Multi-Focus Network to Decode Imaging Phenotype for Overall Survival Prediction of Gastric Cancer Patients

Liwen Zhang, Di Dong, Lianzhen Zhong, Cong Li, Chaoen Hu, Xin Yang, Zaiyi Liu, Rongpin Wang, Junlin Zhou, and Jie Tian, *Fellow, IEEE*

***Abstract*—Gastric cancer (GC) is the third leading cause of cancer-associated deaths globally. Accurate risk prediction of the overall survival (OS) for GC patients shows significant prognostic value, which helps identify and classify patients into different risk groups to benefit from personalized treatment. Many methods based on machine learning algorithms have been widely explored to predict the risk of OS. However, the accuracy of risk prediction has been limited and remains a challenge with existing methods. Few studies have proposed a framework and pay attention to the low-level and high-level features separately for the risk prediction of OS based on computed tomography images of GC patients. To achieve high accuracy, we propose a multi-focus fusion convolutional neural network. The network focuses on low-level and high-level features, where a subnet to focus on lower-level features and the other enhanced subnet with lateral connection to focus on higher-level semantic features. Three independent datasets of 640 GC patients are used to assess our method. Our proposed network is evaluated by metrics of the concordance index and hazard ratio. Our network outperforms state-of-the-art methods with the highest concordance index and hazard ratio in independent validation and test sets. Our results prove that our architecture can unify the separate low-level and high-level features into a single framework, and can be a powerful method for accurate risk prediction of OS.**

***Index Terms*—Overall survival, gastric cancer, multi-level, CT image, deep learning.**

I. Introduction

GASTRIC cancer (GC) is the third leading cause of cancer-associated deaths globally [1]. The tumor-node-metastasis (TNM) staging manual of GC formulated by the American Joint Committee on Cancer (AJCC) is widely-used for prognostic evaluation, which serves as the guidelines to recommend effective treatments to GC patients (e.g., adjuvant chemotherapy, surgical resection) [2]. However, TNM staging is obtained by pathological biopsy, which is invasive and may cause inaccurate result on account of biopsy sampling and subjective judgment. Currently, the rates of 5-year survival still remain poor and the surgical morbidity is high [3]. Therefore, the AJCC Personalized Medicine Core committee has realized that it is crucial to construct image-based risk models to provide better individualized treatment in combination with the TNM staging manual [4].

Recently, emerging field called radiomics has been widely-used for diagnosis and prognosis by extracting amounts of hand-crafted features [5], [6]. Previous studies have demonstrated that an emerging field of radiomics based on hand-engineered features has significant prognostic value [7], which are related to overall survival and can classify all patients into different risk groups. Several studies indicate that hand-crafted features could decode computed tomography (CT) phenotypes in the detection and evaluation of the node stage for GC patients [8], [9]. Previous study shows that radiomics method can provide powerful predictor for overall survival (OS) of GC patients [7]. However, the hand-crafted features are not tailored, but predefined with

Manuscript received August 21, 2020; revised February 24, 2021 and April 13, 2021; accepted June 5, 2021. Date of publication June 8, 2021; date of current version October 5, 2021. This work was supported in part by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant XDB38040200, in part by the National Key R&D Program of China under Grants 2017YFA0205200, 2017YFC1308700, 2017YFC1309100, and 2017YFA0700401, in part by the National Natural Science Foundation of China under Grants 91959130, 81971776, 81771924, 81930053, 81227901, 81671851, 81527805, 82022036, and 62027901, in part by Beijing Natural Science Foundation under Grant L182061, in part by the Project of High-Level Talents Team Introduction in Zhuhai City Zhuhai under Grant HLHPTP201703, and in part by Youth Innovation Promotion Association CAS under Grant 2017175. Liwen Zhang, Di Dong, and Lianzhen Zhong contributed equally to this work. *(Corresponding authors: Z. Liu; R. Wang; J. Zhou, and J. Tian.)*

Liwen Zhang, Di Dong, Lianzhen Zhong, and Cong Li are with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China and the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: zhangliwen2015@ia.ac.cn; di.dong@ia.ac.cn; zhonglianzhen2018@ia.ac.cn; licong2018@ia.ac.cn).

Chaoen Hu and Xin Yang are with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: huchaoen2014@ia.ac.cn; xin.yang@ia.ac.cn).

Zaiyi Liu is with the Department of Radiology, Guangdong General Hospital, Guangzhou 510080, China (e-mail: zyliu@163.com).

Rongpin Wang is with the Department of Radiology, Guizhou Provincial People's Hospital, Guiyang 550002, China (e-mail: wangrongpin@126.com).

Junlin Zhou is with the Department of Radiology, Lanzhou University Second Hospital, Lanzhou 730030, China (e-mail: ery_zhoujl@lzu.edu.cn).

Jie Tian is with the Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing 100191, China, and the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: tian@ieee.org).

This article has supplementary downloadable material available at https://doi.org/10.1109/JBHI.2021.3087634, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2021.3087634

a limited quantity [10]. In addition, some of these radiomics methods for OS, based on the Cox Proportional Hazard (CPH) method, can be applied only if strong assumptions are first made [11]. The mentioned limitation has led to a transformation from the work of hand-crafted features to self-learning features for survival prediction.

Currently, high-capacity convolutional neural networks (CNNs) show remarkable applications in the domain of computer vision [12]–[17], and excellent performance in medical image-based analysis, such as lung cancer [18], and glioblastoma [19]. Prior work shows that CNN can predict survival using colorectal cancer histological images [20]. A previous study employed a shallow CNN to predict prognosis in multi-institutional CT image datasets [21]. The results show that low-level feature maps represented high spatial information [22]. Li *et al.* show that low-level features within CNN could be applied to exploit the intrinsic textural difference for region detection [23]. In our previous study, we have found that high-level features extracted by the residual network (ResNet), a single-level architecture, are capable of predicting the risk score of OS for GC patients [24]. However, our previous work only has demonstrated the availability and superiority of residual network for survival prediction compared with existing methods. Recent studies focus on a feature pyramid network (FPN) [25], [26], which exploits enhanced high-level features for challenging computer vision tasks [27]. Jiang *et al.* proposes an architecture (S-net) to extract high-level features for predicting survival risk and demonstrates that a deep learning method can improve prognostic prediction [28].

However, few studies have proposed a framework focusing on low-level and high-level features to boost performance for OS risk prediction of GC patients with CT images. Therefore, to achieve high performance, we propose a multi-focus and multi-level fusion feature pyramid network (MMF-FPN), which is utilized to unify separate lower level features and fused high-level features [25].

The main contributive purposes of this work are summarized as follows: 1) we propose a multi-focus architecture that consist of two mono-focus fusion subnets to exploit rich information by decoding the phenotype of tumor for OS risk prediction models. 2) We design a new strategy of cascade connection to extract single and fused lower-level features maps in shadow bottom-top pathway. 3) We collect multicenter survival datasets with CT images. 4) The experimental results indicate that our well-design architecture outperforms the competing model such as the baseline of clinical model, radiomics model, and other state-of-the-art deep learning models.

II. Related Work on Survival Analysis

A. Cox Regression and Log-Risk Function

The CPH regression model, also referred to the Cox regression method [29], is widely used in survival analysis. The survival function signifies the probability that each patient has survived beyond time t, which is defined as

$$\mathrm{Surv}(t) = \mathrm{Pro}(T > t) = \int_t^\infty p(T)dT, \tag{1}$$

where Pro denotes the probability, and T denotes the survival time for each patient. $p(T)$ is hazard function. For the Cox regression model, hazard function $\lambda(t)$ is defined as

$$\lambda(t) = \lim_{\delta \to 0} \frac{\mathrm{Pro}(t \leq T < t + \delta \mid T \geq t)}{\delta}, \tag{2}$$

where the hazard function $\lambda(t)$ represents the probability of death for an individual who has already survived up to time t and survives the incremental amount of time δ. A greater value of $\lambda(t)$ denotes a higher risk of death. For the survival analysis, suppose that each patient has the features of $x = (x_1, \ldots.., x_n)$, the hazard function is also defined as:

$$\lambda(t, \boldsymbol{x}) = \lambda_0(t)e^{h_\beta(\boldsymbol{x})}, \tag{3}$$

$$h_\beta(\boldsymbol{x}) = \log \frac{\lambda(t, \boldsymbol{x})}{\lambda_0(t)} = \boldsymbol{\beta}^T \boldsymbol{x}, \tag{4}$$

where vector $\boldsymbol{\beta}$ is $n \times 1$ and represents a set of coefficients and $\lambda_0(t)$ represents the base hazard when $\boldsymbol{x} = 0$. $h_\beta(x)$ is the log-risk (log-hazard) function, which is a linear function performed by the CPH regression. A higher value of $h_\beta(x)$ indicates a greater hazard of occurrence of an endpoint event. Note that the log-hazard function can be applied for hand-crafted features and clinical variables.

In our study, we collect 640 patients in three centers. During the follow-up period, censored patients' current statuses are unknown due to the loss of track. We record the patients' time from the date of operation until the date of the final follow-up as overall survival time. Uncensored events represent that the patients who are observed for the cancer-related death. We record the patients' time from the date of operation until the date of tumor-related death. Totally, 340 patients are censored. The ratios of censored observations are 55% in training set, 49% in validation set, and 56% in test set.

B. Problem Definition for Deep Learning

Our aim is to model the distribution of the log-risk function $h_\beta(x)$ based on CT image information. Survival analysis can be regarded as a regression problem by ranking all the patients. For each patient, the time-to-event model is a "decoding" function, which learns the patterns within the tumor and predicts the survival risk over the log-risk space as $h_\beta(x)$. To easily describe and understand the survival risk concerning the current condition for each patient, we call the predicted value of $h_\beta(x)$ by each method as the risk score $\hat{h}_\beta(x)$.

C. Loss Function

For our proposed network, whereas MMF-FPN is agnostic to the different loss function, we employ the negative log partial likelihood as our loss function to enable a controlled comparison with previous studies(e.g., [24], [28]). We train all the models by minimizing the loss function for optimal estimation of parameters β:

$$L(\beta) = -\frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_\beta\left(x_i\right) - \log \sum_{j \in A(T_j)} e^{\hat{h}_\beta(x_j)}\right). \tag{5}$$

In the equation, the number of patients is N, and each patient is with an uncensored status. Note that $A(T_j)$ is a set of patients still at risk of failure (death) at time t. For a patient with event time T_j regardless of its status, if $T_j > \mathrm{t}$, the patient will be in the set $A(T_j)$ to train the model. Therefore, any censored patient belonging to $A(T_j)$ will be included in the evaluation of the loss function. $\hat{h}_\beta(x)$ is the output of our proposed network.

D. Predictive Accuracy Metrics

1. *Concordance index*

Each model is evaluated using Harrell's concordance index (c-index), which is a widely-used indicator for performance evaluation [30]. The c-index is similar as the indicator of area under the receiver operating characteristic curve (AUC) to time-to-event survival data. The formula of c-index is defined as follows:

$$C = \frac{\sum_{i,j} \lambda_i \times \mathbf{1}\left(\hat{h}_\beta(x_i) < \hat{h}_\beta(x_j)\right) \times \mathbf{1}(T_i < T_j)}{\sum_{i,j} \lambda_i \times \mathbf{1}(T_i < T_j)} \tag{6}$$

In the formula, to evaluate a model in a dataset, the $\hat{h}_\beta(x_i)$ and $\hat{h}_\beta(x_j)$ represent the predicted risk scores for patient i and j in a pair. The T_i and T_j denote the survival time for patient i and j in a pair. The function of $\mathbf{1}(\hat{h}_\beta(x_i) < \hat{h}_\beta(x_j))$ is 1 if the condition of $\hat{h}_\beta(x_i) < \hat{h}_\beta(x_j)$ is true, and 0 otherwise. The function of $\mathbf{1}(T_i < T_j)$ is 1 if T_i is less than T_j, and 0 otherwise. The numerator of the equation counts the number of patient pairs (i, j) where the pair members with greater predicted risk scores have shorter survival time, denoting correspondence between the predicted risk scores and ground-truth survival outcomes. Production by λ_i denotes the requirement for the sum to subject pairs where they are possible to determine who died first (that is, informative pairs). Therefore, the C (c-index) denotes the fraction of informative pairs exhibiting concordance between predictions and outcomes. The c-index of 0.5 denotes that the risk score is no better than a coin-flip for risk prediction of OS. The c-index of 1 denotes that the risk prediction is perfect in determining which patient has a better prognosis [31].

2. *Hazard ratio*

For clinical evaluation, the hazard ratio (HR) is a widely-used indicator to evaluate prognostic value for the method to classify patients into different risk groups [32]. We employ the HR to evaluate the prognostic value of each method. The cutoff (median risk score) is obtained in the training set. We classify patients with risk scores lower than the cutoff into the low-risk group, and high-risk group otherwise. Assume that the outputted risk score is a risk factor in low-risk and high-risk groups. The value of HR represents the ratio of risk functions between the high-risk and low-risk groups. Patients in the high-risk groups are HR times the risk of morbidity of patients in the low-risk group.

In our study, the statistical significance test is performed with R software (http://www.R-project.org). Clinical variables and hand-crafted features are compared using the Mann-Whitney U test. Prognostic difference between different risk groups are compared by the Log-Rank test. Moreover, we employ the G-rho rank test for calculation of the HR [32]. We also compare the C-indexes of our proposed method and other methods by the Student's t-test. The result is considered statistically significant when the P-value (two-sided test) is less than 0.05.

TABLE I
CLINICAL DATA FOR THREE INDEPENDENT DATASETS

Variables	Training (Center 1)	Validation (Center 2)	Test (Center 3)
Total	337	181	122
Age(years)	55±9	59±12	58±12
Follow-up (Month)	30±19	28±15	53±27
Event			
Censored	184	88	68
Uncensored	153	93	54

[01] Continuous and numerical variables are (mean ± std). During the follow-up period, censored patients' current statuses are unknown due to the loss of track. We record the patients' time from the date of operation until the date of the final follow-up as overall survival time. Uncensored events represent that the patients who are observed for the cancer-related death. We record the patients' time from the date of operation until the date of tumor-related death.

III. MATERIALS AND METHODS

A. Multicenter Survival Datasets

Ethical approval was respectively received for the Institutional Review Board of each center, and informed consent from patients was waived. Survival data consist of four parts for each patient $i(x_i, T_i, E_i, I_i)$: a patient's clinical variables x, an observed event time T, a status of event indicator E and CT images $\boldsymbol{I}$. During the patient's follow-up, if a patient is observed (uncensored) with cancer-related death, we define the indicator of E as 1, otherwise patient is lost to follow up and the E is 0. Data from a total of 640 GC patients are collected from three independent centers: 1) Lanzhou University Second Hospital (337 cases), 2) Guizhou Provincial People's Hospital (181 cases), and 3) Guangdong General Hospital (122 cases). We uniform the recruitment criteria for three centers to ensure consistency (Fig. S1). Characteristics and clinicopathological variables in the training, validation, and test sets are shown in Table I and Table SI. Further details about the survival data are provided in the supplementary (Section S1). The time of OS is calculated from the date of operation until the date of tumor-related death or the date of the final follow-up. We follow up all the patients from January 2013 to March 2019, September 2012 to October 2017, and June 2010 to April 2019 in center1, center 2, and center 3, respectively.

B. Tumor Segmentation and Preprocessing

We employ the software ITK-SNAP for segmentation [33]. First, we identify the largest CT image slice in portal venous phase which is the best phase for each patient by two experienced radiologists in each center and outlined with a bounding box. The same operation is applied to draw the nearest upper and lower slices identified with tumor region. All the tumors are covered by the bounding boxes. For impartial comparison, the input image size is compatible with the ResNet (required inputs size: 224 × 224). Fig. 1(a) is a schematic diagram, which shows

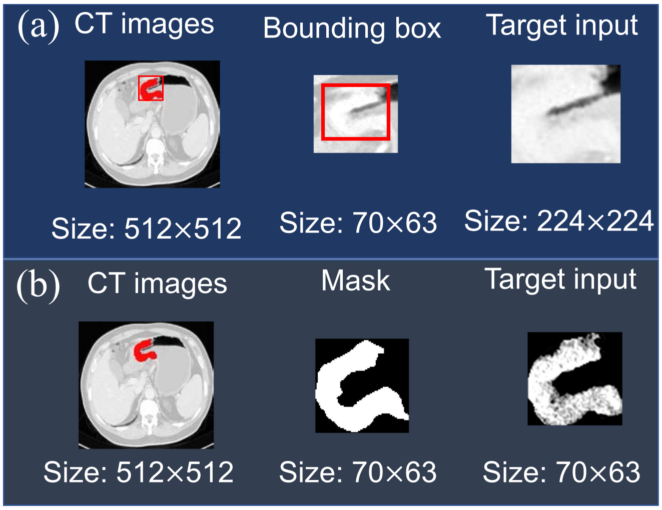

Fig. 1. Image segmentation and preprocessing. (a) An example for deep learning methods. (b) An example for radiomics method.

that the segmented CT images are resized as the target size of 224 × 224. The details for pre-processing methods are described in the supplementary (Section S2). To construct the radiomics model, we once more precisely manually delineate the tumor region of the largest slice for each patient (Fig. 1(b)).

C. Architecture of Our Proposed Network

Our proposed network consists of four core components: (1) a tailored network consisting of a pyramid framework and four residual blocks as the backbone; (2) focusing on lower-level layer-wise feature maps with a pyramid bottom-to-top pathway; (3) focusing on higher-level semantic feature maps with pyramid coarse-to-fine resolution pathway and layer-wise lateral connections; (4) concatenation. We elaborate the details of the core architectures as follows. The backbone of our network is tailored as 10 layers based on four self-defined residual blocks, and we utilize the output of each stage in the backbone (Fig. 2). The outputted feature maps for conv1, conv2, conv3, conv4, and conv5 are represented as {C1, C2, C3, C4, C5}, and their respective pyramid feature maps sizes are {112 × 112, 56 × 56, 28 × 28, 14 × 14, 7 × 7}.

Focusing on the fused lower-level features is to learn representative features from shallow layers. Some lower-level semantic features are equivalent to the basic hand-crafted features, which are related to the prognostic information of OS (e.g., T stage and N stage [8], [9]). As shown in Fig. 2, this subnet unifies the multi-level features maps obtained from different scale receptive field {C1, C2, C3, C4, C5}. The architecture can ensure that each module contains different numbers and combinations of multi-level feature maps in different stages of convolutional blocks. The fused features are conducive to capturing rich information of the tumor phenotype.

The pyramid coarse-to-fine resolution pathway generates feature maps of finer resolution by upsampling the feature maps of lower resolution from the top of the pyramid. The upsampled feature maps are then reinforced with lower-level feature maps from the bottom-to-top pathway via lateral addition. Fig. 2 exposes the multi-level pipelines that build our coarse-to-fine resolution feature maps. In the subnet of focusing on higher-level features, outputted feature maps are upsampled by a factor of 2. The aim of this coarse-to-fine pathway is to generate the finest feature maps after four iterations. The description for each generated feature map is referred to {P1, P2, P3, P4, P5}. One of our design principles is to ensure simplicity and efficiency. We have experimented with more sophisticated frameworks and observed poorer results. We note that designing a better connection approach and sophisticated modules is not the attention of this paper but efficient and suitable architecture for the accurate risk prediction of OS.

The operation of concatenation serves to complement the loss of information brought by fusion and convolution in each layer. In the bottom-to-top pathway, we concatenate all multi-level features to decode the tumor phenotype sufficiently. In the coarse-to-fine pathway, we concatenate all reinforced high-level feature maps and apply a 3 × 3 convolution to eliminate the aliasing distort effect. The concatenation strategy can enforce convolutions to comprehensively collaborate all multi-level semantic features for accurate risk prediction with improved fusional semantic features. We have made our source code of proposed method available at https://github.com/dreamenwalker/Multi-focusNet/.

IV. Experiments and Results

A. Experiments and Existing Methods

The input images are two-dimensional (2D) segmented CT image patches. Taking the cost of data collection and segmentation into consideration (we also train the radiomics model based on hand-crafted features, which requires elaborate delineation by radiologists for each slice of the CT images), we select three available CT slices for each patient. The network architecture is developed for RGB image resulting in a three-channel input. In order to be suitable for the requirement and decode the tumor phenotype entirely, each selected CT image slice is copied twice, and the three slices are stacked as a three-channel image. The average predicted probability is treated as the OS risk probability for each patient.The risk score is the average of the predicted risk values of three slices. We evaluate our proposed network with three independent datasets from three centers. We train our network and the compared methods using 1011 slices of 337 patients from center 1. We select optimal hyper-parameters and weights for each method using 543 images of 181 patients from center 2. We test the trained model using 366 images of 122 patients from center 3. We employ data augmentation to avoid overfitting. For each patient in our datasets, classic augmentation techniques are used including flipping, translation, rotation, crop, sharpen, and linear contrast. More details for data augmentation are illustrated in the supplementary (Section S2).

We evaluate our network MMF-FPN and the existing methods of radiomics [7], FPN [25], S-net [28], residual CNN [24], VGG16 [12], VGG19 [12], Inception [13], DenseNet [14], InceptionResNet [15], NASNetMobile [16], Xception [17], and the clinical model using the c-index, HR, and KM curves as indicators. Further details regarding the construction of the radiomics model are shown in the supplementary (Section

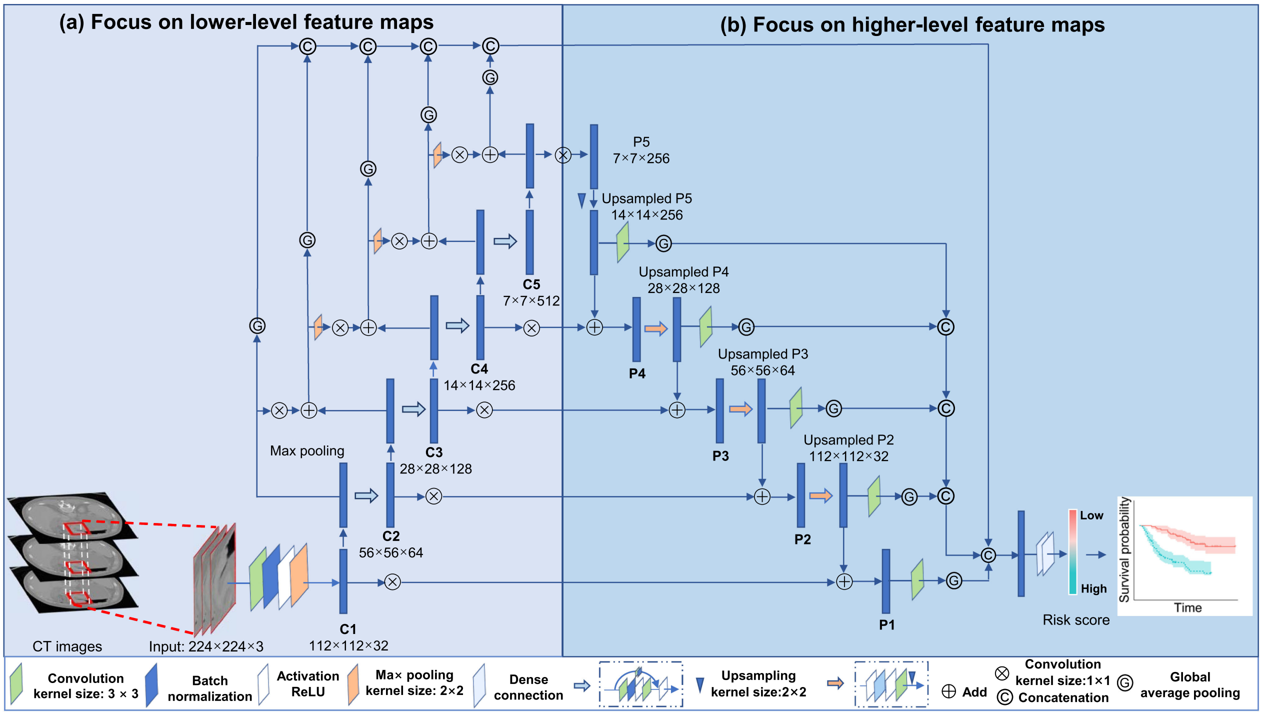

Fig. 2. Architecture of the proposed network. The lower-level and higher-level feature maps are extracted by two mono-focus subnets, respectively. They are fused to generate the multi-level feature maps. (a) The bottom-to-top connection effectively promotes the lower-level alignment of semantic information. (b) The subnet to extract higher-level feature maps. The two mono-focus subnets are further reinforced by fusion and convolution through lateral connections. Finally, the output is adaptively enhanced by the architecture.

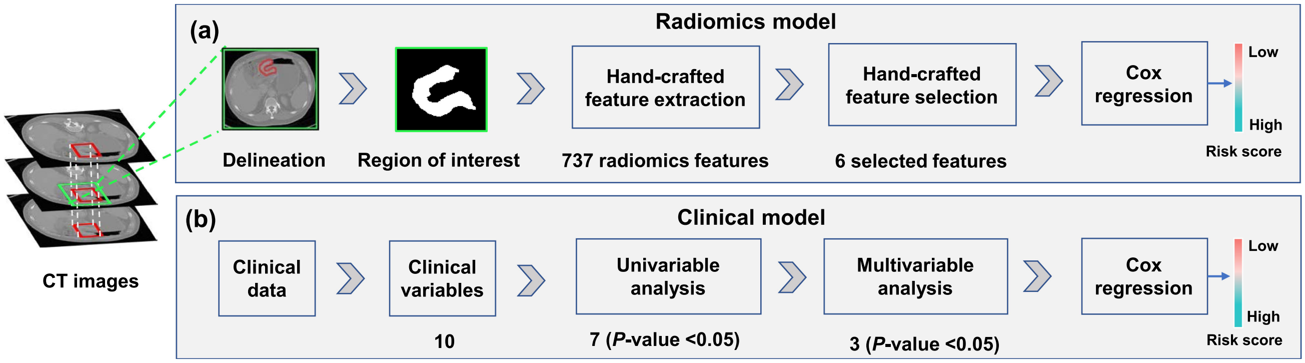

Fig. 3. Flowchart for construction of radiomics and clinical models. (a) radiomics model. To construct a radiomics model, 737 radiomics features are extracted based on the region of interest (ROI) for each patient. Six radiomics features are selected by the LASSO with the 10 folds cross-validation strategy. Radiomics model is constructed by the Cox proportional hazards model with selected six features. (b) clinical model. To construct a clinical model, we collect 10 clinical variables including age, gender, tumor stage, node stage, tumor-node-metastasis (TNM) stage, tumor localization, differentiation, adjuvant chemotherapy, lymphovascular invasion, and tumor size. The clinical model is constructed based on CPH method for univariable and multivariable analysis. The three selected variables of tumor stage, node stage, and adjuvant chemotherapy are used to construct a clinical model according to the P-values (less than 0.05).

S3). The details for the training procedure are provided in the supplementary (Section S4). As shown in Table I, the rate of censored observations in training, validation, and test sets is 55%, 49%, and 56%, respectively. The high rate is a main issue to impact the model performance. To mitigate this problem, we set a large epoch to train deep learning model for convergence. Furthermore, for radiomics method based on hand-crafted features, we set iterations as 1000 with 10-fold cross-validation to train the radiomics model. To demonstrate the incremental improvement of the submodules of multi-focus architectures for OS risk prediction, we perform ablation studies to explore the impact of each mono-focus module. For clear comparisons, we name the architectures that only focus on fused lower-level and higher-level feature maps as FPN+FLL and FPN+FHL, respectively. Fig. 3 describes the pipelines for the construction of radiomics and clinical models. For an impartial comparison of deep learning methods, the image size of input, epoch, batch size, iteration, and the training and validation sets are consistent. Pre-processing and augmentation are applied equally to all image-based methods in the training dataset.

TABLE II
PERFORMANCE EVALUATION FOR EACH METHOD IN THREE DATASETS

Method	P(M)	C-index(95%CI)			Hazard ratio(95%CI)		
		Training	Validation	Test	Training	Validation	Test
VGG16	14.7	0.57 (0.53-0.62)	0.51 (0.45-0.57)	0.5 (0.41-0.59)	1.67 (1.21-2.32)	1.03 (0.69-1.55)	0.79 (0.46-1.35)
VGG19	20.0	0.6 (0.56-0.65)	0.52 (0.46-0.58)	0.59 (0.5-0.67)	1.92 (1.39-2.66)	1.04 (0.67-1.63)	2.01 (1.13-3.58)
DenseNet	7.0	0.62 (0.58-0.67)	0.57 (0.51-0.63)	0.6 (0.52-0.68)	2.2 (1.58-3.06)	1.49 (0.98-2.26)	1.49 (0.82-2.7)
Inception	21.8	0.66 (0.62-0.71)	0.58 (0.52-0.64)	0.58 (0.5-0.65)	2.53 (1.83-3.51)	1.25 (0.83-1.88)	0.9 (0.51-1.57)
InceptionResNet	54.3	0.71 (0.67-0.75)	0.57 (0.5-0.63)	0.68 (0.61-0.76)	3.66 (2.59-5.16)	1.37 (0.91-2.07)	2.11 (1.08-4.09)
NASNet-Mobile	**4.2**	0.57 (0.52-0.61)	0.54 (0.48-0.6)	0.53 (0.45-0.61)	1.37 (0.99-1.88)	1.33 (0.88-2)	1.58 (0.84-2.94)
Xception	20.8	0.66 (0.62-0.7)	0.58 (0.52-0.64)	0.66 (0.59-0.73)	2.76 (1.98-3.87)	1.84 (1.22-2.77)	2.51 (1.13-5.56)
ResNet-18	11.2	0.68 (0.64-0.72)	0.53 (0.47-0.59)	0.68 (0.61-0.76)	2.89 (2.06-4.05)	0.97 (0.63-1.5)	2.44 (1.19-4.99)
ResNet-18+FPN	13.8	0.68 (0.64-0.72)	0.58 (0.52-0.64)	0.71 (0.64-0.78)	3.15 (2.24-4.43)	1.63 (1.08-2.45)	0.71 (0.64-0.78)
ResNet-50	23.5	0.74 (0.71-0.77)	0.51 (0.45-0.57)	0.57 (0.48-0.65)	3.5 (2.47-4.95)	1.39 (0.90-2.14)	1.21 (0.67-2.16)
ResNet-50+FPN	26.9	**0.79** **(0.75-0.82)**	0.59 (0.53-0.65)	0.62 (0.54-0.7)	5.01 (3.53-7.11)	1.76 (1.17-2.66)	1.73 (0.96-3.1)
Clinical	-	0.75 (0.71-0.79)	0.70 (0.64-0.76)	0.68 (0.61-0.76)	4.38 (2.99-6.40)	2.99 (1.66-5.41)	2.71 (1.36-5.38)
Radiomics	-	0.66 (0.62-0.70)	0.54 (0.48-0.60)	0.73 (0.66-0.80)	2.34 (1.68-3.25)	1.35 (0.80-2.29)	7.73 (1.88-31.78)
S-net	27.8	0.65 (0.60-0.69)	0.58 (0.52-0.64)	0.68 (0.61-0.76)	2.25 (1.62-3.12)	1.77 (1.17-2.68)	6.12 (2.21-16.99)
MMF-FPN	5.6	0.77 (0.74-0.81)	**0.74** **(0.69-0.79)**	**0.76** **(0.70-0.82)**	**5.57** **(3.89-7.99)**	**3.50** **(2.27-5.37)**	**9.46** **(2.30-38.91)**

[01] 95%CI: 95% confidence interval; clinical and radiomics represents the model constructed by the Lasso-Cox method using clinical variables and hand-crafted features, respectively; the ResNet-18-FPN and ResNet-50-FPN represent that the models apply the ResNet-18 and ResNet-50 as the backbone in combination with FPN, respectively. The range of the c-index is from 0.5 to 1. The higher c-index denotes better risk prediction of OS. Generally, HR is greater than one for evaluation of survival model. The higher HR is, the better model performance is.

B. Comparisons With Existing Methods

We compare our proposed method with the clinical model, radiomics model, and other methods in Table II. The results indicate that MMF-FPN exhibits robust performance in three sets (training set: c-index: 0.77, 95% confidence interval (CI): 0.74-0.81; Validation set: 0.74, 95%CI: 0.69-0.79; Test set: 0.76, 95%CI: 0.70-0.82). P-values for the comparison between the c-index of our method and the other mentioned methods are significant (all P-values < 0.05) except for the overfitted ResNet-50-FPN method compared in training set (P-value is 0.81). Our method shows the best discrimination capability with the highest c-index in the validation set, for which the comparison of c-index is significant (P-value $< 8.6e\text{-}06$) except for the clinical model (P-value $= 0.11$). The MMF-FPN also outperforms other methods in the external test sets with a significant difference (P-value < 0.05) except for the radiomics model (P-value $=$ 0.1).

C. Assessment With Hazard Ratio and KM Curves

To evaluate the prognostic value for each method (model), all patients in each set are classified into either low-risk or high-risk groups based on the cutoff of the median risk score obtained from the training set. The P-value indicates the difference between the two risk groups (P-value < 0.05 indicates that the two risk groups that have a discrepant prognosis). As is shown in Table II, the clinical indicator of the HR (highest in the training set: 5.57, 95%CI: 3.89-7.99; highest in the validation set: 3.50, 95%CI: 2.27-5.37; highest in the test set: 9.46 95%CI: 2.30-38.91) demonstrates three points: 1) the risk score outputted by our proposed method for each patient is the best signature for risk classification; 2) our method is the most powerful model to dividing people into two risk groups compared with other models; 3) The high-risk group identified by our method has the highest risk of death. Moreover, we also interestingly find that our method shows a higher HR than that of ResNet-50-FPN, despite the lower c-index. The KM curves for major comparisons are plotted for the three datasets respectively in Fig. 4. Our multi-focus network outperforms the other methods and demonstrates its capability about OS to stratify GC patients into low-risk and high-risk groups with discrepant prognosis in three datasets.

D. Impact of Each Mono-Focus Component

To understand which mono-focus subnet is critical for improvement of model performance, we analyze the subnets on the three datasets. Table III shows that two mono-focus subnets improve robustness in the validation and test sets. In the validation set, the risk score predicted only by the backbone (no mono-focus subnet) is 0.62 (0.56-0.68) and 0.67 (0.60-0.75) in test set. The backbone in combination with the mono-focus

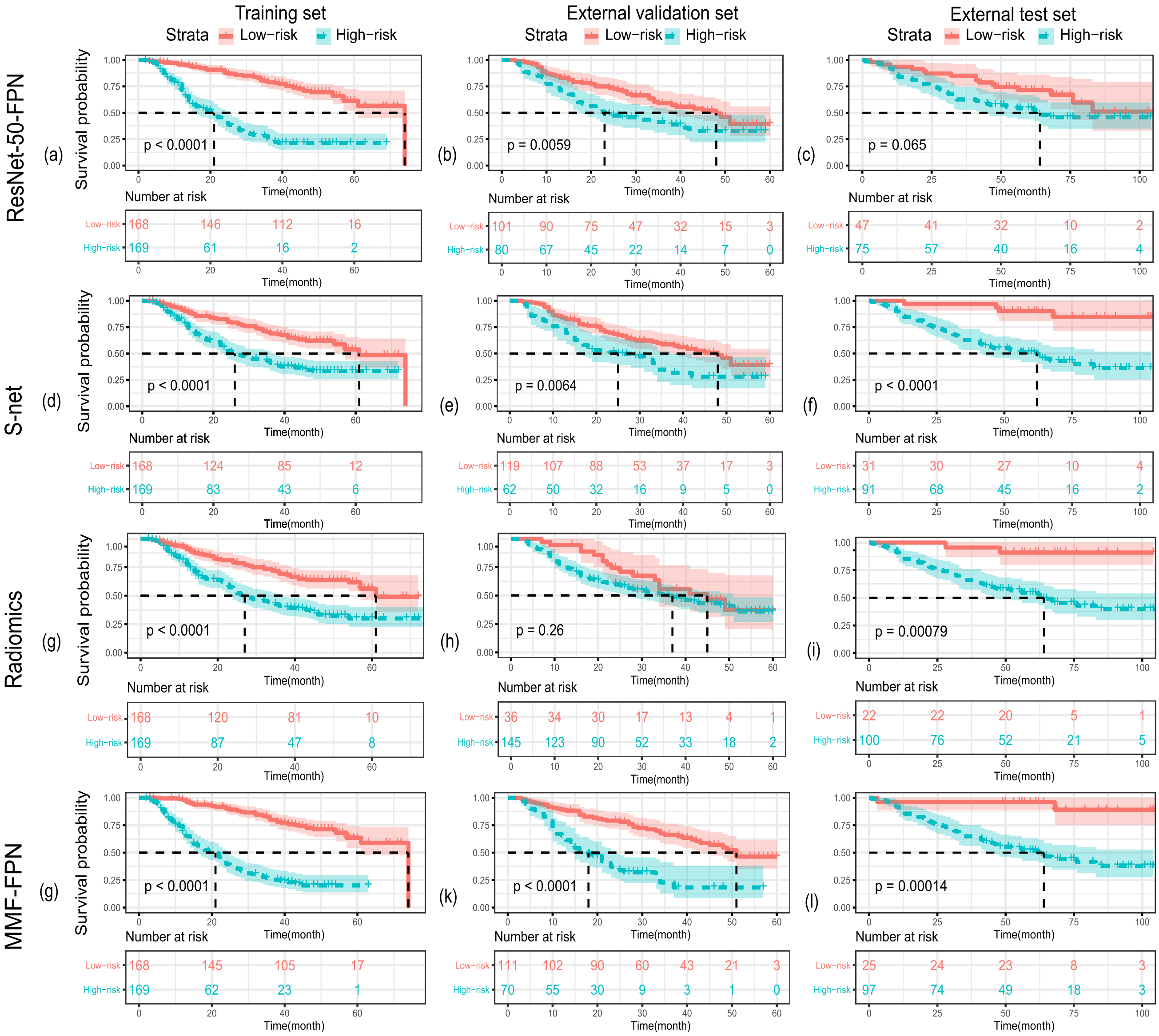

Fig. 4. Evaluation of prognostic value with KM curves for each method. For each survival curve, the dotted line represents the number of patients in the different risk groups who would survive with a survival probability of 0.5. The P-values is calculated by Log-Rank test, which shows the prognostic differences between high-risk and low-risk groups.

subnet (focus on lower-level semantic features) exhibits a higher c-index (0.71, 95%CI:0.67-0.74) and a better prognostic value (HR: 3.61, 95%CI: 2.57-5.08). The other mono-focus subnet (focus on higher-level semantic features) also provides an incremental margin for the c-index (c-index in the validation set: 0.65 vs 0.62, test set: 0.69 vs 0.67) and prognostic value (HR in validation set: 2.08 vs 1.46, test set: 3.1 vs 2.8). The MMF-FPN achieves the highest c-indexes in validation and test sets (validation set: 0.74 (MMF-FPN) vs 0.67 vs 0.65; test set: 0.76 (MMF-FPN) vs 0.72 vs 0.69). The KM curves (Fig. 5) also indicate a significant difference between two risk groups in the validation and test sets, and the significance level in the two sets is improved compared to the backbone (validation set: 0.00 045 (focusing on higher-level) vs 0.00 017 (focusing on lower-level) vs 0.073 (baseline); test set: 0.0012 vs 0.00 052 vs 0.0033).

V. Discussion

In this study, the proposed network, MMF-FPN, outperforms other competing methods, based on the evaluation of the c-index, HR, and KM curves in the validation and test sets. The good performance of MMF-FPN indicates that our architecture unifies the separate low-level and high-level features into a single framework, and reasons global information about the multi-level features to predict survival risk accurately.

Currently, CNNs attract much attention in survival analysis [24], [34]. For the existing CNN methods applied to survival prediction, the architectures may be limited for processing CT images. Because these methods can only extract single high-level features for local detailed information. The conventional CNNs do not fully utilize single and fused multi-level features. Although the S-net exploit the different high-level features for

TABLE III
Ablation Experiments of Performance Evaluation for the Mono-Focus Components

Method	c-index (95%CI)			Hazard ratio(95%CI)		
	Training	Validation	Test	Training	Validation	Test
Backbone	0.70 (0.66-0.74)	0.62 (0.56-0.68)	0.67 (0.60-0.75)	2.67 (1.91-3.73)	1.46 (0.96-2.22)	2.8 (1.37-5.74)
Backbone + FLL	**0.71** **(0.67-0.74)**	**0.67** **(0.62-0.73)**	**0.72** **(0.66-0.78)**	**3.61** **(2.57-5.08)**	**2.21** **(1.45-3.37)**	**3.32** **(1.62-6.81)**
Backbone + FHL	0.69 (0.65-0.73)	0.65 (0.59-0.7)	0.69 (0.62-0.76)	2.51 (1.81-3.49)	2.08 (1.37-3.17)	3.1 (1.51-6.34)

[01] Backbone + FLL and Backbone + FHL represent the model consist of backbone combined with the mono-focus component for lower-level feature maps and the mono-focus component for higher-level feature maps, respectively.

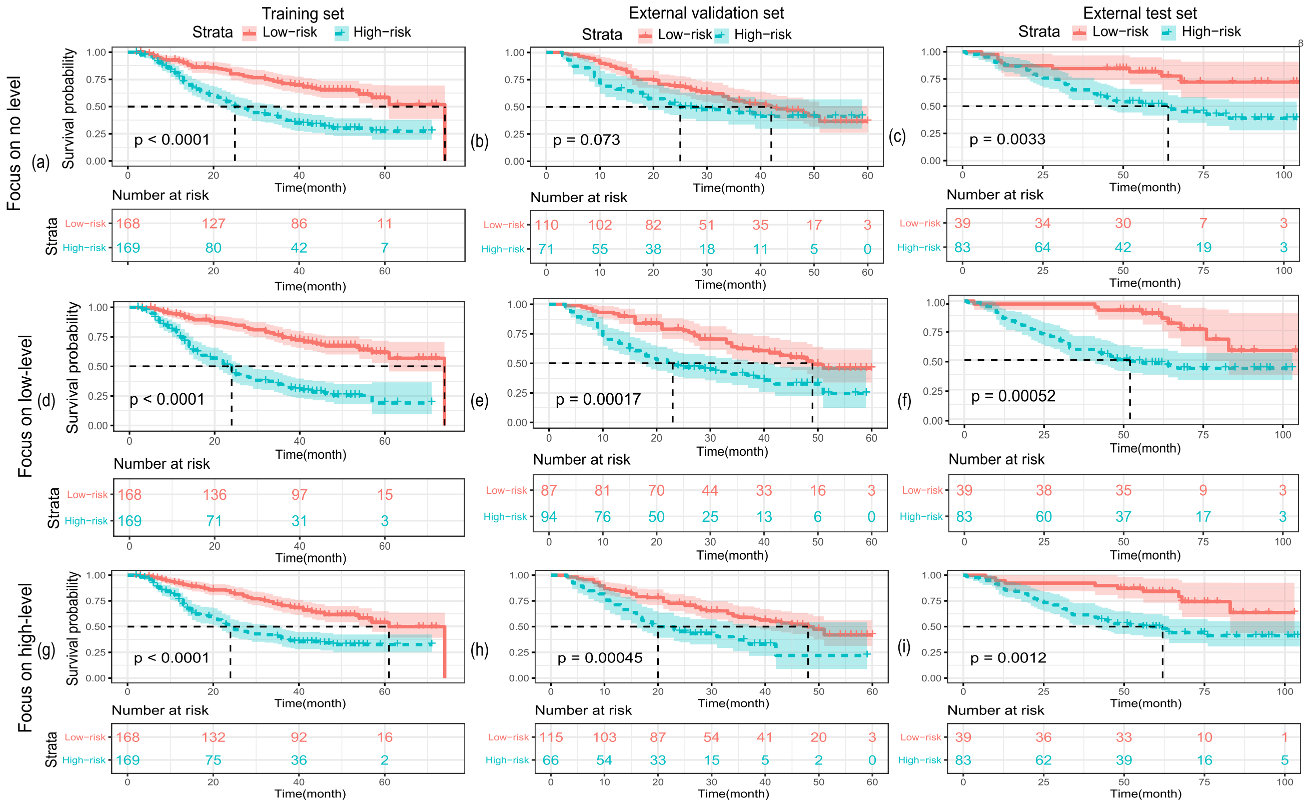

Fig. 5. Evaluation of prognostic value with KM curves for each component.

survival prediction [28], it does not consider the shallow useful information with the continuous operations of pooling and convolution. Although the network proposed in our previous study has better performance, the network is not tailored but modified based on ResNet-18 architecture, and the network is only evaluated in training and validation sets. The network is no independent test set to evaluate the model objectively. In order to further explore whether a tailored deep learning method can establish a stable model for survival prediction, a new network framework is proposed to predict the survival risk. The results of this work indicate that the multi-focus network is robust for prediction of survival risk.

Our ablation studies indicate that the performance for the mono-focus networks in the prognosis analysis follows the order of MMF-FPN > backbone+FLL > backbone+FHL > backbone (table II and III). Our multi-focus network outperforms the standard mono-focus networks and the backbone (FPN), which illustrates that the multi-focus architecture is effective for the improvement of the OS risk prediction. Our network reveals that lower-level features in the shallow layers contain prognostic information to improve the risk prediction of OS. We also observe that the separate lower-level and enhanced higher-level features can boost performance for risk prediction of OS.

Our proposed network can extract more relevant features to decode the tumor phenotype. In clinical practice, some radiomic phenotypic features (e.g., tumor size, tumor shape) are important for radiologists to specify surgical plans [7], [34]. However, these features are predefined, and the number of hand-crafted features is limited, which caused the limited ability to analyze diverse patients. Actually, the hand-crafted features (e.g., wavelet features, texture features) are extracted by a single operation of convolution, which are similar as the feature maps extracted in shallow layers. Therefore, we design the cascaded bottom-to -up subnetwork to extract the low-level features, which can extract some basic information like radiomics features. Focusing on low-level features is compelling due to the easy-to-understand semantic information and the attachment of stationary weights to specific locations.

In recent years, inspired by human perception that focuses on a sequence of several important parts to process a whole scene better, deep learning shows powerful ability to capture the discriminative local details by continuous operations of pooling and convolution [25]. However, the subsampling of the operations causes the loss of global information. For the overall survival prediction, the single high-level deep learning features are limited to extract discriminative features. Compared to the existing CNNs, the cascaded bottom-to-up subnet extracts the low-level features. The second subnet, upsampling architecture, has greater expressive power with fused high-level feature maps than single high-level feature maps in the last few convolutional layers retaining these invariance properties of details. In particular, our upsampling subnet fuses the low-level information separately and avoid the information consumption in deep layers and keeps the feature aggregation to adapt to OS risk prediction.

We note some limitations for our study. Although our datasets are collected from three centers, our multicentric datasets exist unbalanced distribution of clinical data (e.g., gender, survival time), which may cause the impact on our results. Our datasets also have a high rate of censored observations, which is more than fifty percent in training and test sets. The censored patients are the samples with no label, which may cause poor performance of the proposed network and cost long time for training. In our study, we found that the deep learning model is not robust with the small epoch (e.g., epoch = 20–50), and the radiomics model is not converged with small iteration (e.g., iteration = 100). Thus, we set a large epoch and iteration to train the model for convergence. Besides, further work should be done to investigate the impact of censored observations. In our study, considering that the workload is huge for elaborate delineation, we only select three slices for each patient. Further work should be done to train the model with whole tumor slices or more slices. We evaluate the generalization and applicability of our method on lung cancer data from TCIA(Fig. S2) [35], our network can also significantly divide all lung cancer patients into different risk groups. Therefore, whether a CNN can be efficient for different cancer types (e.g., lung cancer, head and neck cancer) simultaneously should be explored.

VI. Conclusion

In conclusion, we propose a multi-focus network, MMF-FPN, to fuse multi-level features for OS risk prediction of GC patients. The MMF-FPN outperforms the radiomics method and existing deep learning networks. Notably, the MMF-FPN can provide OS-related prognostic risk scores and classify GC patients into different risk groups with the highest HR compared with the competing methods. Our results prove that our architecture can unify the separate low-level and high-level features into a single framework, and can be a powerful method for accurate risk prediction of OS.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: A Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018.

[2] M. B. Amin *et al.*, “The eighth edition ajcc cancer staging manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging,” *CA: A Cancer J. Clin.*, vol. 67, no. 2, pp. 93–99, 2017.

[3] J. J. Tegels, M. F. De Maat, K. W. Hulsewé, A. G. Hoofwijk, and J. H. Stoot, “Improving the outcomes in gastric cancer surgery,” *World J. Gastroenterol.: WJG*, vol. 20, no. 38, 2014, Art no. 13692.

[4] M. W. Kattan *et al.*, “American joint committee on cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine,” *CA: A Cancer J. Clin.*, vol. 66, no. 5, pp. 370–374, 2016.

[5] P. Lambin *et al.* “Radiomics: The bridge between medical imaging and personalized medicine,” *Nature Rev. Clin. Oncol.*, vol. 14, no. 12, pp. 749–762, 2017.

[6] D. Dong *et al.*, “Development and validation of a novel mr imaging predictor of response to induction chemotherapy in locoregionally advanced nasopharyngeal cancer: A randomized controlled trial substudy (nct01245959),” *BMC Med.*, vol. 17, no. 1, pp. 1–11, 2019.

[7] W. Li *et al.*, “Prognostic value of computed tomography radiomics features in patients with gastric cancer following curative resection,” *Eur. Radiol.*, vol. 29, no. 6, pp. 3079–3089, 2019.

[8] D. Dong *et al.*, “Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: An international multi-center study,” *Ann. Oncol.*, vol. 31, no. 7, pp. 912–920, 2020.

[9] D. Dong *et al.*, “Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer,” *Ann. Oncol.*, vol. 30, no. 3, pp. 431–438, 2019.

[10] J. J. Van Griethuysen *et al.*, “Computational radiomics system to decode the radiographic phenotype,” *Cancer Res.*, vol. 77, no. 21, pp. e 104–e107, 2017.

[11] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, “Deepsurv: Personalized treatment recommender system using a cox proportional hazards deep neural network,” *BMC Med. Res. Methodol.*, vol. 18, no. 1, pp. 1–12, 2018.

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

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[15] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, pp. 4278–4284, 2017.

[16] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning transferable architectures for scalable image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8697–8710.

[17] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.

[18] S. Wang *et al.*, “Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning,” *Eur. Respir. J.*, vol. 53, no. 3, 2019, Art. no. 1800986.

[19] Z. Tang *et al.*, “Deep learning of imaging phenotype and genotype for predicting overall survival time of glioblastoma patients,” *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2100–2109, Jun. 2020.

[20] J. N. Kather *et al.* “Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study,” *PLoS Med.*, vol. 16, no. 1, 2019, Art no. e 1002730.

[21] P. Mukherjee *et al.*, “A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets,” *Nature Mach. Intell.*, vol. 2, no. 5, pp. 274–282, 2020.

[22] G. Ghiasi and C. C. Fowlkes, “Laplacian pyramid reconstruction and refinement for semantic segmentation,” in *Proc. Eur. Conf. Comput. Vis.*. Cham: Springer, 2016, pp. 519–534.

[23] H. Li, J. Chen, H. Lu, and Z. Chi, “Cnn for saliency detection with low-level feature integration,” *Neurocomputing*, vol. 226, pp. 212–220, 2017.

[24] L. Zhang *et al.*, “A deep learning risk prediction model for overall survival in patients with gastric cancer: A multicenter study,” *Radiother. Oncol.*, vol. 150, pp. 73–80, 2020.

[25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.

[26] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8759–8768.

[27] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis.,*, 2017, pp. 2961–2969.

[28] Y. Jiang *et al.*, “Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: A multicenter, retrospective study,” *Ann. Surg.*, 2020.

[29] D. R. Cox, “Regression models and life-tables,” *J. Roy. Statist. Soc.: Ser. B. (Methodol.)*, vol. 34, no. 2, pp. 187–202, 1972.

[30] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati, “Evaluating the yield of medical tests,” *JAMA*, vol. 247, no. 18, pp. 2543–2546, 1982.

[31] V. C. Raykar, H. Steck, B. Krishnapuram, C. Dehing-Oberije, and P. Lambin, “On ranking in survival analysis: Bounds on the concordance index,” in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 2008, pp. 1209–1216.

[32] M. A. Hernán, “The hazards of hazard ratios,” *Epidemiol.* (Cambridge, Mass.), vol. 21, no. 1, pp. 13–15, 2010.

[33] P. A. Yushkevich *et al.*, “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[34] W. Zhang *et al.*, “Development and validation of a ct-based radiomic nomogram for preoperative prediction of early recurrence in advanced gastric cancer,” *Radiother. Oncol.*, vol. 145, pp. 13–20, 2020.

[35] H. J. Aerts *et al.*, “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach,” *Nature Commun.*, vol. 5, 2014, Art. no. 4006.