] combined clinical and radiomic features extracted from CT images for preoperative MaVI prediction, which suggested that clinical experience and image information were useful features for MaVI prediction. Although promising prediction results were achieved by using traditional radiomic features [6], [7], this kind of features is hand crafted, simple, and presentative, and thus traditional radiomic features are insufficient to represent the image information [8]. In exploring the image information, deep learning methods can be used to extract the informative features from images, which is a potential tool for MaVI prediction.

Among the deep learning methods, convolutional neural network (CNN) is widely used in image classification [9]–[11]. However, using CNN to locate the key position related to the prior knowledge of the task is difficult because the output of the last convolutional layer is fed into a pooling layer to obtain the features that can be used to represent the whole image in the CNN for classification [12]. In a previous study, we have demonstrated that two kinds of information are highly correlated with vascular invasion in HCC: one is related to tumor marginal heterogeneity, such as invasive shape, HCC capsule, and corona enhancement (Fig. 2 (a)); the other is associated with tumor internal heterogeneity, such as mosaic architecture, nodule-in-nodule architecture, and necrosis (Fig. 2 (b)) [13]. Therefore, incorporating priori knowledge related to tumor margin and tumor region into the neural network is crucial to automatically exploit deep features for MaVI prediction.

In this study, a prior knowledge-aware fusion network, referred to as PKAFnet, is presented to combine tumor marginal and internal heterogeneity information for MaVI prediction. First, a 2.5D-based graph convolution network (GCN) is used as a specific perception module for feature extraction of tumor marginal heterogeneity in the graph domain, where rotation invariance and pixel relationships among tumor margin are incorporated into the deep learning framework to extract useful features from the tumor margin. Second, a tumor segmentation-based CNN network is adopted to exploit important features associated with tumor internal heterogeneity in the image domain. Therefore, the features related to tumor marginal and internal heterogeneity can be obtained from the graph and image domains, respectively. Moreover, a multi-domain attentional feature fusion module, namely, MdaFF, is used to combine these multi-domain features for MaVI prediction. Thus, an end-to-end network can be built to comprehensively extract image information related to tumor marginal and internal heterogeneity. The contributions of this work are as follows:

- We introduced PKAFnet for accurate MaVI prediction, where crucial features related to tumor marginal and internal heterogeneity were exploited and combined by using a self-attention module to provide a comprehensive description of the tumors. Based on previous reports, MaVI prediction has yet to be systematically investigated while incorporating deep learning frameworks to exploit variations among tumor margin and tumor region. Therefore, the proposed method may provide a new way to capture tumor-related image variations of MaVI and achieve improved prediction performance for MaVI.
- We proposed a novel GCN-based perception module to explore information of tumor marginal heterogeneity. By extracting the tumor margin and building it as a graph, GCN can utilize translation and rotation invariance within the extracted tumor margin region of interest (ROI). Moreover, GCN can be used to exploit spatial correlation among pixels and extract crucial features from the tumor margin.
- The proposed method was evaluated on a multi-center clinical dataset, which contained 374 patients with HCC collected from five hospitals. The proposed method was trained on patients from four out of five hospitals and tested on patients from the remaining hospital.

Moreover, good prediction performance was achieved by using the proposed method, which demonstrated that important information can be extracted from the HCC images, and thus good generalization can be obtained.

II. Related Work

A. Information Exploration for Tumor Classification

Similar to MaVI prediction, information related to lesion is relatively crucial in lesion classification. Considerable research has been proposed to explore information on lesion classification. Zhou *et al.* [9] presented a multi-task learning method, which included tumor classification and tumor segmentation tasks, to achieve high tumor classification accuracy. The extra task of tumor segmentation can make the CNN focus on the tumor region and distinguish the intensity variations within the tumor region, which is beneficial to extract information related to tumor internal heterogeneity by using the CNN [14]. Liu *et al.* [10] proposed a Siamese network architecture with a margin ranking loss to capture the heterogeneity of lung nodules, which showed good performance in distinguishing the positive and negative samples. Moreover, Afshar *et al.* [11] used a coarse lesion mask, which was generated by a segmentation network, as an extra input to increase network's attention on the lesion region and achieve good lung nodule classification performance. Although potential information of tumor internal heterogeneity can be discovered by using these methods [9]–[11], [14], variant information related to tumor marginal heterogeneity may be ignored because verifying whether the network accurately concerns on the tumor margin is difficult. Hence, making the network focus on the tumor margin is an important research direction.

B. Rotation Invariance

Rotation invariance is a desired property of machine learning methods for medical image analysis [15]–[18]. Most of the existing research developed for rotation invariance of medical images can be roughly grouped into two categories. The first category is to modify the operation of the standard convolution to ensure that the features with rotation invariance can be extracted by the modified convolution. For example, Andrearczyk *et al.* [15] used steerable filters to replace standard filters in the CNN to classify benign and malignant pulmonary nodules. Lafarge *et al.* [16] used a special Euclidean motion group convolution layer to replace the standard convolution layer for multi-organ nuclei segmentation. Although some rotation invariances can be learnt by incorporating these novel filters into the CNN, rotation invariant performance cannot be naturally exhibited in the CNN [19]. The second category is to modify the network inputs to obtain features with rotation invariance. For example, Ebrahim *et al.* [17] adopted data augmentation with rotation in the training stage and made the network learning rich information from rotating data. Moreover, Zhu *et al.* [18] proposed a self-supervised learning by splitting a 3D image to many sub-cubes and applying cube orientation to extract deep features with rotation invariance. Rotation invariance can be included in the CNN by using data augmentation, but few studies incorporated rotation invariance directly into the data as their characteristics.

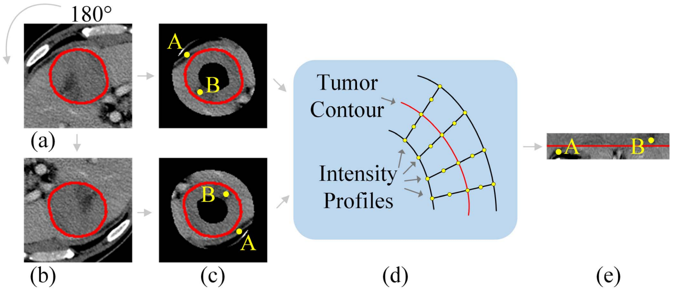

Fig. 3. (a) Cropped HCC CT image, where HCC is indicated by a red curve. (b) Cropped CT image obtained from (a) with 180° rotation. (c) Extraction of the tumor margin area. (d) Extraction of intensity profiles. (e) Construction of the proposed tumor margin ROI. When HCC is rotated 180°, the same tumor margin ROI is obtained. Therefore, rotation invariance is incorporated into tumor margin ROI. However, spatial locations of pixels A and B in (c) are damaged within the extracted ROI in (e).

The abovementioned research has demonstrated the effectiveness of using rotation invariance for medical image analysis. In this paper, a tumor margin extraction method proposed in our previous study [20], which extracted intensities of some pixels within the tumor to outside of the tumor and arranged all of these pixels as a ROI, was introduced to extract features related to tumor marginal heterogeneity and incorporate rotation invariance into the proposed method (Fig. 3). However, information of spatial location will be lost within the extracted ROI because the tumor margin area (Fig. 3 (c)) is reshaped to be a rectangle (Fig. 3 (e)). As illustrated in Fig. 3 (c) and (e), spatial locations of pixels A and B are damaged. The CNN is skilled in exploiting the relationships of spatial locations among image pixels, and thus the CNN may be unsuitable for data whose spatial location is damaged. Although the spatial location information is lost within the tumor margin ROI, the spatial connection of the tumor margin is retained. The tumor margin can be constructed into a graph to utilize the spatial connection, and the spatial connection of the tumor margin can be preserved. Different from the CNN, the GCN is an advanced tool for dealing with graph-structured data. Furthermore, compared with the CNN, the GCN contains rotation invariant behavior in computer vision [19]. Therefore, the GCN can be used to extract the deep features from the graph constructed by the tumor margin, which can retain the spatial connections and utilize rotation invariance.

C. Feature Fusion

As the features extracted from different parts have diverse contributions to the final prediction performance, many useful feature fusion methods have been presented to automatically fuse features from multi-parts and enhance the performance of image classification [21]–[23]. Majumder *et al.* [21] implemented a context-aware bimodal feature fusion by gradually aggregating the features extracted from multiple modalities, which can filter out conflicting or redundant information obtained from different modalities. Chen *et al.* [22] fused

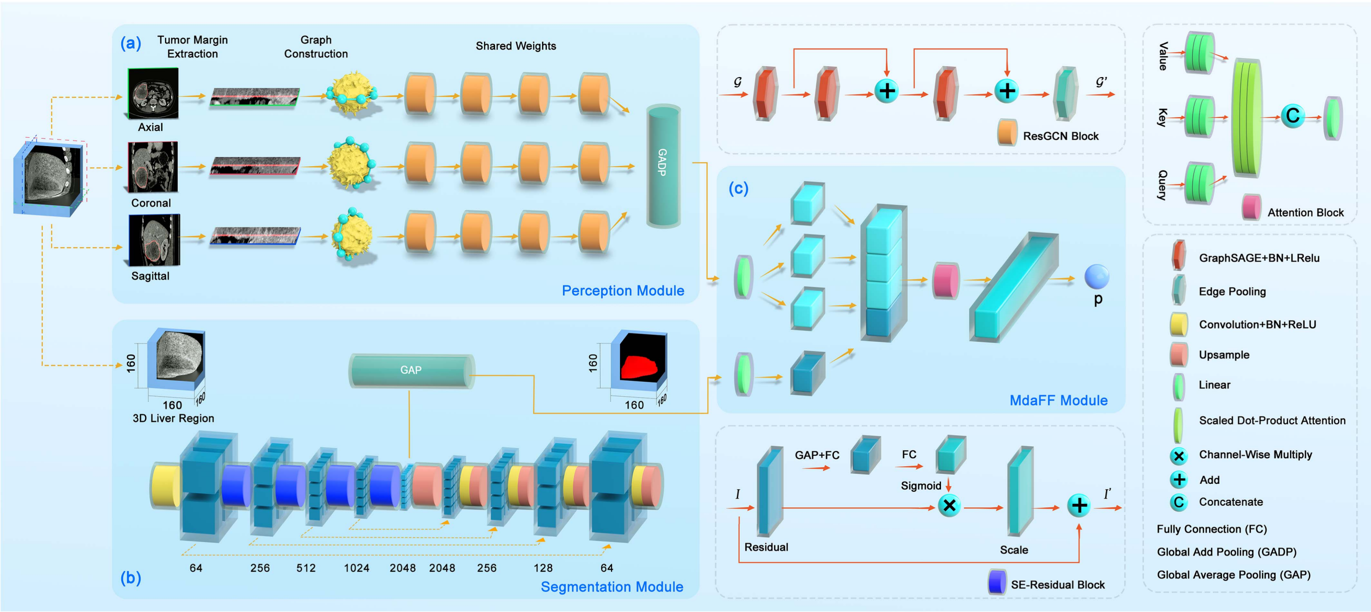

Fig. 4. Overview of the proposed PKAFnet for MaVI prediction.

histopathology and genomic features using Kronecker Product, which could capture the interactions of inter-modalities. Zhu *et al.* [23] parallelly used different pooling operators to generate different feature weights and achieve additive attention for the fusion of multi-instance features, which can consider the specific weights of multi-instance features.

Although the abovementioned feature fusion methods have achieved good performance, two key issues must be addressed to improve performance. The first key issue is inter-correlations among features extracted from different parts. Different information can be provided from different inputs or modalities, which may be complementary and potentially redundant. Therefore, exploring a strategy that can further identify inter-correlation of features extracted from different parts is necessary. The second key issue is calculation errors within feature fusion. Considering that learning fusion weights in feature fusion is an important step in decision-making, the calculation errors of feature fusion should be considered in weight learning. Therefore, a method to reduce the calculation errors in feature fusion should be proposed.

Transformer is designed to explore the correlation of long-range sequences, learn the expressive representations, and reduce the calculation errors by using a multi-head mechanism [24]. It has been introduced to image analysis and replace the convolution operator, which achieves excellent performance [25]. In this paper, we proposed a transformer-style feature fusion module, namely, multi-domain attention feature fusion module (MdaFF), to effectively explore the synergy and correlation of multi-domain features and reduce the calculation errors during feature fusion.

III. Method

The overview of the proposed PKAFnet is illustrated in Fig. 4, which consists of three modules for accurate MaVI prediction. The first perception module is used to extract features related to tumor marginal heterogeneity (Fig. 4 (a)). The second segmentation module is applied to explore features associated with tumor internal heterogeneity (Fig. 4 (b)). The third MdaFF module is adopted to effectively combine features extracted from the tumor margin and tumor region (Fig. 4 (c)). The details of the proposed PKAFnet are provided in the following subsections.

A. Perception Module for Feature Extraction of Tumor Margin

1) Extraction of Tumor Margin ROI: A clever strategy is introduced for the extraction of the tumor contour (red curve in Fig. 3 (a)) to effectively extract features related to tumor marginal heterogeneity and incorporate rotation invariance into the deep learning network. Given a 2D tumor image, the pixels of the tumor margin are initially extracted. In details, the contours of the tumor margin were smoothened by a Gaussian filter to prevent noise disturbance, which can be denoted as follows:

$$\hat{X} = X * G_{1D}(0, \xi), \hat{Y} = Y * G_{1D}(0, \xi) \tag{1}$$

where $G_{1D}(\cdot)$ is the 1-dimensional convolution kernel with zero means, and ξ is the standard deviation. $(X, Y) \in \{(x_i, y_i)\}_{i=1\cdots n_r}$ are the original coordinates of the pixels of the tumor contour, where n_r is the pixel number of the tumor contour for the rth subject. $(\hat{X}, \hat{Y}) \in \left\{\left(\hat{x}_i, \hat{y}_i\right)\right\}_{i=1\cdots n_r}$ are the coordinates smoothened by using the Gaussian filter. Then, for each pixel of the tumor contour, the pixels along the normal of tumor contour are sampled from inside to outside of the tumor to extend the receptive field of the tumor margin. The angles of the normal of tumor contour are calculated by using

$$\theta_i = \arctan\left(\hat{y}_i / \hat{x}_i\right) \tag{2}$$

Using the angle θ_i, the coordinates of the pixels along the normal of tumor contour for the ith pixel are denoted by

$$\tilde{x}_{ij} = \hat{x}_i + l_{ij} \times \cos\theta_i, \tilde{y}_{ij} = \hat{y}_i + l_{ij} \times \sin\theta_i \tag{3}$$

where $j = 1, \ldots, q$, q is the pixel number along the normal of tumor contour, and l_{ij} is the distance between the ith pixel of the tumor contour and the jth pixel along the normal of tumor contour. For a given coordinate $\left(\tilde{x}_{ij}, \tilde{y}_{ij}\right)$, the pixels number of tumor margin may not be the same as that shown in the image; thus, linear interpolation is implemented to extract the corresponding pixel in the image. Using these procedures, the pixels along the normal of tumor contour can be sampled, and intensities of the sampled pixels can be extracted as an intensity profile for each pixel of the tumor contour. Finally, the intensity profiles of all pixels of the tumor contour are paralleled to form a tumor margin ROI $\mathcal{V} \in \mathbb{R}^{n_r \times q}$ and represent the tumor marginal heterogeneity (Fig. 3).

2) Graph Construction for Tumor Margin ROI: Clinically, tumor sizes vary from different patients, and thus the sizes of tumor margin ROIs vary from different tumors. If CNN is used to analyze the ROIs with varied sizes, then ROI interpolation is required to standardize the inputs into similar sizes for the CNN. In this process, noise may be introduced, which can decrease classification accuracy. In addressing this disadvantage of the CNN, a graph can be constructed on the extracted tumor margin ROI, whose spatial connection can be retained. Moreover, the GCN can be subsequently applied on the constructed graph to exploit crucial features related to tumor marginal heterogeneity.

The tumor margin ROI extracted from each 2D slice image can be built as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ represent the nodes and edges, respectively. For a pixel of the tumor contour, its corresponding intensity profile can be regarded as a node representation of the graph. Moreover, the spatial connection of the tumor contour can be used to construct the edge of each node. Based on the relationships among pixels of the tumor contour, an edge exists if two nodes are connected directly, which is set to 1 or 0. Using the constructed nodes and edges, the tumor margin ROI extracted from each slice can be built as a graph, which contains important information of the tumor margin and retains the inherent spatial connection of the tumor contour.

3) 2.5D-Based ResGCN for Feature Exploration of Tumor Margin: In general, radiologists use the axial, coronal, and sagittal views of a 3D CT image to comprehensively analyze tumor information within the 3D image (Fig. 4) [26]. In particular, three 2D images can be extracted from three views and combined as a 2.5D image to approximate the 3D image [26]. Therefore, a 2.5D method is used to extract adequate information of the tumor margin from a 3D image. For each view of the 3D CT image, only the slice with the largest tumor area is used to extract information from tumor margin ROI for graph construction. The 2.5D method can not only extract sufficient pixels from the tumor margin, but also reduce redundant information within all slices of a 3D image, which provides a comprehensive description of the tumor margin region.

For 2D images in each of the three views, tumor margin ROIs of the 2D images are extracted, and then graphs of the tumor margin ROIs are constructed. A residual GCN, namely, ResGCN, is built for 2D images in each view to extract crucial features from the constructed graphs, and thus three ResGCN networks with shared parameters are built. In particular, GraphSAGE [27] and residual modules [28] are included in the proposed ResGCN, where Resnet18 is used as the backbone. Specifically, the number of channels in ResNet18 from the first to fourth blocks are 64, 128, 256, and 512, respectively. GraphSAGE, a widely used type of GCN, is introduced with local neighborhood sampling, aggregation, and combination for inductive node embedding, whose computation and memory complexity are constrained with regard to the size of a graph [27]. Therefore, GraphSAGE was used in this study to address the large graph size. In addition, residual module [28] is used to explore the deep features and address gradient disappearance caused by a deepening network. Moreover, the detail structure of the proposed ResGCN is shown in Fig. 4. The features extracted from three ResGCN with shared parameters can be presented as follows:

$$f_t = \text{ResGCN}\left(\mathcal{G}_t\right) \tag{4}$$

where $\mathcal{G}_t$ is the graph built from the tth view, $\{f_t\}_{t=1}^{3} \in \mathbb{R}^{1\times d_t}$ are deep features extracted by using ResGCN of the tth view in the graph domain, and d_t is the dimension of the extracted features.

B. Segmentation Network for Exploration of Tumor Internal Heterogeneity

In the proposed framework, a MaVI prediction (i.e., classifying patients into with or without subsequent MaVI during follow-ups) is combined with a tumor segmentation task to explore tumor internal heterogeneity. In particular, a 3D Unet-like architecture, which contains a SE-Resnet50-based encoder [29] and a decoder, is used as a subnetwork in the proposed framework for tumor segmentation. A squeeze-and-excitation (SE) module is proposed to obtain expressive representations by explicitly modelling the correlations among the channels of its convolutional features. Therefore, a SE-Resnet50-based network is selected as the encoder of Unet to embrace the advantages of the residual and SE module in the proposed method. The encoder is regarded as a feature extractor, and the decoder is used to locate and classify the voxels. With the guidance of the decoder, the encoder will pay attention to the target region of a segmentation task. In this study, the segmentation target is the tumor region; thus, important information related to tumor internal heterogeneity from the tumor region can be well explored, and deep features can be extracted by using the encoder for the following classification task. The outputs of UNet can be presented as follows:

$$\begin{aligned} [f_1, f_2, f_3, f_4, f_5] &= \text{Encode}\,(I) \\ U &= \text{Decode}\,([f_1, f_2, f_3, f_4, f_5]) \end{aligned} \tag{5}$$

where Encode(·) and Decode(·) respectively represent the functions of the encoder and decoder in the UNet. $I \in \mathbb{R}^{C\times D\times H\times W}$ is the input of UNet, where C, D, H, and W are the channel number, depth, height, and width of the input image, respectively. Skip connection was also used in the proposed method. Thus, $f_i \in \mathbb{R}^{d_i\times\frac{D}{2^i}\times\frac{H}{2^i}\times\frac{W}{2^i}}$ (the output of Encode(·)) is a list of features of different scales, where $i = 1,\ldots,5$, and d_i is the channel number of f_i in the ith block (Fig. 4 (b)). $U \in \mathbb{R}^{M\times D\times H\times W}$ is the estimated probability of tumor segmentation, where M is the segmentation channel of liver tumor (including two classes: 0 for background and 1 for liver tumor). Moreover, global average pooling (GAP) is implemented on f_5 to utilize the extracted features from the encoder.

$$f_p = \mathrm{GAP}(f_5) \tag{6}$$

where $f_p \in \mathbb{R}^{1\times d_5}$ is the deep features that include important information of tumor internal heterogeneity in the image domain, which will be applied in the following feature fusion.

C. MdaFF Module for Fusion of Multi-Domain Deep Features

In utilizing the multi-domain deep features extracted from the tumor margin and tumor region by using the 2.5D-based ResGCN and 3D ResUnet, respectively, a transformer-style MdaFF module is implemented, which is used to fuse the multi-domain features and preserve useful information generated from different subnetworks. Inspired by the idea of transformer [24], the proposed MdaFF module also consists of a scaled dot-product attention and a multi-head attention. The proposed MdaFF module captures correlations among multi-domain features of different subnetworks and adaptively learns the weights of multi-domain features; thus, this module can well fuse the multi-domain features for decision-making.

1) Scaled Dot-Product Attention: Before performing feature fusion, the deep features $\{f_t\}_{t=1}^{3}$ and f_p generated from four subnetworks (i.e., three 2D ResGCNs and a segmentation subnetwork) are first mapped into a unified feature space, and then the mapped features are concatenated to form the multi-domain features $L \in \mathbb{R}^{s\times d}$, where s and d are the subnetwork number and feature dimension, respectively. Subsequently, L is translated to query $Q \in \mathbb{R}^{s\times d_k}$, and key-value pairs $K \in \mathbb{R}^{s\times d_k}$ and $V \in \mathbb{R}^{s\times d_v}$ by using three projection matrices (i.e., $W^Q \in \mathbb{R}^{d\times d_k}$, $W^K \in \mathbb{R}^{d\times d_k}$, and $W^V \in \mathbb{R}^{d\times d_v}$), where d_k and d_v are the feature dimensions of projection matrices ($d_k = d_v = d/4$). Finally, the output of the self-attention is a scaled dot-product, which can be formulated as follows:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(QK^T/\sqrt{d_k}\right)V \tag{7}$$

In the MdaFF module, multi-domain features can be automatically weighted and fused by using the scaled dot-product attention, which conduces to explore the correlation of multi-domain features.

2) Multi-Head Attention: Considering that performing a single attention function may lead to calculation error, a parallel multi-head attention strategy is implemented on the MdaFF module. First, L is translated to Q_i, K_i, and V_i by h times with different projection matrices ($W_i^Q \in \mathbb{R}^{d\times d_k}$, $W_i^K \in \mathbb{R}^{d\times d_k}$, $W_i^V \in \mathbb{R}^{d\times d_v}$, $i = 1,\ldots,h$), which are recoded as heads. Second, all heads are fed into a scaled dot-product attention for parallel calculation. Finally, the outputs of the self-attention from all heads are concatenated for final projection by using $W^O \in \mathbb{R}^{(hd_v)\times d}$, which can be defined as follows:

$$f_m = \mathrm{Concat}\left(\mathrm{head}_1,\ldots,\mathrm{head}_h\right)W^O \tag{8}$$

where $\mathrm{head}_i = \mathrm{Attention}(Q_i, K_i, V_i)$, and $f_m \in \mathbb{R}^{s\times d}$ is the fused features by using the MdaFF module. In the MdaFF module, multi-head attention can be used to generate precise fusion features for decision-making. Then, f_m is reshaped to a vector $f_v \in \mathbb{R}^{(sd)\times 1}$ followed by a fully connection layer and a non-linear activation function; thus, the final classification prediction can be defined as follows:

$$p = \sigma\left(W^c f_v\right) \tag{9}$$

where $W^c \in \mathbb{R}^{1\times(sd)}$ is the weights of the fully connection layer; $\sigma(\cdot)$ is a sigmoid activation function, and $p \in [0, 1]$ is the estimated probability of the classification network.

D. Loss Function

Our proposed method is a multi-task learning framework, including MaVI prediction and tumor segmentation tasks. Moreover, the MaVI prediction task can be regarded as a classification task (i.e., classifying patients into with or without subsequent MaVI during follow-ups). Therefore, the total loss function consists of two parts, including a classification loss $\mathcal{L}_{class}$ for tumor classification and a segmentation loss $\mathcal{L}_{seg}$ for tumor segmentation.

$$\mathcal{L}_{total} = \mathcal{L}_{class} + \mathcal{L}_{seg} \tag{10}$$

1) Classification Loss Function: Classification loss $\mathcal{L}_{class}$ consists of a focal loss $\mathcal{L}_{focal}$ [30] and a margin ranking loss $\mathcal{L}_{mr}$ [10], which can be defined as follows:

$$\mathcal{L}_{class} = \mathcal{L}_{focal} + \mathcal{L}_{mr} \tag{11}$$

As listed in Table I, MaVI prediction suffers from class imbalance issue in this study. In addressing this issue, focal loss, which can alleviate class imbalance and improve the ability to distinguish difficultly classified samples, is applied as part of classification loss in this study:

$$\mathcal{L}_{focal} = \frac{1}{N}\sum_{i=1}^{N} -\alpha\left(1-p\right)^{\gamma}\log\left(p\right) \tag{12}$$

where $\alpha \in [0, 1]$ and $1-\alpha$ are the weighting factor for class 1 and class 0 in class imbalance, respectively, which are adjusted on the basis of the ratio between positive and negative samples; γ is the focusing parameter to smoothen the weights of difficultly classified samples, and N is the training number.

The margin ranking loss is implemented to capture the differences between positive and negative samples, which can be defined as follows:

$$\mathcal{L}_{mr} = \frac{1}{2N}\sum_{i=1,j=1}^{N} \max\left(0, \varphi - \delta\left(p_i, p_j\right)\times\left(c_i - c_j\right)\right) \tag{13}$$

TABLE I
SUBJECT NUMBER (WITH/WITHOUT SUBSEQUENT MAVI) IN DIFFERENT HOSPITALS

Training set				Independent testing set
Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5
98 (10/88)	119 (14/105)	72 (10/62)	44 (2/42)	41 (9/32)
333 (36/297)				

where $c_i \in [0, 1]$ and $c_j \in [0, 1]$ denote the ground truth of the classification task for the ith and jth samples, respectively. p_i and p_j are the estimated probabilities of the network for the ith and jth samples, respectively. φ is the margin parameter. $\delta(\cdot)$ is the indicator function, which can be defined as follows:

$$\delta\left(p_i, p_j\right) = \begin{cases} 1 & p_i \geq p_j \\ -1 & p_i \leq p_j \end{cases} \tag{14}$$

All paired samples are calculated on a min-batch for every iteration to simply implement the margin ranking loss.

2) Segmentation Loss Function: Segmentation loss $\mathcal{L}_{seg}$ consists of a cross-entropy loss $\mathcal{L}_{ce}$ and a dice loss $\mathcal{L}_{dice}$, which are useful loss functions in image segmentation [31], and it can be formulated as follows:

$$\mathcal{L}_{seg} = \lambda \mathcal{L}_{ce} + (1-\lambda)\mathcal{L}_{dice} \tag{15}$$

$$\mathcal{L}_{ce} = \frac{1}{NDHW}\sum_{i=1}^{N}\sum_{k=1}^{DWH} -\log\left(\mu_k\right) \tag{16}$$

$$\mathcal{L}_{dice} = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{2\sum_{k=1}^{DWH}\mu_k \upsilon_k}{\sum_{k=1}^{DWH}\mu_k + \sum_{k=1}^{DWH}\upsilon_k}\right) \tag{17}$$

where λ is used to balance the effects of $\mathcal{L}_{ce}$ and $\mathcal{L}_{dice}$, and $\mu_k \in U$ and $\upsilon_k \in V$ are the estimated probability and ground truth of the kth voxel, respectively. $V \in \mathbb{R}^{M \times D \times H \times W}$ is the ground truth of tumor segmentation.

IV. EXPERIMENTAL RESULTS

In this section, we presented the studied materials, implementation settings, and experimental results of the proposed method for MaVI prediction on a multi-center clinical dataset. Moreover, we implemented our network by using PyTorch and performed all experiments on a server with one NVIDIA GeForce 2080Ti GPU.

A. Materials

The CT scans of HCC used in the proposed method were collected from five Chinese hospitals: Nanfang Hospital (Hospital 1), Yangjiang People's Hospital (Hospital 2), Zhuhai People's Hospital (Hospital 3), Zhongshan City People's Hospital (Hospital 4) and Shenzhen People's Hospital (Hospital 5). The sample number and CT parameters obtained from different hospitals are listed in Tables I and II, respectively. All of the samples were diagnosed of HCC between April 2007 and November 2016, and they were followed up until December 2019. Only those who met the following conditions were accepted: 1) HCC was diagnosed clinically or pathologically; 2) CT images were recorded at the time of diagnosis; 3) patients were initially treated by liver resection, transarterial chemoembolization, or ablation as recommended by the guidelines; 4) no extrahepatic metastasis or MaVI was identified during diagnosis; 5) subsequent MaVI was identified during follow-ups or no subsequent MaVI for at least 1 year was identified unless death occurred.

This retrospective study was approved by the Ethics Review Committee of the Zhuhai People's Hospital and written informed consent was waived for its retrospective design. All patients' data were anonymized before analysis.

The liver tumor and liver region in CT scans were manually outlined by radiologists for every patient, where the manually outlined tumor contours were used for the extraction of tumor margin ROIs. Radiologists can better identify the HCC capsule in the portal phase than in the arterial phase [32], which may influence the segmentation accuracy; hence, tumor segmentation was performed in the portal phase. According to the modified response evaluation criteria in solid tumor assessment [33], one target lesion was used for segmentation depending on the longest diameter and suitability when multiple lesions were presented.

The determination of MaVI was based on the typical CT findings: enhancement of arterial phase imaging, expanding vessel, and/or direct extension into the vasculature [34]. All MaVIs were assessed independently by two radiologists (Sirui Fu and Jie Zhang), who have 10 years of work experience. When disagreement occurs, a third radiologist (Ligong Lu with over 20 years of working experience) performed another independent assessment. The final result was made according to the agreement of at least two of the three radiologists.

Patients from one hospital were used as an independent testing set to verify the performance of PKAFnet. The patients in the remaining four hospitals were combined as a training set to train and tune the network. All of the samples were preprocessed as follows. First, the intensity values of CT scans were truncated to the range of $[-17, 201]$ HU to eliminate disturbance of irrelevant information, which was obtained by using the intensity values at 0.5% and 99.5% of the histogram of the voxels within the tumor areas in the training set. Second, the mean and deviation of the training set were calculated, and all samples in the training and testing sets were standardized by subtracting the mean and dividing the deviation. Third, the multi-center dataset has the same in-plane resolution of 0.645 mm but different slice spacing from 1 mm to 5 mm. Therefore, all CT scans were resampled to the medium resolution (i.e., 3 mm) of the five centers by using bilinear interpolation. Following these pre-processed steps, three slices with the largest tumor area were selected from three views, which can be used to extract tumor margin ROIs and build graphs, and the graphs can be used as inputs of the ResGCNs. Furthermore, for the input of the segmentation subnetwork, only the liver and liver tumor region were retained in the 3D CT scans, and all CT scans were resized to 160 × 160 × 160.

B. Experimental Setup

In MaVI prediction experiments, four out of five hospitals were selected as the training set (total 333 subjects), whereas

TABLE II

CT Parameters in Different Hospitals, Where PB, SSDF, TK, TC, RT, DC, FOV, PM, PSH, FST, and ST Respectively Denote Philips Brilliance, Siemens SOMATOM Definition Flash, Tube Voltage, Tube Current, Rotation Time, Detector Collimation, Field of View, Pixel Matrix, Filter Sharp, Filter Standard, and Slice Thickness

Hospital	Scanner	TV	TC	RT	DC(mm)	FOV(mm)	PM	Reconstruction	PST(mm)
1	PB	120	142	0.75	128 × 0.625	300 × 300	512 ×512	FSH(C)	5
2	PB	120	300	0.75	64 × 0.625	350 × 350	512 × 512	FSH(C)	2
3	SSDF	120	160	0.5	64 × 0.625	350 × 350	512 × 512	FSH(C)	2/5
4	PB	120	250	0.5	128 × 0.625	350 × 350	512 × 512	FST(B)	5
5	PB	120	250	0.5	64 × 0.625	500 × 500	1024 × 1024	FSH(C)	5
	SSDF	120	160	0.5	64 × 0.625	350 × 350	512 × 512	FSH(C)	5

the one remaining hospital was selected as the independent testing set (total 41 subjects). Moreover, a fivefold cross-validation strategy was used in the training set to select the best hyperparameters. All subjects within the training set were divided into five subsets with the same proportion of each class label. One of the five subsets was successively selected as a validation set for each run of the fivefold cross-validation to optimize hyperparameters and prevent overfitting, and the four remaining subsets were combined for model training. Finally, the best hyperparameters were obtained, and the model with the best hyperparameters was retrained on the entire training set for testing on the independent testing set. Specifically, the aforementioned strategy was applied in all experiments in this study for a fair comparison. For extraction of tumor margin ROIs, pixel number along the normal of tumor contour q was set to 31 unless otherwise specified. For the Unet, the channel number C, the segmentation number M, depth D, height H, and width W were set to 1, 1, 160, 160, 160, respectively. For the MdaFF module, the unified dimension d and the head number h were set to 1024 and 4, respectively. For the loss function, γ, λ, and α were set to 2, 0.6, and 0.8, respectively, in the experiments. In the training stage, stochastic gradient descent with momentum was used as the optimizer, whose learning rate, batch size, and hyper-parameter of momentum were 0.001, 4, and 0.9, respectively. Some simple online data augmentation operators, including random flipping, rotating, zooming, and Gaussian noise, were used on the training set to alleviate overfitting. Codes are available at https://github.com/Meiyan88/PKAFnet.

Severe class imbalance was observed in the MaVI dataset; thus, area under the receiver operating characteristic curve (AUROC), area under the precision recall curve (AUPRC), balance accuracy (BACC), weighted F1 score (W-F1), sensitivity (SEN) and specificity (SPE) were adopted as the quantitative metrics to assess the prediction performance, which have been proven effective in addressing class imbalance problem [35]–[37].

Equations (18)-(25) were used to calculate the aforementioned metrics, where TP, TN, FP, and FN were regarded as true positive, true negative, false positive, and false negative values, respectively. In calculating AUROC, we first defined the receiver operating characteristic curve (ROC), which used the values of true positive rate (18) as the y-axis and values of false positive rate (19) as the x-axis. Then, AUROC can be obtained by calculating the area under the ROC. Similarly, in achieving AUPRC, we first defined precision recall curve (PRC), which used the precision values (20) as the y-axis and recall values (21) as the x-axis. Subsequently, AUPRC can be achieved by calculating the area under the PRC. Moreover, SEN, SPE, BACC, and W-F1 can be calculated by using (18), (21), (22) and (23), respectively. Specifically, Youden index [38] was applied on the training set to achieve the threshold of high- and low-risk populations. With the achieved threshold, samples with predicted probabilities higher than the threshold were defined as high-risk population (i.e., predicted positive), otherwise, the samples were defined as low-risk population (i.e., predicted negative). Therefore, TP, TN, FP, and FN can be determined, respectively, for the calculation of SEN, SPE, BACC, and W-F1.

$$\mathrm{TPR} = \mathrm{recall} = \mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{18}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}} \tag{19}$$

$$\mathrm{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{20}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{21}$$

$$\mathrm{BACC} = \frac{\mathrm{SEN} + \mathrm{SPE}}{2} \tag{22}$$

$$\mathrm{W\text{-}F1} = \frac{\mathrm{PS} \times \mathrm{F1}_{\mathrm{PS}} + \mathrm{NS} \times \mathrm{F1}_{\mathrm{NS}}}{\mathrm{PS} + \mathrm{NS}} \tag{23}$$

where PS and NS are the number of positive and negative samples on the independent testing set, respectively. $\mathrm{F1}_{\mathrm{PS}}$ and $\mathrm{F1}_{\mathrm{NS}}$ can be defined as follows:

$$\mathrm{F1}_{\mathrm{PS}} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{24}$$

$$\mathrm{F1}_{\mathrm{NS}} = \frac{2 \times \mathrm{TN}}{2 \times \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{25}$$

In our MaVI dataset, the number of negative samples is more than that of positive samples, and the accurate prediction of patients with MaVI in the future is important in this study. On the one hand, maximizing AUROC aims to rank the prediction score of any positive samples higher than any negative samples, which is suitable for handling imbalanced data distribution and uses 0.500 as baseline for evaluation. On the other hand, AUPRC aims to evaluate the ability of the model in classifying positive samples correctly in an imbalanced data distribution. In particular, the baseline of

TABLE III
CLASSIFICATION PERFORMANCE OF DIFFERENT FEATURE EXTRACTION METHODS ON TUMOR MARGIN FOR MAVI PREDICTION

Method	AUROC	AUPRC	BACC	W-F1	SEN	SPE
FE1	0.785	0.627	0.748	0.753	0.719	**0.778**
FE2	0.861	0.716	0.722	0.859	**1.000**	0.444
PKAFnet	**0.865**	**0.758**	**0.802**	**0.875**	0.938	0.667

AUPRC can be calculated by using (26), which is equal to 0.220 in the independent testing set.

$$\text{baseline}_{\text{AUPRC}} = \frac{\text{PS}}{\text{PS} + \text{NS}} \tag{26}$$

Moreover, maximizing BACC and W-F1 aims to distinguish high- and low-risk populations well in an imbalanced data distribution. Finally, the balance between SEN and SPE is required to achieve good performance in distinguishing high- and low-risk populations.

C. Performance of Different Feature Extraction Methods on Tumor Margin

In this section, the proposed method was compared with two other feature extraction methods on the tumor margin to evaluate the effectiveness of tumor margin extraction and ResGCN. For the first method (denoted as FE1), the tumor margin area was extracted directly (Fig. 3 (c)), and then the extracted area was reshaped to a unified size of 160 × 160 and used as input for the following feature extraction network. Given the spatial locations of the extracted area in this case, the CNN can be used to explore the information provided by the tumor margin. Moreover, a 2.5D-based ResCNN (Resnet18 was used as backbone) was applied to replace the 2.5D-based ResGCN in the proposed perception module. For the second method (denoted as FE2), the tumor margin ROI was first extracted by using the proposed method (Fig. 3). Then, zero padding was performed on the extracted tumor margin ROIs, and all ROIs had a similar size of 1000 × 31. Finally, a 2.5D-based ResCNN (Resnet18 was used as backbone) was constructed to exploit information from the tumor margin ROIs with the same size. In these two compared methods, the segmentation subnetwork and MdaFF module were also included, and all hyper-parameters were turned carefully as in the proposed method to make fair comparison. As shown in Table III, the prediction performance of FE2 is better than that of FE1, which indicates the effectiveness of applying the proposed tumor margin extraction method in incorporating rotation invariance into the ResCNN network. Furthermore, a good prediction performance was observed in the proposed method, which indicated that spatial connections of image pixels can be well captured, and tumor marginal heterogeneity can be exploited by using the proposed ResGCN.

D. Robustness of Tumor Margin Extraction

Three experiments were implemented to assess the robustness of the proposed tumor margin extraction method. First, some artificial segmentation errors were introduced into the manual tumor segmentation masks (denoted as disturbed segmentation), and tumor margins were extracted on the basis of modified inaccurate tumor contours for the independent testing set. Specifically, the original manual segmentation masks were randomly enlarged or shrunk by shifting the tumor contour points along the normal tumor contour (Fig. 5). Point movements were set to 1–8 mm. Moreover, the proposed method, which was trained on the basis of the original manual segmentation masks, was tested on the data with artificial segmentation errors. Second, 10% of patients (39 samples) were randomly selected from five hospitals. Every tumor was then manually outlined three times: one radiologist outlined twice (defined as intra-observer segmentation), and another radiologist outlined once (defined as inter-observer segmentation). Therefore, three kinds of manual segmentation masks can be obtained for each of the 39 samples (Fig. 6). Subsequently, all 39 samples with their three corresponding kinds of manual segmentation masks were used as inputs for our proposed PKAFnet, which was trained on the basis of the entire training set, to achieve three groups of prediction probabilities. Dice loss $\mathcal{L}_{dice}$ (17) was used for these experiments to evaluate the manual segmentation variability. Moreover, paired Wilcoxon rank sum test [39] was applied to compare the prediction differences between manual and disturbed segmentations as well as intra- and inter-observer segmentations. Third, the pixel number along the normal tumor contourr q was varied (21, 31, and 41), and tumor margins were extracted on the basis of these varied q values. The proposed PKAFnet was conducted on the extracted tumor margins in the three experiments, and all hyperparameters were turned carefully as in the proposed method for a fair comparison.

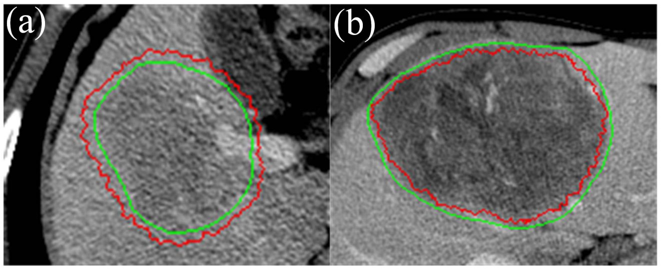

Fig. 5. Green curves represent tumor contours outlined by radiologists, whereas red curves represent (a) enlarged or (b) shrunk tumor contours.

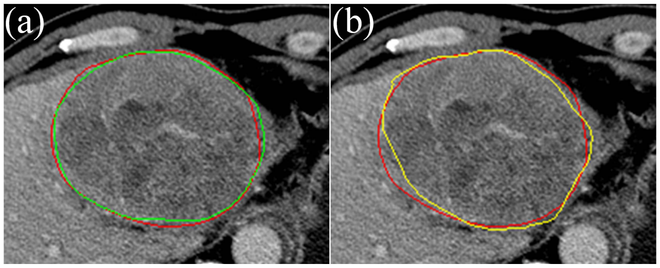

Fig. 6. (a) Intra-observer segmentations, wherein the red and the green curves represent tumor contours outlined twice by a radiologist. (b) Inter-observer segmentation, wherein the yellow and the red curves represent tumor contours outlined by two radiologists.

TABLE IV
EFFECTS OF MANUAL SEGMENTATION VARIABILITY ON PKAFNET, WHERE IQR IS SHORT FOR INTERQUARTILE RANGE

	Dice loss(median (IQR))	p-value
Manual and disturbed segmentation	0.287 (0.084, 0.651)	0.848
Intra- and inter-observer segmentation		
Intra-observer segmentation	0.145 (0.086, 0.276)	0.715
Inter-observer segmentation	0.169 (0.134, 0.297)	0.871

TABLE V
PREDICTION PERFORMANCE OF PAKFNET WITH DIFFERENT q VALUES

q	AUROC	AUPRC	BACC	W-F1	SEN	SPE
21	0.819	0.717	0.748	0.753	0.719	**0.778**
31	**0.865**	**0.758**	**0.802**	**0.875**	0.938	0.667
41	0.837	0.747	0.762	0.869	**0.969**	0.556

Table IV shows that the influence of the proposed method with disturbed segmentation has no statistical significance ($p = 0.848$) compared with the results achieved by the proposed method with manual segmentation. Moreover, a similar prediction probability was achieved by using the proposed method ($p = 0.715$ and $p = 0.871$) with intra- and inter-observer segmentations. Differences might exist in the segmentation process ($\mathcal{L}_{dice} > 0$); however, these results suggest the robustness of the proposed method on manual segmentation variability. The prediction performance improved (except for SPE) for the third experiment by increasing q from 21 to 31, whereas the prediction performance decreased (except for SEN) by increasing q from 31 to 41 (Table V). The reason may lie in the insufficient information within the tumor margin when q was small and the presence of redundant information within the tumor margin when q was large (especially in the small tumor cases). Therefore, an appropriate q value was selected in this study. Moreover, the prediction performance of the proposed method with different q values fluctuated slightly, which further proves the robustness of the proposed tumor margin extraction method.

E. Performance of the Proposed Method With Different Backbones

Two experiments were conducted to evaluate the effects of replacing the backbone with other popular architectures on the performance of the proposed method. First, ResNet34 and ResNet50 were used to replace ResNet18, respectively, in the perception module. Moreover, the architectures used in the segmentation subnetwork and MdaFF module were the same as the proposed method to make a fair comparison. Second, ResNeXt50(32 × 4d) and DenseNet169 were applied to replace SE-ResNet50, respectively, in the segmentation network [40], [41]. Moreover, the architectures used in the perception and MdaFF modules were the same as the proposed method for fair comparison. On the one hand, expressive representations without redundancy can be achieved by ResNeXt- and DenseNet-based networks using group convolution and feature reusing, respectively [40], [41]. Moreover, the expressive representations without redundancy can be also achieved using SE-ResNet-based networks as mentioned in Section III.B. On the other hand, the performance of SE-ResNet50, ResNeXt50(32 × 4d), and DenseNet169 are better than that of ResNet50 based on previous works [40], [41]. Therefore, ResNeXt50(32 × 4d) and DenseNet169 were applied to compare with SE-ResNet50 in this study. As shown in Table VI, as GCN deepened, the prediction performance decreased (the first two rows and last row in Table VI), which may result from the large parameter number and gradient vanishing in deep networks. For the second experiment, the prediction performance of using DenseNet169 as backbone was better than that of using ResNeXt50 as backbone in segmentation subnetwork, which may be due to compacting representations and reducing feature redundancy using features reusing in DenseNet169. Moreover, the best prediction performance was achieved using SE-ResNet50 as the backbone in the segmentation subnetwork, which may benefit from the expressive representations obtained by explicitly modelling the correlations among the channels of its convolutional features used in the SE module.

TABLE VI
PREDICTION PERFORMANCE OF USING DIFFERENT BACKBONES IN PKAFNET FOR MAVI PREDICTION

Method	AUROC	AUPRC	BACC	W-F1	SEN	SPE
Resnet34+SEResnet50	0.781	0.669	0.747	0.847	**0.938**	0.556
Resnet50+SEResnet50	0.764	0.640	0.708	0.750	0.750	0.667
Resnet18+Densenet169	0.861	0.756	0.755	0.811	0.844	0.667
Resnet18+ResneXt50	0.840	0.750	0.733	0.731	0.688	**0.778**
Resnet18+SEResnet50	**0.865**	**0.758**	**0.802**	**0.875**	**0.938**	0.667

F. Ablation Experiments

Each module was discarded from the proposed method to investigate the influence of each module (i.e., perception, segmentation, and MdaFF modules) on the performance of MaVI prediction. First, when the perception module was removed from the proposed method, only tumor segmentation network was retained, and deep features extracted from the encoder of the segmentation network were used for prediction. Moreover, the MdaFF module was discarded in this case. Second, when the segmentation network was removed, deep features extracted from three 2D ResGCNs were fused by using the MdaFF module, and the fused features were fed to a classifier for MaVI prediction. Third, when the MdaFF module was discarded, deep features provided by the perception module and the encoder of tumor segmentation network were concatenated directly and then fed to a classifier for MaVI prediction. The ablation results are shown in Table VII. As shown in Table VII, the following results are observed: first, a relatively low prediction performance is obtained by using segmentation network alone, which illustrates that the information of tumor internal heterogeneity may be insufficient for MaVI prediction. Second, prediction performance can be improved by combining perception and MdaFF modules, which demonstrates that the discriminative feature representations can be effectively learnt from the tumor margin by using ResGCN for

TABLE VII
ABLATION EXPERIMENTS OF THE PROPOSED METHOD FOR MAVI PREDICTION

Perception	Segmentation	MdaFF	AUROC	AUPRC	BACC	W-F1	SEN	SPE
-	√	-	0.744	0.603	0.708	0.750	0.750	**0.667**
√	-	√	0.826	0.700	0.731	0.826	0.906	0.556
√	√	-	0.761	0.639	0.667	0.823	**1.000**	0.333
√	√	√	**0.865**	**0.758**	**0.802**	**0.875**	0.938	**0.667**

TABLE VIII
PREDICTION PERFORMANCE OF THE PROPOSED METHOD BY USING DIFFERENT HOSPITAL DATASETS AS THE INDEPENDENT TESTING SET

Testing set	AUROC	AUPRC (baseline)	BACC	W-F1	SEN	SPE
Hospital 1	0.802	0.526 (0.102)	0.729	0.783	0.758	0.700
Hospital 2	0.783	0.521 (0.118)	0.719	0.806	0.796	0.643
Hospital 3	0.802	0.651 (0.139)	0.749	0.860	0.898	0.600
Hospital 5	0.865	0.758 (0.220)	0.802	0.875	0.938	0.667

MaVI prediction. Third, compared with the use of segmentation network alone, prediction performance is enhanced slightly by combining perception module and segmentation network. Simple concatenation of different domain features may result in generation of redundant features and limited increase of prediction accuracy. Finally, the best prediction performance is achieved by combining three modules in the proposed method, which demonstrates that the correlations of multi-domain features can be explored by using the MdaFF module, and information extracted from the tumor margin and tumor region can contribute to MaVI prediction.

G. Robustness of the Proposed Method on Multi-Center Dataset

Each of the five hospitals was regarded as an independent testing set (except Hospital 4) to evaluate the generalization ability of the proposed PKAFnet. Specifically, only two positive samples were found in Hospital 4 (the ratio of positive and negative samples is 1:21), which suffers from the most severe class imbalance issue among the five hospitals. As shown in Table VIII, the worst and best prediction performance were achieved when using Hospitals 2 and 5 as the independent testing sets, respectively. The maximum differences in AUROC, BACC, and W-F1 values were less than 0.100, whereas the ratio of negative and positive samples ranged from 3.6 to 8.8 on the multi-center dataset (except Hospital 4). These results indicate that relatively stable performance can be achieved by using the proposed PKAFnet on a multi-center dataset for MaVI prediction; thus, the result illustrates the good generalization ability of the proposed PKAFnet.

H. Comparison With Other Methods

The proposed method was compared with some common classification methods, including traditional machine learning and two deep learning methods, to further evaluate the effectiveness of the proposed method on MaVI prediction.

TABLE IX
COMPARISON OF DIFFERENT CLASSIFICATION METHODS ON THE MAVI DATASET

Method	AUROC	AUPRC	BACC	W-F1	SEN	SPE
Radiomics+SVM	0.642	0.291	0.622	0.684	0.688	0.556
VGG	0.631	0.375	0.628	0.737	0.813	0.444
SE-ResNet50	0.667	0.421	0.675	0.796	0.906	0.444
PKAFnet	**0.865**	**0.758**	**0.802**	**0.875**	**0.938**	**0.667**

1) Traditional Machine Learning Method: We adopted the strategy of combining radiomic features and support vector machine (SVM) as traditional method for comparison. First, 1220 radiomic features were extracted from tumor regions within three kinds of images (original images and images smoothened by using Gaussian and wavelet filters, respectively), including 3D shape-based features, gray level cooccurrence matrix, gray level run length matrix, gray level size zone matrix, and gray level dependence matrix. Then, fivefold cross-validation was implemented in the training set to select the best hyper-parameters. Univariate logistic regression analysis was applied for feature selection, which can be used to analyze the correlation between each feature and MaVI, and p-values of all features were obtained for MaVI. Only the features with p-value smaller than 0.10 were kept. Next, the retained features were fed into a SVM using the balance strategy to search the optimal hyper-parameters. After fivefold cross-validation, univariate logistic regression analysis was performed, and a SVM with the optimal hyper-parameters was retrained on the training set and tested on the independent testing set for final evaluation.

2) Deep Learning Methods: We compared the proposed method with two frequently used classification networks, including VGG [42] and SE-Resnet50 [29]. For these two compared networks, the extracted liver and liver tumor region with a size of 160 × 160 × 160 was used as input. Focal loss and margin ranking loss were implemented as the training loss. Stochastic gradient descent with momentum was used as the optimizer. Moreover, fivefold cross-validation was performed to turn the hyper-parameters of the compared networks for fair comparison.

Table IX presents the prediction performance of different classification methods. First, the prediction performance of all deep learning methods is better than that of traditional machine learning method on AUPRC, which indicates the advantages of deep learning methods in detecting positive samples. Second, the AUROC of the traditional machine learning method is higher than that of VGG. In this study, small data size and high class imbalance are observed in the MaVI dataset. Compared with traditional machine learning methods, serious overfitting may be observed in a simple network (e.g., VGG). Therefore, higher AUROC is obtained by using a traditional machine learning method than using VGG. Next, better prediction performance is achieved by using SE-Resnet50 than using VGG probably because important and discriminative features can be learnt by the residual connection and SE modules within SE-Resnet50. Finally, our proposed network has superior performance compared with the other three methods, indicating the robust performance of the proposed method on our dataset for MaVI prediction.

V. Discussion

First, the effectiveness of each of the three modules in the proposed method is evaluated in this section. Second, the effectiveness of tumor margin extraction and ResGCN used in the perception module is confirmed. Third, the competitive performance of PKAFnet with other methods is presented. Fourth, the generalization capability of the proposed method is validated using different hospital data as the independent testing set. Fifth, the interpretability of PKAFnet is illustrated. Finally, the limitations of the current study and possible future solutions are analyzed.

As shown in Fig. 2, tumor margin and tumor region are crucial for the prediction of patients with or without MaVI subsequent MaVI during follow-ups. Therefore, a perception module for tumor margin and a segmentation network for tumor region are presented in this study to fully exploit the effective information within these two regions. As listed in Tables VII and IX, higher prediction results are achieved by using the segmentation network alone in the proposed method than using the SE-Resnet50 classification network (0.744 vs. 0.667 for AUROC and 0.603 vs. 0.421 for AUPRC, and 0.708 vs. 0.675 for BACC), which may contribute to the tumor internal heterogeneity exploited by using the segmentation network. Moreover, the prediction performance of combining the perception module and segmentation network is better than that of using the segmentation network alone (Table VII), indicating that combining information extracted from tumor region and tumor margin can boost the prediction performance. However, the performance of combining perception and MdaFF modules (the segmentation network is ignored) is better than that of directly concatenating the deep features extracted from the perception module and segmentation network (the MdaFF module is discarded). This result indicates that information redundancy may be found among deep features extracted from multi-domain, which leads to suboptimal prediction results by simply concatenating multi-domain features. In addition, representative information of multi-domain features can be exploited by using the MdaFF module, which can adaptively explore the synergy and adaptive weights of features. As shown in Table VII, compared with the proposed method with MdaFF, high SEN (1.000 vs. 0.938) and low SPE (0.333 vs. 0.667) were observed in the proposed method without MdaFF module, which suggests that MdaFF module can alleviate overestimation (i.e., tends to estimate all samples as high-risk population) for MaVI prediction. Moreover, delong test [43] was implemented on the AUROC values of the proposed method with and without MdaFF module. The result illustrated that the performance of the proposed method with MdaFF module is significantly better than that of the proposed method without MdaFF module (0.865 vs. 0.761, $p = 0.009$).

The proposed perception module consists of three steps, including ROI extraction, graph construction, and feature exploitation of tumor margin. Two ROI extraction methods (FE1 and FE2) are initially designed to investigate the effectiveness of different ROI extraction methods on MaVI prediction. In FE1 and FE2 methods, ResCNN is applied to explore potential information among the tumor margin ROIs. As mentioned in the Introduction section, rotation invariance is incorporated into the proposed ROI extraction method. Moreover, rotation invariance is desired by the CNN-based network, which can enhance feature extraction [16]. Therefore, better prediction performance is achieved by using FE2 than using FE1, where the proposed ROI extraction method is used in FE2. However, spatial location information is lost in the proposed ROI extraction method, which may decrease the prediction performance of the CNN-based network. Therefore, instead of using ResCNN, ResGCN is applied in the proposed method to exploit important features related to tumor marginal heterogeneity, which can utilize the rotation invariance and spatial connection of image pixels. As shown in Table III, prediction performance is further improved by using the proposed method than using the FE2 method.

The liver tumors in CT scans were manually outlined in this study to ensure segmentation accuracy. Two experiments were conducted to assess the influence of segmentation variability (manual and disturbed segmentations as well as intra- and inter-observer segmentations) on the prediction performance of the proposed method considering that PKAFnet was designed to extract information from tumor margins. Table IV shows that a similar prediction probability was achieved by using the proposed method (all p-values > 0.050) despite some differences among various kinds of manual segmentation masks ($\mathcal{L}_{dice}$ equals to 0.287, 0.145, and 0.169). These results may contribute to the robustness of tumor margin extraction; thus, complete information related to tumor marginal heterogeneity can be retained. Moreover, the size of tumor margins may be another important factor that influences the performance of the perception module. Thus, q was varied to assess its effects on the performance of the proposed method. A large tumor margin (i.e., larger q value) generally leads to additional discriminative information for capturing tumor marginal heterogeneity. However, redundant information may exist within the extracted tumor margins with large q values when tumors are small. Therefore, a moderate q value was more applicable in the proposed method than other q values.

In traditional radiomic methods, hand-crafted features are initially extracted, and then feature selection and classification are performed separately, which is suitable for a specific task with small data size. On the contrary, in CNN-based classification networks, related features are learnt automatically, and feature extraction and classification are performed in an end-to-end manner, which is more acceptable than radiomic method for a task with large data size. However, in the field of medical image analysis, dataset size is often hundreds, which may result in serious overfitting by simply deepening the classification networks (higher AUROC is obtained in radiomic method than that in VGG (Table IX)). Therefore, introducing certain prior knowledge is necessary to guide network for task-related features extraction and prevent overfitting [44]. In the proposed method, tumor marginal and

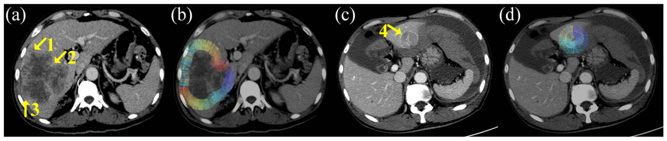

Fig. 7. Interpretability of the perception module on tumor marginal heterogeneity. The patients have no MaVIs at diagnosis. (a) shows patient with subsequent MaVI during follow-ups, whereas (c) displays HCC patient without subsequent MaVI. (b) and (d) show their corresponding Grad-CAM. Corona enhancement, invasive shape, and HCC capsule breakthrough are shown in (a), which are indicated by yellow arrows with 1, 2, and 3, respectively. HCC is indicated by a yellow arrow with 4 in (c), which presents complete capsule and smooth margin.

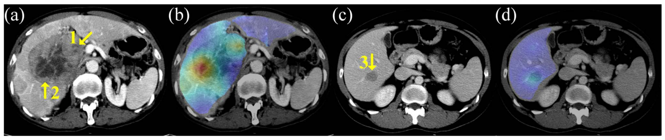

Fig. 8. Interpretability of the segmentation network for tumor internal heterogeneity. The patients have no MaVIs at diagnosis. (a) shows patients with subsequent MaVI during follow-ups, whereas (c) displays HCC patients without subsequent MaVI. (b) and (d) show their corresponding Grad-CAM. Necrosis and mosaic architecture are indicated by yellow arrows with 1 and 2 in (a), respectively. HCC with low heterogeneity is indicated by a yellow arrow with 3 in (c).

internal heterogeneity is included and regarded as clinically prior knowledge, which can guide the whole network to exploit targeted and discriminative features from CT images, thereby achieving good performance of MaVI prediction. Moreover, when different hospital data were used as the independent testing set, a satisfying prediction performance was achieved using the proposed method, which can further prove the good generalization ability of the proposed PKAFnet for MaVI prediction.

The potential clinical interpretation is crucial for computer-aided prediction. Therefore, gradient of class activation map (Grad-CAM) [45], [46] is applied to illustrate the important regions captured by using perception module and segmentation network in PKAFnet. Specifically, the implementation details of Grad-CAM for perception module are as follows. First, given the GCN layer l before the global add pooling layer, the output of layer l can be defined as $m \in \mathbb{R}^{n_r' \times f_t}$, and the gradients between class c and m can be calculated and denoted as $g \in \mathbb{R}^{n_r' \times f_t}$, where n_r' is the number of nodes in layer l, and f_t is the features dimension of nodes in layer l. Then, average operation was implemented on g to obtain the class-specific weights of m, which can be defined as $g^a \in \mathbb{R}^{n_r' \times 1}$. Next, the weights of nodes $\mathcal{V}^l$ in layer l can be calculated by:

$$w^{n_r'} = \mathrm{ReLU}\left(\sum_f g^a \odot m\right) \tag{27}$$

where $\odot$ is the dot product operation and $w^{n_r'} \in \mathbb{R}^{n_r' \times 1}$. Finally, the nodes $\mathcal{V}$ of $\mathcal{G}$ can be mapped by the nodes $\mathcal{V}^l$, where $\mathcal{V}^l$ was reduced by using edge pooling layer [47] on $\mathcal{V}$. Therefore, the weights of nodes $\mathcal{V}^l$ can also be mapped to the weights of nodes $\mathcal{V}$, and the weights of nodes $\mathcal{V}$ are the Grad-CAM of GCN [46].

For HCC patient with subsequent MaVI during follow-ups (Figs. 7 (a) and (b)), the perception module pays attention to the areas (highlight areas) containing corona enhancement, invasive shape, and HCC capsule breakthrough, which has been proven to be related to MaVI in prior knowledge [13]. For HCC patient without subsequent MaVI, complete capsule is observed in Fig. 7 (c), and the color variation of Grad-CAM is smooth (Fig. 7 (d)), which implies that the variation trend of the tumor margin can be detected by using the perception module. Therefore, as shown in Fig. 7, tumor marginal heterogeneity related to MaVI can be well captured by using our proposed perception module based on Grad-CAM (the implementation details of Grad-CAM used in tumor margin are provided in supplementary materials). In evaluating the interpretation of tumor internal heterogeneity by using the segmentation network, Figs. 8 (b) and (d) show the Grad-CAM of the segmentation network on CT images of a HCC patient with subsequent MaVI during follow-ups and a HCC patient without subsequent MaVI. As shown in Figs. 8 (a) and (b), the highlighted colormap covers on the area of necrosis and mosaic architecture, which illustrates that the proposed segmentation network may capture the important features related to prior knowledge of tumor internal heterogeneity. Moreover, uniform density of Grad-CAM can be observed in HCC with slight heterogeneity (Figs. 8 (c) and (d)). Therefore, the effectiveness of PKAFnet can be proven by using Grad-CAMs, and thus the proposed PKAFnet can be used to assist MaVI prediction from CT images.

Although PKAFnet achieves good performance in MaVI prediction, limitations are still identified. First, a multi-center dataset was used in our paper, but the number of samples in the MaVI dataset was still insufficient to fully exploit the potential of deep learning. Therefore, more samples will be collected to improve and evaluate the performance of PKAFnet in future research. Second, the liver and tumor regions are manually outlined in advance, which is time consuming. Therefore, auto-segmentation models for the liver tumor and liver region must be introduced to our proposed method in future work to achieve an auto pipeline for MaVI prediction. In the future, an auto-computer-aided prediction (PKAFnet) for MaVI will show great potential in clinical practical applications and help improve the survival time of HCC patients.

VI. Conclusion

We have presented an accurate and effective framework, namely, PKAFnet, for MaVI prediction. In PKAFnet, the perception module can be used to explore information related to tumor marginal heterogeneity, and segmentation network is implemented to focus on tumor internal heterogeneity. In addition, the information of tumor marginal and internal heterogeneity is fused by using the MdaFF module for decision-making. Prior knowledge of clinical experience for MaVI prediction is introduced to PKAFnet, which can alleviate overfitting and increase the credibility of our proposed model. Consequently, the proposed PKAFnet achieves remarkable performance in MaVI prediction in our multi-center data, implying its potential application for clinical practices.

References

[1] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, A Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.

[2] J. Tian *et al.*, "Deep learning-based aggressive progression prediction from ct images of hepatocellular carcinoma," *Proc. SPIE*, vol. 11597, Mar. 2021, Art. no. 115972Y.

[3] R. S. Finn *et al.*, "Atezolizumab plus bevacizumab in unresectable hepatocellular carcinoma," *New England J. Med.*, vol. 382, no. 20, pp. 1894–1905, 2020.

[4] Z. J. Brown, T. F. Greten, and B. Heinrich, "Adjuvant treatment of hepatocellular carcinoma: Prospect of immunotherapy," *Hepatology*, vol. 70, no. 4, pp. 1437–1442, Oct. 2019.

[5] E. K. Paulson, "Evaluation of the liver for metastatic disease," *Seminars Liver Disease*, vol. 21, no. 2, pp. 225–236, 2001.

[6] J. Wei *et al.*, "Development and validation of a radiomics-based method for macrovascular invasion prediction in hepatocellular carcinoma with prognostic implication," *Proc. SPIE*, vol. 10950, Mar. 2019, Art. no. 109501N.

[7] J. Wei, J. Tian, S. Fu, and L. Lu, "Noninvasive prediction of future macrovascular invasion occurrence in hepatocellular carcinoma based on quantitative imaging analysis: A multi-center study," *J. Clin. Oncol.*, vol. 37, no. 15, p. e14623, May 2019.

[8] Q. Zhang *et al.*, "Differentiation of recurrence from radiation necrosis in gliomas based on the radiomics of combinational features and multimodality MRI images," *Comput. Math. Methods Med.*, vol. 2019, pp. 1–12, Dec. 2019.

[9] Y. Zhou *et al.*, "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101918.

[10] L. Liu, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, "Multi-task deep model with margin ranking loss for lung nodule analysis," *IEEE Trans. Med. Imag.*, vol. 39, no. 3, pp. 718–728, Mar. 2020.

[11] P. Afshar, K. N. Plataniotis, and A. Mohammadi, "Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1368–1372.

[12] D. Castelvecchi, "Can we open the black box of AI?" *Nature News*, vol. 538, no. 7623, p. 20, 2016.

[13] S. Fu *et al.*, "Deep learning-based prediction of future extrahepatic metastasis and macrovascular invasion in hepatocellular carcinoma," *J. Hepatocellular Carcinoma*, vol. 8, pp. 1065–1076, Sep. 2021.

[14] M. Zhou, L. O. Hall, D. B. Goldgof, R. J. Gillies, and R. A. Gatenby, "Exploring brain tumor heterogeneity for survival time prediction," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 580–585.

[15] V. Andrearczyk, J. Fageot, V. Oreiller, X. Montet, and A. Depeursinge, "Local rotation invariance in 3D CNNs," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101756.

[16] M. W. Lafarge, E. J. Bekkers, J. P. W. Pluim, R. Duits, and M. Veta, "Roto-translation equivariant convolutional networks: Application to histopathology image analysis," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101849.

[17] M. Ebrahim, M. Alsmirat, and M. Al-Ayyoub, "Performance study of augmentation techniques for HEp2 CNN classification," in *Proc. 9th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2018, pp. 163–168.

[18] J. Zhu, Y. Li, Y. Hu, K. Ma, S. K. Zhou, and Y. Zheng, "Rubik's Cube+: A self-supervised feature learning framework for 3D medical image analysis," *Med. Image Anal.*, vol. 64, Aug. 2020, Art. no. 101746.

[19] N. Anh Mac and H. Son Nguyen, "Rotation variance in graph convolutional networks," in *Proc. 16th Conf. Comput. Sci. Intell. Syst. (FedCSIS)*, Sep. 2021, pp. 81–90.

[20] M. Huang, W. Yang, M. Yu, Z. Lu, Q. Feng, and W. Chen, "Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced MRI images," *Comput. Math. Methods Med.*, vol. 2012, pp. 1–17, Oct. 2012.

[21] N. Majumder, D. Hazarika, A. Gelbukh, E. Cambria, and S. Poria, "Multimodal sentiment analysis using hierarchical fusion with context modeling," *Knowl.-Based Syst.*, vol. 161, pp. 124–133, Dec. 2018.

[22] R. J. Chen *et al.*, "Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis," *IEEE Trans. Med. Imag.*, vol. 41, no. 4, pp. 757–770, Apr. 2022.

[23] W. Zhu, L. Sun, J. Huang, L. Han, and D. Zhang, "Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2354–2366, Sep. 2021.

[24] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[25] N. Parmar *et al.*, "Image transformer," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4055–4064.

[26] F. Ciompi *et al.*, "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Med. Image Anal.*, vol. 26, no. 1, pp. 195–202, Dec. 2015.

[27] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[29] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze- and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[31] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "NNU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, p. 203, 2021.

[32] J.-Y. Choi, J.-M. Lee, and C. B. Sirlin, "CT and MR imaging diagnosis and staging of hepatocellular carcinoma: Part II. Extracellular agents, hepatobiliary agents, and ancillary imaging features," *Radiology*, vol. 273, no. 1, pp. 30–50, Oct. 2014.

[33] R. Lencioni and J. Llovet, "Modified RECIST (mRECIST) assessment for hepatocellular carcinoma," *Seminars Liver Disease*, vol. 30, no. 1, pp. 052–060, Feb. 2010.

[34] S. Cheng *et al.*, "Chinese expert consensus on multidisciplinary diagnosis and treatment of hepatocellular carcinoma with portal vein tumor thrombus (2018 edition)," *Liver Cancer*, vol. 9, no. 1, pp. 28–40, 2020.

[35] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0118432.

[36] J. Lin, S. Pan, C. S. Lee, and S. Oviatt, "An explainable deep fusion network for affect recognition using physiological signals," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 2069–2072.

[37] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216–231, Oct. 2019.

[38] R. Fluss, D. Faraggi, and B. Reiser, "Estimation of the youden index and its associated cutoff point," *Biometrical J., J. Math. Methods Biosci.*, vol. 47, no. 4, pp. 458–472, 2005.

[39] P. C. O'Brien and T. R. Fleming, "A paired prentice-wilcoxon test for censored paired data," *Biometrics*, vol. 43, no. 1, pp. 169–180, Mar. 1987.

[40] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.

[41] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[43] O. V. Demler, M. J. Pencina, and R. B. D'Agostino, "Misuse of DeLong test to compare AUCs for nested models," *Statist. Med.*, vol. 31, no. 23, pp. 2577–2587, Oct. 2012.

[44] C. Chen, Y. Wang, J. Niu, X. Liu, Q. Li, and X. Gong, "Domain knowledge powered deep learning for breast cancer diagnosis based on contrast-enhanced ultrasound videos," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2439–2451, Sep. 2021.

[45] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[46] P. E. Pope, S. Kolouri, M. Rostami, C. E. Martin, and H. Hoffmann, "Explainability methods for graph convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10772–10781.

[47] F. Diehl, "Edge contraction pooling for graph neural networks," 2019, *arXiv:1905.10990*.