Neural Networks 152 (2022) 394–406

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Knowledge-guided multi-task attention network for survival risk prediction using multi-center computed tomography images

Liwen Zhang [a,b,1], Lianzhen Zhong [a,b,1], Cong Li [a], Wenjuan Zhang [c], Chaoen Hu [a], Di Dong [a,b,*], Zaiyi Liu [d,**], Junlin Zhou [c,**], Jie Tian [a,e,f,g,*,2]

[a] CAS Key Laboratory of Molecular Imaging, Beijing Key Laboratory of Molecular Imaging, the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[b] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
[c] Department of Radiology, Lanzhou University Second Hospital, Lanzhou 730030, China
[d] Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510080, China
[e] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing 100191, China
[f] Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University, Xi'an, Shaanxi, 710126, China
[g] Key Laboratory of Big Data-Based Precision Medicine (Beihang University), Ministry of Industry and Information Technology, Beijing, 100191, China

ARTICLE INFO

Article history:
Received 3 November 2021
Received in revised form 2 April 2022
Accepted 22 April 2022
Available online 28 April 2022

Keywords:
Overall survival
Deep learning
Computed tomography (CT)
Neural network

ABSTRACT

Accurate preoperative prediction of overall survival (OS) risk of human cancers based on CT images is greatly significant for personalized treatment. Deep learning methods have been widely explored to improve automated prediction of OS risk. However, the accuracy of OS risk prediction has been limited by prior existing methods. To facilitate capturing survival-related information, we proposed a novel knowledge-guided multi-task network with tailored attention modules for OS risk prediction and prediction of clinical stages simultaneously. The network exploits useful information contained in multiple learning tasks to improve prediction of OS risk. Three multi-center datasets, including two gastric cancer datasets with 459 patients, and a public American lung cancer dataset with 422 patients, are used to evaluate our proposed network. The results show that our proposed network can boost its performance by capturing and sharing information from other predictions of clinical stages. Our method outperforms the state-of-the-art methods with the highest geometrical metric. Furthermore, our method shows better prognostic value with the highest hazard ratio for stratifying patients into high- and low-risk groups. Therefore, our proposed method may be exploited as a potential tool for the improvement of personalized treatment.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

Lung and gastric cancers are respectively the first and third leading causes of cancer-associated mortality worldwide (Bray et al., 2018). Although therapeutic plans (e.g., radiotherapy and adjuvant chemotherapy) for patients are continuously explored (Ajani et al., 2016), 5-year survival rates still remain poor (Hirsch et al., 2017; Tegels, De Maat, Hulsewé, Hoofwijk, & Stoot, 2014). Hence, it is crucial to explore feasible methods (e.g., risk prediction models) to provide personalized treatment for varying prognoses. The American Joint Committee on Cancer (AJCC) Staging Manual has become a guideline for diagnosing cancer patients, determining the schedule of treatment. Particularly, the tumor, lymph node, and metastasis (TNM) staging manual has been widely accepted as a guideline of cancer classification for individualized treatment (Amin & Edge, 2017). The AJCC also points out that overall survival (OS) prediction based on accurate risk models is more significant for personalized treatment than the conventional cancer staging systems (Kattan et al., 2016).

Computed tomography (CT) is a routinely used modality to facilitate survival risk prediction for gastric and lung cancers in clinical practice (Hu, Shen, & Sun, 2018; Liu, Johns, & Davison, 2019; Liu, Wang, Liu, Yang, & Tian, 2021; Woo, Park, Lee, & So Kweon, 2018). CT scans can provide rich information such as the location, shape, and size of lesions, and display the extent of

* Correspondence to: CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.
** Corresponding authors.
E-mail addresses: di.dong@ia.ac.cn (D. Dong), zyliu@163.com (Z. Liu), ery_zhoujl@lzu.edu.cn (J. Zhou), jie.tian@ia.ac.cn (J. Tian).
[1] These authors contributed equally to this work.
[2] FAIMBE, FIAMBE, FIEEE, FSPIE, FOSA, FIAPR.

https://doi.org/10.1016/j.neunet.2022.04.027
0893-6080/© 2022 Elsevier Ltd. All rights reserved.

tumors on morphological manifestations for patients. The AJCC cancer staging manual points out that routinely used modality of CT images can provide primary information for assigning clinical TNM (cTNM) stages, which is invaluable for guiding biopsies and surgical resections for gastric and lung cancers (Amin & Edge, 2017; Huang et al., 2016; Li et al., 2019). However, accurate assessment of CT images is limited to the information obtained by experts' analysis. Recently, an emerging field called radiomics has shown encouraging results in medical image analysis by CT-based machine learning, demonstrating that some hand-crafted features generally contain rich information that is complementary to the radiologists' evaluation (Aerts et al., 2014; Hong, Tomé, & Harari, 2012; Li et al., 2019; Warfield, Zou, & Wells, 2008; Zhang et al., 2018). However, these methods can only extract limited pre-defined features and require elaborate segmentation for tumor regions, which is time-consuming and may result in poor prediction performance.

Recent advances in convolutional neural networks (CNNs) have attracted considerable attention for quantitative medical image analysis applications in prognostic prediction models, and have shown remarkable performance for gastric and lung cancers (Dong et al., 2020; Jiang et al., 2020; Kather et al., 2019; Kim, Yoon, Choi, & Suk, 2019; Lin et al., 2017; Liu, Qi, Qin, Shi, & Jia, 2018; Mukherjee et al., 2020; Tang et al., 2020; Wang et al., 2019; Zhang et al., 2020). Lu et al. proposed a novel neighboring aware graph neural network (NAGNN) based on neighboring aware representation (NAR) for detecting COVID-19 using chest CT scans. The authors demonstrated that NAGNN outperformed state-of-the-art methods in terms of generalization ability (Lu, Zhu, Gorriz, Wang, & Zhang, 2022). Mukherjee et al. designed a shallow CNN (LungNet**)** to improve the accuracy of survival prediction (Mukherjee et al., 2020). A previous study showed that CNNs can also learn discriminative features from histological images for survival prediction tasks in colorectal cancer (Kather et al., 2019). In our previous work, we employed a residual network to extract high-level features and found that they were capable of predicting the risk of overall survival (OS) in patients with gastric cancer (Zhang et al., 2020). Moreover, previous study proposed four novel abnormal brain diagnosing methods based on deep learning for brain MRI, which shows robust performance and high accuracy (Lu, Wang, & Zhang, 2020). This method maybe a potential framework to search the optimal feature layers. Recent studies have also focused on new architecture called feature pyramid networks (FPNs), which include semantically strong features for different computer vision tasks and have shown excellent performance (Lin et al., 2017; Liu et al., 2018). Jiang et al. proposed an architecture (S-net) based on FPN architecture to predict survival risk, and demonstrated that FPN architecture was efficient for the improvement of prognostic prediction (Jiang et al., 2020).

Although all of the above-mentioned studies have shown great application of medical images for OS prediction, they largely fail to explore a knowledge-guided network to combine with experts' evaluations based on CT images (e.g., clinical TNM staging). The key challenge in automated medical image analysis is to learn representative features from scarce data. Currently, a promising subfield known as multi-task learning (MTL) has shown remarkable successes in many deep learning applications, such as computer vision (Dai, He, & Sun, 2016; Dorado-Moreno et al., 2020; Girshick, 2015; He, Gkioxari, Dollár, & Girshick, 2017; Liu et al., 2019), medical image analysis (Tang et al., 2020), and speech recognition (Ruder, 2017).

In recent years, inspired by human perception, which focuses on a sequence of several important parts to better process a whole scene, attention mechanisms have been widely explored to facilitate network model optimization. They bring out encouraging improvements in the performance of different tasks such as detection and segmentation (Hu et al., 2018; Liu et al., 2019, 2021; Pang, Du, Orgun, Wang, & Yu, 2021; Woo et al., 2018). The purpose of attention networks is to reinforce representative features and suppress redundant features. A previous study proposed an attention module of the Squeeze-and-Excitation Network (SENet) to extract discriminative features using a channel attention network (Hu et al., 2018). However, this study did not focus on spatial attention. Woo et al. proposed both channel and spatial attention networks based on pooling and pixel-wise production (Woo et al., 2018), which has been demonstrated as an efficient module in classification and detection performance. Liu et al. proposed a multi-task attention network (MTAN) for vision tasks, and the results showed the efficiency of their proposed network (Liu et al., 2019). Although the backbone provides shared features for each task, the attention module in MTAN was not shared but rather was designed for each task. Furthermore, regarding survival risk prediction of human cancers, most studies have only focused on capturing rich information using CT images for OS prediction.

In clinical practice, the clinical TNM stages are the primary guidelines for treatment, which are implemented with the involvement of multiple radiologists' assessments by evaluating CT images. However, few studies have explored the development of powerful machine learning methods to integrate prior knowledge from human expert labeling for OS prediction. Moreover, few studies have proposed a tailored multi-task CNN architecture to pay attention to extracting rich information from CT imaging data and improving the performance of OS prediction models guided by radiologists' experience.

In this study, as shown in Fig. 1, we exploit clinical TNM stages as knowledge-guide information and proposed a novel multi-task attention pyramid network (KMAP-Net) to improve the performance of OS risk prediction for lung and gastric cancer patients. To achieve this, we designed sibling subnetworks to capture rich information from 2D CT images of multiple scales, both from the region of interest (ROI) and ROI with the peritumoral region. To improve the learning ability of the proposed CNN framework, we designed a novel attention module called a selection and reinforcement attention module (SRAM) to select important features and reinforce them. SRAM comprises of a channel attention module (CAM) and a spatial attention module (SAM). The attention submodules are designed to sequentially focus on information in both the channel and spatial axes, so that each convolutional block can find and emphasize important channels and regions. We designed a multi-task CNN framework to focus on obtaining rich information from CT images guided by cTNM with the experience of radiologists to improve OS risk prediction.

We summarize our main contributions as follows.

1. We propose a knowledge-guided multi-task attention pyramid network (KMAP-Net) to improve the performance of OS prediction methods in lung and gastric cancer patients, where both human expert labeled information for clinical TNM stages and survival times were provided.
2. We propose a learning strategy based on scale-adaptive inputs to capture rich information from 2D CT images of multiple scales. The proposed strategy enables the network to capture related features within the tumor and the immediately surrounding contexts with as little noise as possible.
3. We propose a novel attention module called SRAM to select and reinforce important features in both channel and spatial axes.

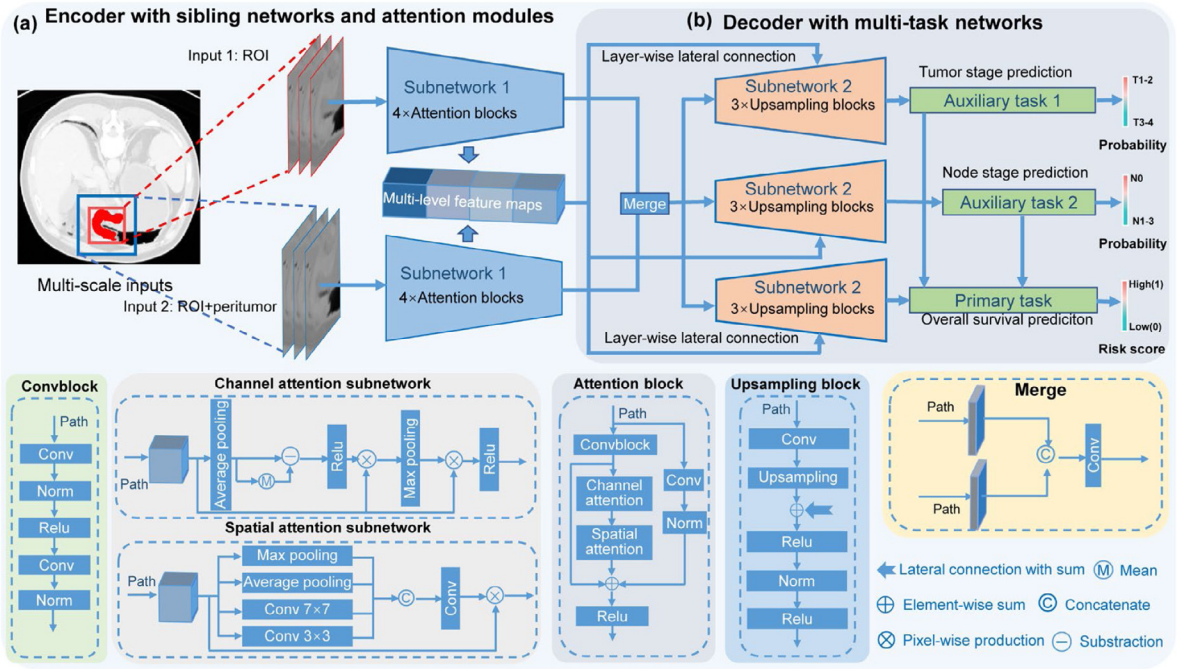

Fig. 1. Overview diagram of our proposed KMAP-Net. (a) Scale-adaptive inputs and sibling encoders. One input includes the ROI colored in red with a red bounding box. The other input is the scale-adaptive ROI with a peritumoral region with a blue bounding box. The different scale inputs are designed to focus on different representative features in the two regions, which may contain abundant information for predicting correlation with tumor prognosis. The peritumoral region is determined by its ROI size. The encoder consists of sibling subnetworks and attention modules. The multi-scale input is designed to focus on different representative features in both an ROI and an ROI with a peritumoral region for each CT slice. Subnetwork 1 in the two branches has the same architecture of four attention modules with channel and spatial attention subnetworks. (b) Decoder architecture and multi-task networks. We concatenate the output feature maps in each attention block in the sibling subnetworks, and multi-level features are also added to the corresponding feature maps in subnetwork 2 with the same size as subnetwork 1. A merging strategy is applied for the output feature maps in the last attention block. Subnetwork 2 consists of several up-sampling blocks. Three prognostic related tasks include OS risk prediction, tumor stage prediction and node stage predictions.

2. Methodology

An overview of the proposed KMAP-Net is presented in Fig. 1. KMAP-Net is designed to employ CT images of different scales as inputs with the same target size ($224 \times 224 \times 3$) to capture representative features for OS prediction and clinical cancer stage prediction. The inputs of the KMAP-Net are 2D images cropped from each slice of CT scans. Our proposed knowledge-guided multi-task network architecture is composed of two sibling encoders and three task-specific decoders. The sibling encoders are equipped with multi-attention modules to perform discriminative feature extraction based on the ROIs. The second set of task-specific decoders performs feature up-sampling and reinforcement.

2.1. Scale-adaptive inputs

We design scale-adaptive inputs with multi-scale cropped CT images to feed them into sibling subnetworks. For each patient, we selected a CT slice of the largest ROI and delineated its boundary precisely. Then, we cropped the tumor region using a rectangular bounding box according to the delineated boundary. To compromise the workload for segmentation and sample size, we also cropped the tumor regions from the nearest upper and lower slices of the selected slice using the same operation for each patient. For impartial comparison, the input image size was set to 224×224. As shown in Fig. 2, the input for one of the subnetworks of the encoder includes the ROI colored in red with a red bounding box for each CT scan. The other input is the scale-adaptive ROI with a peritumoral region rather than a fixed size mentioned in a previous study (Wu et al., 2020). Considering that previous studies have demonstrated that peritumoral regions also contain rich information for prognosis analysis (Pak et al., 2015; Wang et al., 2020), we designed an adaptive scaling strategy for the input of a scale-adaptive ROI with the peritumoral region. Fig. 2 shows the details used to obtain the scale-adaptive input 2. For each patient, each ROI input 1 is defined and its boundary is delineated by experienced radiologists and outlined it with a tumor-centered rectangle box. For the input 2 of ROI and peritumor, its optimal size is confirmed by our proposed learning strategy of scale-adaptive inputs. According to its ROI coordinates (Fig. 2a) defined by experienced radiologists, we can obtain the corresponding expansion increments of d_x and d_y to find the optimal size of the peritumoral area (Fig. 2b). In our study, our controlled experiments show that we obtained the best optimal size for input 2 when a parameter of scaling factor k was set as 0.2 by our scale-adaptive strategy.

2.2. Selection-and-reinforcement attention module

Our proposed selection-and-reinforcement attention module (SRAM) can be an independent component embedded in common architectures such as SENet and convolutional block attention module (CBAM) (Hu et al., 2018; Woo et al., 2018). Given an intermediate input feature map $\mathbf{X} \in \mathbf{R}^{C\times H\times W}$, the output of the channel attention module (CAM) is a tensor of $\mathbf{X_c} \in \mathbf{R}^{C\times H\times W}$ and the output of the spatial attention module (SAM) is a tensor of $\mathbf{X_s} \in \mathbf{R}^{C\times H\times W}$. We use the following equation to show the overall process of operation for the attention mechanism:

$$\mathbf{X}_c = \mathbb{F}_c(\mathbf{X}) \otimes \mathbf{X}, \tag{1}$$

$$\mathbf{X_s} = \mathbb{F}_s(\mathbf{X_c}) \otimes \mathbf{X_c}, \tag{2}$$

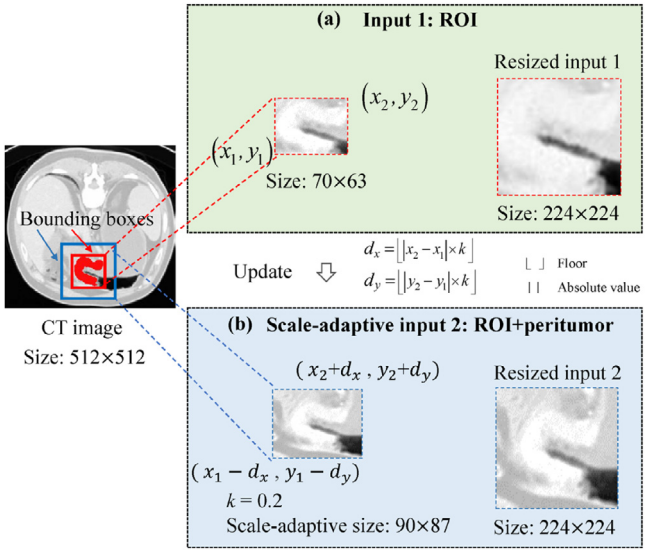

Fig. 2. Scale-adaptive inputs with multi-scale cropped images. (a) The input for the subnetwork of the encoder includes the ROI colored in red with a red bounding box for each CT scan. (b) The other input is the scale-adaptive ROI with a peritumoral region. The input image size was set to 224 × 224. For each image, according to its ROI coordinates of (x_1, y_1) and (x_2, y_2), we can obtain the corresponding expansion increments of d_x and d_y to find the optimal size of the peritumoral area. The proposed strategy enables the network to capture related features within the tumor and the immediately surrounding contexts with as little noise as possible.

where $\otimes$ represents element-wise production. The functions of $\mathbb{F}_{\mathbf{c}}$ and $\mathbb{F}_{\mathbf{s}}$ represent the operation of the CAM and SAM, respectively. Further details with respect to CAM and SAM are given as follows.

2.2.1. Channel attention module (CAM)

Our CAM attention module is proposed to select important information and reinforce it. The tensor of $\mathbf{X}$ can be a transformation, mapping the input tensor of $\mathbf{U}$. We take an operator of convolution to implement the transformation as

$$\mathbf{X} = Conv\,(\mathbf{U})\,, \tag{3}$$

and we describe $\mathbf{X_N}$ as

$$\mathbf{x}_N = \mathbf{c}_N * \mathbf{U} = \sum_{s=1}^{N} \mathbf{c}_N^s * \mathbf{u}^s, \tag{4}$$

where * represents convolution operation, $\mathbf{c}_N = \left[\mathbf{v}_N^1, \mathbf{v}_N^2, \ldots, \mathbf{v}_N^{N'}\right]$, $\mathbf{x}_N \in \mathbf{R}^{H \times W}$. $\mathbf{c}_N^s$ is a convolutional kernel denoting a channel $\mathbf{c}_N$ operating with its corresponding channel $\mathbf{U}$. The bias for the convolution is omitted for simplicity. For *Conv*, $\mathbf{x}_N$ is a summation of all the previous channels obtained by $\mathbf{c}_N$. Thus, each output feature map consists of global information of $\mathbf{X}$. However, the global information is present redundantly in various channels due to the operations in each local receptive field, which is dependent on the local relationship learned by the filters (Hu et al., 2018). To address these drawbacks, we propose a CAM to select and reinforce representative information from global information to improve the efficiency of the network and enhance its representational capacity to capture the relevant information.

As shown in Fig. 3a, to implement the design, we employed global average pooling (GAP) as a simple and efficient method to aggregate channel-wise statistics, which has been demonstrated in previous studies (Hu et al., 2018; Woo et al., 2018; Zhou, Khosla, Lapedriza, Oliva, & Torralba, 2016). GAP obtains an individual signal (tensor of $\mathbf{X}'$) for each channel. The process of operation for CAM is denoted as follows:

$$\mathbf{X}_c = \mathbb{F}_{\mathbf{c}}\,(\mathbf{X}) \otimes \mathbf{X} = GMP\,(ReLU\,(GAP(\mathbf{X}) - Mean(GAP(\mathbf{X}))) \otimes \mathbf{X}) \otimes \mathbf{X}, \tag{5}$$

where the functions of *GMP*, *ReLU*, *GAP*, and *Mean* denote the operations of GAP, ReLU, GAP, and the average of the value of $\mathbf{X}'$, respectively. $\otimes$ represents element-wise production. The Eq. (5) is used for selecting important features and reinforce them. We implement GAP operation to calculate retain representative features in all transformed feature maps and reduce feature dimensions. The output of GAP operation is an individual signal (tensor of $\mathbf{X}'$) for each channel. We then compute the mean value of the aggregated signals in all channels. ReLU is used to select the high-expression signals (tensor of $\mathbf{X}''$). Then, we obtain the selected feature maps by element-wise production of the vectors of $\mathbf{X}''$ and $\mathbf{X}$. When we obtain all representative regions by the GAP operation, we sequentially exploit global max pooling (GMP) to further capture local high-expression signals from all representative feature maps ($\mathbf{X}_c'$) aggregated by GAP. Then, we reinforce the input feature maps $\mathbf{X}$ by identifying the local high-expression signals and the operation of element-wise production.

Note that our CAM differs from the attention module known as CBAM that uses the parallel operations of GAP and GMP (Woo et al., 2018). The designed CAM is able to select representative features and reinforce them. CAM calculates the global information of channels to remove channels that do not contain important information, which can improve the efficiency of model calculation. However, the CBAM only can reinforce the all the channels rather than selection. We experimentally demonstrated that our proposed method is superior to CBAM for OS prediction, and the details are described below.

2.2.2. Spatial attention module (SAM)

The architecture of the proposed SAM is illustrated in Fig. 3b. Woo et al. indicated that spatial attention (a module in CBAM) is also crucial to locate the discriminative region (Woo et al., 2018). However, CBAM only aggregates spatial information with average and max pooling to capture the representative information or the overall spatial content of the corresponding pixel locations of different channels, which miss the representative information, or the overall spatial content obtained in the local receptive field of all of the channels. Hence, we tailor SAM not only to focus on the most informative points of spatial feature maps, but also to complement spatial attention with local representative information. The abovementioned process can be described as

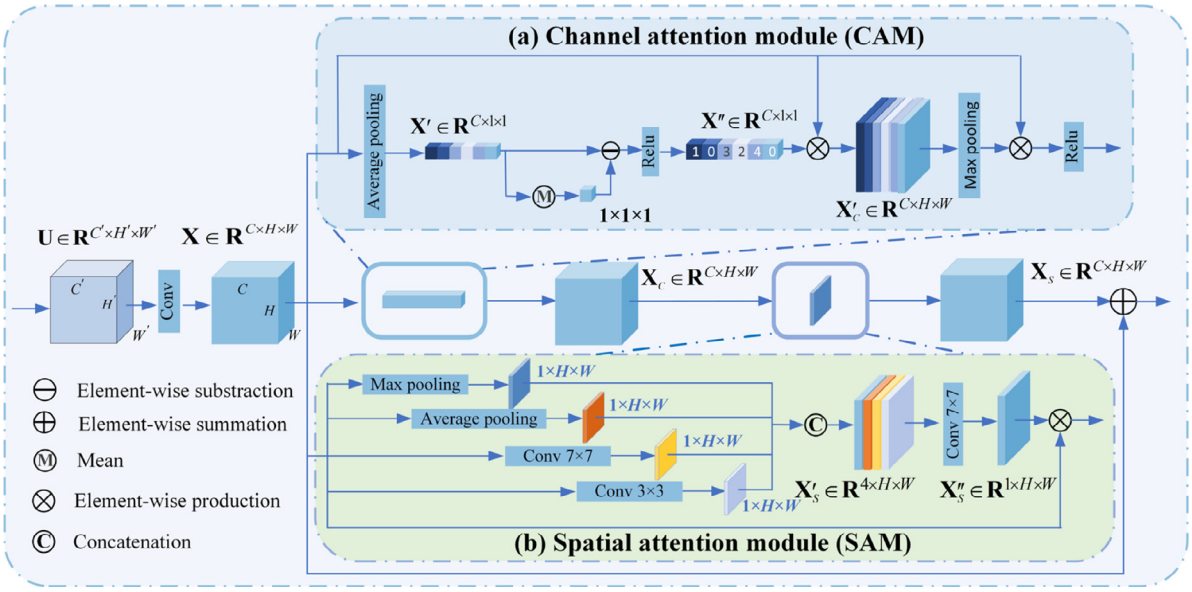

Fig. 3. Architecture of our proposed selection-and-reinforcement attention module (SRAM). The proposed attention modules enable the network to select and reinforce important features in both channel and spatial axes. (a) Channel attention module (CAM). In CAM, we apply the global average pooling and global max pooling to aggregate the 3D feature maps of **X** to a 1D vector. We achieve the selection of high-expression signals by simple operation of the rectified linear unit (ReLU) activation function. Then, we exploit the signals to reinforce the input feature maps. (b) Spatial attention submodule (SAM). SAM is designed to focus the informative points of spatial feature maps and complement spatial attention with local representative information.

follows:

$$\mathbf{X}_s' = Conc[AP(\mathbf{X}_c); MP(\mathbf{X}_c); Conv_{7\times7}(\mathbf{X}_c); Conv_{3\times3}(\mathbf{X}_c)], \tag{6}$$

$$\mathbf{X_s} = \mathbb{F}_s(\mathbf{X_c}) \otimes \mathbf{X_c} = Conv_{7\times7}(\mathbf{X}_s') \otimes \mathbf{X}_c, \tag{7}$$

where *Conc, AP, MP,* $Conv_{7\times7}$, and $Conv_{3\times3}$ denote the operations of concatenation, average-pooling, max pooling, convolution with a kernel size of 7 × 7, and convolution with kernel size of 3 × 3, respectively. The submodule of SAM is designed to focus on the most informative points of spatial feature maps, but also to complement spatial attention with local representative information. To achieve this, we exploit the operations of average pooling and max pooling to aggregate all channels into a single channel. We also exploit convolution with kernel sizes of 7 × 7 and 3 × 3 to generate a local informative combination. We then concatenate them as feature maps of $\mathbf{X}_s'$. Subsequently, we apply a convolution operation (kernel size 7 × 7) for the feature maps of $\mathbf{X}_s'$ to generate a spatial informative map $\mathbf{X}_s''$, which captures the spatial locations to be reinforced.

2.3. KMAP-Net architecture

Our proposed network comprises of two components. As shown in Fig. 4, the first part is a shared sibling network equipped with attention modules. Multi-level features in each block of sibling networks are concatenated, and we exploit layer-wise lateral connections to share multi-level feature maps for each task. Note that the branch for node stage prediction is the same as that for tumor stage prediction, which is omitted considering its simplicity and exploitation. We tailor two subnetworks with the same architecture. Our source codes of the proposed method will be available soon at https://github.com/dreamenwalker/KMAP-Net/.

2.4. Multi-level layer-wise lateral connection

We can extract low-level feature maps in the shallow layers of the decoder, which generalizes the all-side fusion of low-level semantic feature maps in the shadow bottom-top path. Focusing on low-level features is compelling due to the easy-to-understand semantic information and the attachment of stationary weights to specific locations. The task-oriented encoder with fused high-level feature maps has greater expressive power than single high-level feature maps in the last few convolutional layers. In particular, our network fuses low-level information separately and avoids information consumption in deep layers. This maintains the feature aggregation to adapt to OS prediction.

2.5. Loss function

As shown in Fig. 1, the proposed network includes three tasks. The tasks were trained jointly with different loss functions. For the survival task, we employed the negative log partial likelihood as a loss function to enable a controlled comparison with previous studies (e.g., Jiang et al. (2020) and Li et al. (2019)). We trained all the models by minimizing the loss function for the optimal estimation of parameter β, as given below:

$$L_{OS}(\beta) = -\frac{1}{N}\sum_{i=1}^{N}\left(\hat{h}_\beta(x_i) - \log\sum_{j\in A(T_j)} e^{\hat{h}_\beta(x_j)}\right), \tag{8}$$

where N is the number of patients with an observed status, and $A(T_j)$ is a set. For each patient, Tj is the survival time during the follow-up such that $T_j \geq T_i$. $\hat{h}_\beta(\boldsymbol{x})$ is the output of the proposed network.

The proposed network model was also used to perform clinical stage prediction, which is a classification task. Considering the limited number of patients, we conducted secondary tasks for clinical tumor and node stage prediction as binary classification, which was obtained via the evaluation of CT images by experienced radiologists. The total loss of the proposed network is defined as

$$L_{Total} = L_{OS} + L_{cTstage} + L_{cNstage}, \tag{9}$$

where L_{OS}, $L_{cTstage}$, and $L_{cNstage}$ represent the loss functions for the tasks of OS prediction, clinical tumor prediction, and node stage prediction, respectively.

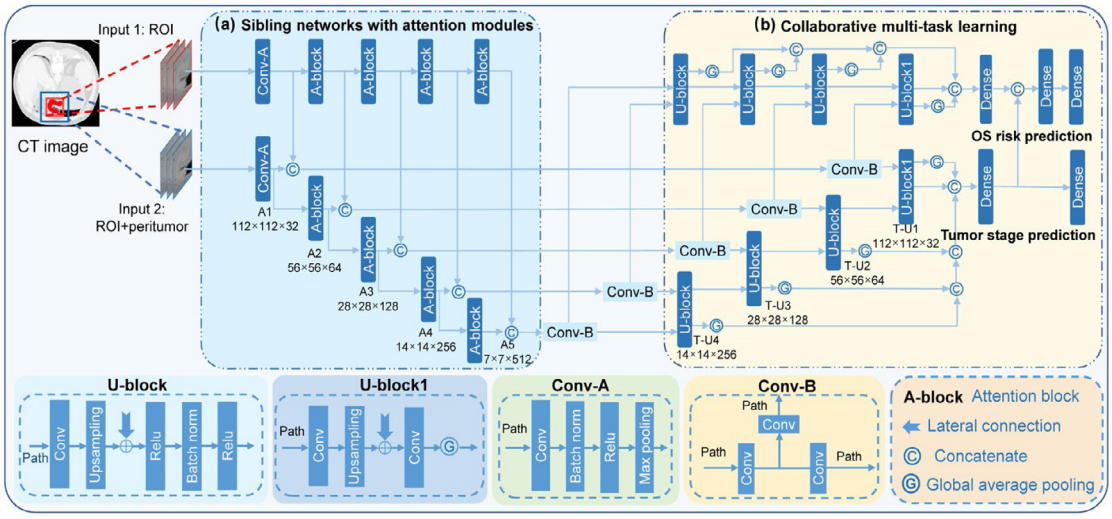

Fig. 4. Architecture of our KMAP-Net. (a) Decoder with pyramid sibling networks equipped with attention modules. Multi-level features in each block of sibling networks are concatenated, and we exploit layer-wise lateral connections to share multi-level feature maps for each task. (b) Collaborative multi-task learning with cascade connection. Note that the subnetwork for node stage is omitted considering the simplicity and explicitation, which is similar to the subnetwork for tumor stage prediction.

2.6. Implementation of training parameters

Two pre-processing methods were performed for each input ROI, including (1) min max normalization. For an image x, the pre-processed image was $x_{pre-processed} = (x\text{-}min)/(max\text{-}min)$, where min is the minimum gray value and max was the maximum gray value in the image x. This method was used to accelerate the convergence speed of the proposed network and improve its performance empirically, and (2) image resampling. The method was able to resample CT images to a target size of 224 × 224 for the input of our network based on a residual convolutional neural network. For our deep learning model, the network architecture is suitable for RGB images, resulting in a three-channel input. To be suitable for the requirement and decode the tumor phenotype entirely, each selected single-channel CT slice was copied twice, and the three single-channel slices were stacked as a three-channel image. To avoid the danger of overfitting, we employed data augmentation. For each CT slice in our datasets, classic augmentation techniques were used, including flipping, translation, rotation, scaling, adding Gaussian noise, and cropping. The average predicted probability was treated as the OS probability for each patient.

In this study, the network was trained for two types of tasks, including survival prediction and classification. We employed a stochastic gradient descent (SGD) algorithm, set the mini-batch size as 8 and trained the model for 300 epochs. We set the initial learning rate to 0.001. We performed network training using the TensorFlow and Keras libraries and trained the proposed network on eight NVIDIA 2080Ti GPUs with a 24TB buffer. Other parameters in the Keras library were left at default values, unless otherwise indicated.

3. Experimental setup and results

OS prediction has been investigated for decades and remains a challenge in existing model owing to the necessity of the laborious collection of survival data for each patient. The long-term purpose of this study is to explore a powerful model to accurately predict risk probability given the observed and censored patents with the status, time duration, clinical TNM staging information, and image data. In this study, we evaluated our proposed method and compared it on three independent datasets. The first two datasets consisted of patients with gastric cancer collected in two hospitals. The third dataset was a public lung cancer dataset. Associated program code will be made available for reproducibility.

In this study, we collected data on 879 lung cancer and gastric cancer patients including CT scans, follow-up information, and clinical stage information. Survival data included four parts for patient i (x_i, T_i, E_i, I_i): a patient's clinical variable x, an observed event time T, a status of event indicator E, and CT images I. The time was recorded for OS from the operation date to tumor-related death or final follow-up date. If the status of a patient (e.g., death) was observed, we called this complete survival data labeled $E = 1$. The corresponding time T denotes the duration from the operation date to tumor-related death. If the status of a patient was censored, we called this censored survival data and labeled it as $E = 0$. The corresponding time T denotes the duration from the operation date to the final follow-up date.

3.1. Multi-centric gastric cancer datasets

All gastric cancer patients enrolled in this study were pathologically confirmed. A total of 459 consecutive gastric cancer patients were collected from two independent centers: (1) Lanzhou University Second Hospital (337 cases) and (2) Guangdong General Hospital (122 cases). Baseline demographic and clinicopathological characteristics were retrospectively collected from the electronic medical records of each patient. CT imaging data were obtained from the picture archiving and communication system (PACS) of each hospital. The characteristics and clinicopathological variables used in the training and external validation sets are shown in Table 1.

We used the software ITK-SNAP (http://www.itksnap.org/) for segmentation. For each patient, each ROI is defined by using a tumor-centered rectangular bounding box according to the delineated boundary.

3.2. Public lung cancer dataset

To further reflect the generalizability and superior performance of the proposed model, we evaluated our method and the state-of-the-art methods on the public lung cancer dataset (Aerts et al., 2014; Mukherjee et al., 2020). In the public lung

Table 1
Clinical data for independent three hospitals datasets.

Clinical variables	Gastric (center 1)	Gastric (center 2)	Lung (public)
Number of patients	122	337	422
age (mean ± SD)	58 ± 12	55 ± 9	68 ± 10
Gender (%)			
Male	84 (68.9)	254(75.4)	290 (68.7)
Female	38 (31.1)	83 (24.6)	132 (31.3)
Clinical tumor stage (cTstage) (%)			
T1	5 (4.1)	0 (0.0)	93 (22.1)
T2	18 (14.8)	57 (16.9)	156 (37.1)
T3	68 (55.7)	183(54.3)	53 (12.6)
T4	31 (25.4)	97 (28.8)	119 (28.3)
Clinical node Nstage (cNstage) (%)			
N0	35 (28.7)	82 (24.3)	170 (40.3)
N1	37 (30.3)	71 (21.1)	23 (5.5)
N2	36 (29.5)	70 (20.8)	141 (33.4)
N3	14 (11.5)	114(33.8)	88 (20.9)
Clinical tumor-node-metastasis stage (cTNMstage) (%)			
I	24 (19.7)	101(30.0)	133 (31.5)
II	28 (23.0)	147(43.6)	112 (26.5)
III	70 (57.4)	89 (26.4)	177 (41.9)

cancer dataset, data on 422 patients with lung cancer are available for download. Among them, the CT images of two patients (numbered LUNG1-85 and LUNG1-192) had fewer layers skipped and fewer layers (including two tumor sections with small tumor area), which were excluded because they were ineligible for inclusion in the multi-center data based on the appropriate inclusion criteria. Finally, 420 patients were used to evaluate the proposed method and other competing methods. We randomly divided the data into a validation set (108 patients, IDs from 1–108) and a training set (312 patients, IDs from to 109–422) according to a fixed serial number and a ratio of one to four pairs.

3.3. Geometrical and clinical metric (assessment criteria)

3.3.1. Concordance index

Our method and existing methods were evaluated using Harrell's concordance index (c-index), which is a widely used indicator for performance evaluation (Harrell, Califf, Pryor, Lee, & Rosati, 1982), the formula was defined as:

$$C\left(\hat{h}, x_i\right) = \frac{1}{N_{observed\ T_j > T_i}} \sum_{observed\ T_j > T_i} \mathbf{1}_{\hat{h}_\beta(x_i) < \hat{h}_\beta(x_j)} \quad . \tag{10}$$

In the formula, the function of $\hat{h}_\beta(x_i)$ represents risk score for patient i for each model. The function $\mathbf{1}_{\hat{h}_\beta(x_i) < \hat{h}_\beta(x_j)} = 1$ if $\hat{h}_\beta(x_i) < \hat{h}_\beta(x_j)$, and 0 otherwise. The $N_{observed\ T_j > T_i}$ represents the number of pairs in the order of $T_j > T_i$, where T_i and T_j are the survival times for patients i and j, respectively, during the follow-up, and the T_j is observed. The c-index estimates the probability that, of two randomly chosen patients, the patient with a higher prognostic score will outlive the patient with a lower prognostic risk score (Raykar, Steck, Krishnapuram, Dehing-Oberije, & Lambin, 2007).

3.3.2. Hazard ratio

Hazard ratio is a widely used indicator to evaluate the prognostic value of the method to classify patients into different risk groups (Hernán, 2010). The widely accepted method was adopted to obtain the cut-off of the median risk score in the training set. Patients with risk scores lower than the cut-off were classified into the low-risk and the high-risk group otherwise. Assume that the output risk score is a risk factor in the low-risk and high-risk groups. The value of risk score 0 represents low-risk patients, and 1 represents high-risk patients. The hazard ratio (HR) is formulated as:

$$HR = \lambda_1(t, \boldsymbol{x}) / \lambda_2(t, \boldsymbol{x}) = \lambda_0(t)\, e^{h_\beta(\beta_1 \times 1)} / \lambda_0(t)\, e^{h_\beta(\beta_1 \times 0)} = e^{\beta_1}. \tag{11}$$

In this formula, the value of HR represents the ratio of risk functions between the high-risk group and the low-risk group; that is, the risk of morbidity in the high-risk group is HR times than that of the low-risk group. The larger the HR is, the higher the prognostic value of the method or model is. The clinical explanation of the regression coefficient β is the logarithm of the relative risk of the low-risk group and high-risk group covariates of x. If $\beta > 0$, the increase in the value of the corresponding covariate may be expected to increase the probability of death; if $\beta < 0$, the value of the corresponding covariant will reduce the probability of death; if $\beta = 0$, it indicates that the corresponding covariables are independent of the occurrence of the event.

3.3.3. Kaplan–Meier (KM) curve

To visualize and represent the prognostic value of the mentioned methods, Kaplan–Meier (KM) curves were depicted, showing time t as the horizontal axis and survival $Pro(T>t)$ as the vertical axis. In our study, we compared two risk groups divided by the mentioned method for personalized treatment using KM curves. The difference between the low-risk and high-risk groups was evaluated by log-rank test (Kleinbaum & Klein, 2012).

3.4. Model performance evaluation

3.4.1. Performance comparison with competing methods

We evaluated the proposed network and state-of-the-art work (e.g., clinical method (Li et al., 2019), ResNet. (Zhang et al., 2020), S-net (Jiang et al., 2020)) on gastric cancer datasets and public lung cancer datasets. We compared existing work with our proposed method, including clinical method (Li et al., 2019), deep learning networks of ResNet (Zhang et al., 2020), S-net (Jiang et al., 2020), VGG16 (Simonyan & Zisserman, 2014), VGG19 (Simonyan & Zisserman, 2014), DenseNet (Huang, Liu, Van Der Maaten, & Weinberger, 2017), ResNet50 (He, Zhang, Ren, & Sun, 2016), Inception (Szegedy et al., 2015), InceptionResNet (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017), NASNetMobile (Zoph, Vasudevan, Shlens, & Le, 2018), NASNetLarge (Zoph et al., 2018), and Xception (Chollet, 2017). As shown in Table 2, in the independent validation set of gastric cancer patients, our network outperformed the state-of-the-art methods for OS prediction of gastric cancer patients with the highest metrics (c-index: 0.74, 95% confidence interval (CI): 0.67–0.80, HR: 3.39, 95% CI: 1.53–7.51). The differences for comparison (KMAP-Net vs. ResNet vs. S-net = 0.74 vs. 0.62 vs. 0.61) were significant between our method and other methods (p-value<0.05, except for the clinical model). Our multi-task network demonstrated an accuracy of 0.81 for clinical tumor stage prediction (classifying patients into early or advanced stages) in the validation set. Meanwhile, the accuracy for clinical node stage prediction (with or without node metastasis) was 0.72 on the validation set. Furthermore, KM curves demonstrated that KMAP-Net was the most significant model in stratifying gastric cancer patients at high-risk versus low-risk (log-rank $p = 0.0014$, Fig. 5).

In the validation sets of public lung cancer patients, our model exhibited better performance than the existing methods, with the highest metrics (c-index: 0.66, 95% CI: 0.60–0.71; HR: 2.10, 95% CI: 1.37–3.21). The accuracies for the secondary tasks of the clinical tumor and node stages were 0.65 and 0.61, respectively. Meanwhile, KMAP-Net was the most significant model in stratifying lung cancer patients at high-risk versus low-risk (log-rank $p = 0.00054$, Fig. 5).

Table 2
Performance comparison of the KMAP-Net against existing methods of survival risk prediction in human cancer datasets.

Method	Gastric cancer Primary task of OS		Lung public dataset Primary task of OS	
	c-index	HR	c-index	HR
VGG16 (Simonyan & Zisserman, 2014)	0.50 (0.41–0.59)	0.88 (0.51-1.50)	0.55 (0.49–0.61)	1.36 (0.89-2.07)
VGG19 (Simonyan & Zisserman, 2014)	0.62 (0.53–0.70)	1.85 (1.02-3.35)	0.54 (0.48–0.60)	1.28 (0.85-1.94)
DenseNet (Huang et al., 2017)	0.68 (0.61–0.75)	2.29 (1.21-4.36)	0.62 (0.56–0.68)	1.54 (1.03-2.30)
ResNet50 (He et al., 2016)	0.66 (0.58–0.73)	1.72 (0.95-3.12)	0.57 (0.50–0.63)	1.82 (1.20-2.75)
Inception (Szegedy et al., 2015)	0.58 (0.51–0.66)	1.14 (0.66-1.98)	0.61 (0.55–0.67)	1.66 (1.10-2.49)
InceptionResNet (Szegedy et al., 2017)	0.69 (0.62–0.76)	2.31 (1.21-4.39)	0.60 (0.54–0.66)	1.48 (0.99-2.22)
NASNetMobile (Zoph et al., 2018)	0.61 (0.52–0.70)	1.88 (1.03-3.41)	0.62 (0.57–0.68)	1.99 (1.32-3.00)
NASNetLarge (Zoph et al., 2018)	0.67 (0.59–0.75)	2.80 (1.41-5.58)	0.60 (0.54–0.66)	1.54 (1.02-2.32)
Xception (Chollet, 2017)	0.71 (0.64–0.77)	3.16 (1.49-6.71)	0.60 (0.53–0.66)	1.56 (1.03-2.36)
ResNet (Zhang et al., 2020)	0.62 (0.54–0.70)	2.24 (1.18-4.26)	0.56 (0.50–0.62)	1.55 (1.02-2.37)
S-net (Jiang et al., 2020)	0.61 (0.53–0.70)	1.98 (1.04-3.76)	0.47 (0.42–0.52)	0.92 (0.69-1.24)
Clinical stage(Li et al., 2019)	0.70 (0.63–0.76)	2.23 (1.30-3.81)	0.57 (0.50–0.64)	1.56 (1.03-2.36)
Ours	**0.74** (0.67–0.80)	**3.39 (1.53–7.51)**	**0.66*** **(0.60–0.71)**	**2.10 (1.37–3.21)**

Note:Clinical stage represents the clinical survival model constructed by clinical tumor stage, node stage, and TNM stage. Acc-T and Acc-N represent the accuracy of the prediction of clinical tumor and node stages, respectively. * indicates that our method outperformed other competing methods with a significant difference ($p<0.05$). HR: hazard ratio. c-index: concordance index.

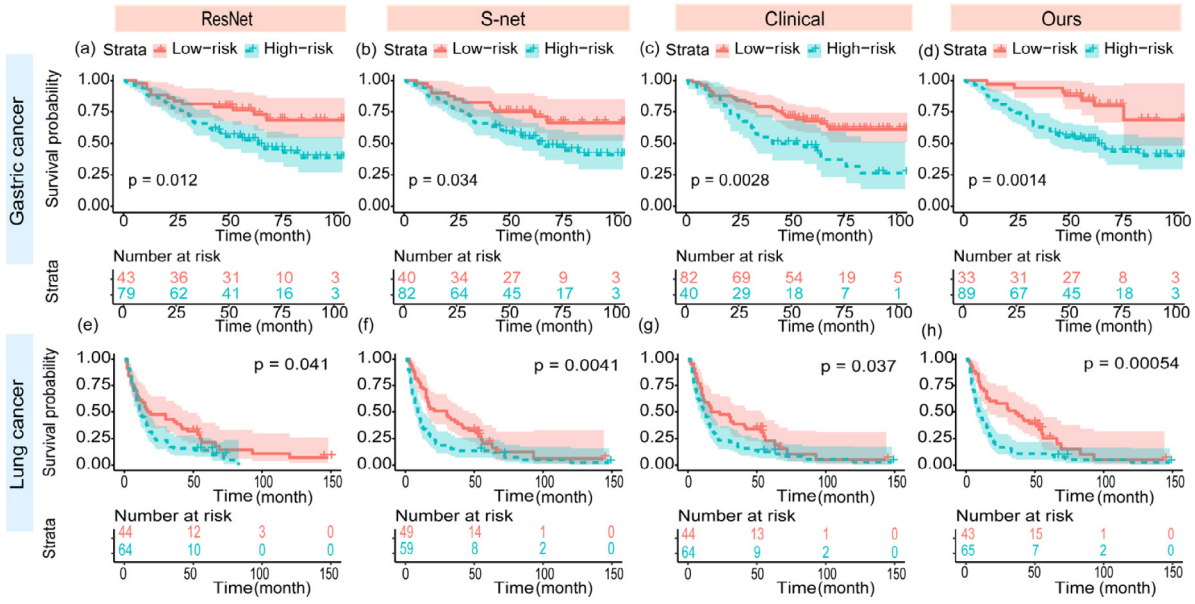

Fig. 5. Prognostic value evaluation using Kaplan–Meier (KM) curves of KMAP-Net against existing methods of survival risk prediction in the gastric and lung cancer datasets, respectively. All the model can stratify all patients into high-risk and low-risk groups. The result shows that KMAP-Net was the most significant model in stratifying gastric cancer patients at high-risk versus low-risk (log-rank p=0.0014). For each survival curve, a p-value is calculated by the log-rank test, which shows the differences in prognosis between high-risk and low-risk groups. The median predicted risk score of each method was applied to divide patients into low-risk and high-risk groups. For each survival curve, a p-value is calculated by log-rank test, which shows the differences in prognosis between high-risk and low-risk groups.

3.4.2. Effectiveness of our proposed attention modules

To further investigate whether the module proposed in this study is effective, we compared our proposed attention module with the backbone (no attention module), state-of-the-art attention modules of Squeeze-and-Excitation Network (SENet) and CBAM on the datasets of gastric cancer and lung cancer patients (Hu et al., 2018; Woo et al., 2018), respectively. In the validation set of GC patients (Table 3), the baseline (no attention module) showed the poorest performance (c-index: 0.65, 95% CI: 0.58–0.72; HR: 2.15, 95% CI: 1.17–3.95). The baseline network equipped with our attention module showed better performance than the baseline equipped with SENet, and the baseline equipped with CBAM, respectively (c-index: Ours vs. CBAM vs. SENet: 0.74 vs. 0.66 vs. 0.66;HR: 3.39 vs. 2.61 vs. 2.06).

In the public dataset of lung cancer patients, our proposed attention module was able to significantly improve the performance of the baseline network. The model performance of the baseline was slightly better than that of random prediction that achieved a c-index of 0.58 (95% CI: 0.52–0.64). When the baseline was equipped with our proposed attention modules, the performance showed a significant incremental margin compared to the existing modules with the c-index (Ours vs. CBAM vs. SENet = 0.66 vs. 0.59 vs. 0.62; HR: 2.10 vs. 1.62 vs. 1.90, and the p-values for comparison of the c-index were less than 0.05. The KM curves demonstrated that our module can boost model performance better in stratifying gastric and lung cancer patients at high-risk versus low-risk (Fig. 6).

To further investigate whether the module proposed in this study is effective, we compared our proposed attention module with the state-of-the-art attention modules of Squeeze-and-Excitation Network (SENet) and CBAM on the datasets of gastric cancer and lung cancer patients (Hu et al., 2018; Woo et al., 2018), respectively. In the validation set of GC patients (Table 3), the baseline (no attention module) showed the poorest performance (c-index: 0.65, 95% CI: 0.58–0.72; HR: 2.15, 95% CI: 1.17–3.95). The baseline network equipped with our attention

Table 3
Performance comparisons of the baselines equipped with different attention modules.

Method	Gastric cancer dataset				Lung public dataset			
	Primary task of OS		Secondary tasks		Primary task of OS		Secondary tasks	
	c-index	HR	Acc-T	Acc-N	c-index	HR	Acc-T	Acc-N
Baseline	0.65 (0.58–0.72)	2.15 (1.17–3.95)	0.73	0.73	0.58 (0.52–0.64)	1.45 (0.96–2.18)	0.62	0.58
Baseline + SENet	0.66 (0.58–0.73)	2.06 (1.10–3.85)	**0.81**	**0.74**	0.62 (0.56–0.69)	1.90 (1.24-2.91)	0.62	**0.62**
Baseline + CBAM	0.66 (0.58–0.73)	2.61 (1.35–5.07)	0.67	0.66	0.59 (0.52–0.65)	1.62 (1.08–2.45)	0.65	0.55
Baseline + SRAM (ours)	**0.74** (0.67–0.80)	**3.39** (1.53–7.51)	**0.81**	0.72	**0.66** (0.60–0.71)	**2.10** (1.37–3.21)	**0.65**	0.61

Note: The baseline represents a network that is not equipped with the attention module. The SRAM is the proposed attention module.

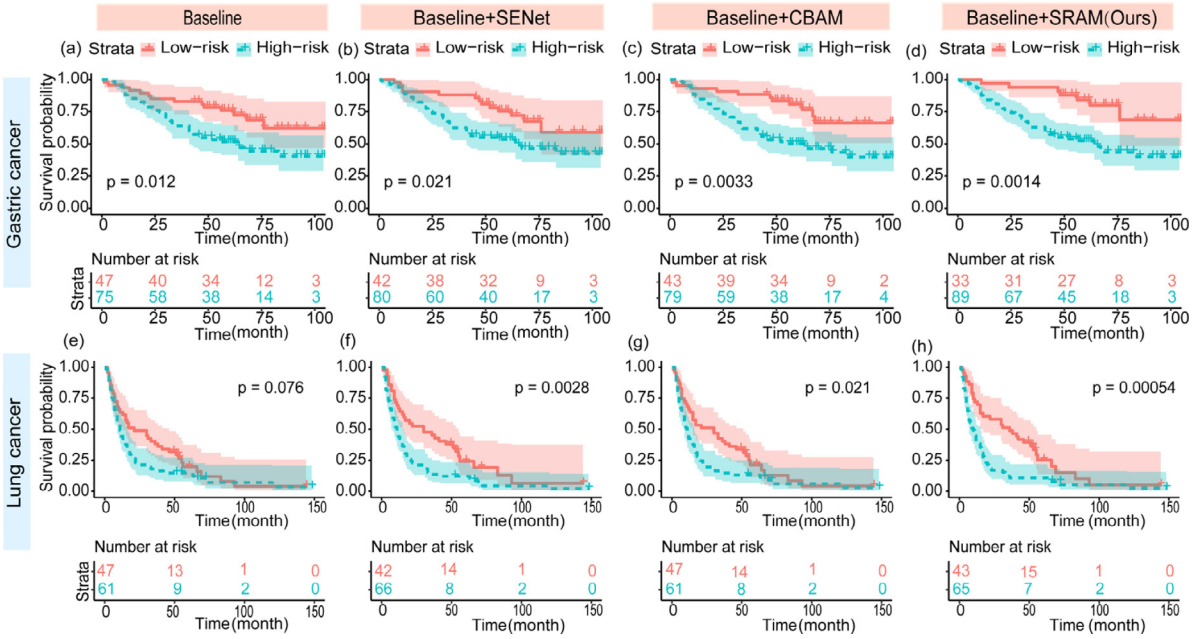

Fig. 6. Prognostic value evaluation using KM curves of the baselines equipped with different attention modules in the gastric and lung cancer datasets, respectively. We found that baseline model showed poor performance for OS prediction of lung cancer patients (no significance between high-risk and low-risk groups, log-rank p=0.076). The results showed that not all the attention modules can boost baseline model performance. The comparisons demonstrated that our proposed attention module can boost model performance better in stratifying gastric and lung cancer patients at high-risk group versus low-risk group.

module showed better performance than the baseline equipped with SENet, and the baseline equipped with CBAM, respectively (c-index: Ours vs. CBAM vs. SENet: 0.74 vs. 0.66 vs. 0.66;HR: 3.39 vs. 2.61 vs. 2.06). The p-values for comparison of c-index were less than 0.05.

3.4.3. Effectiveness of different knowledge-guided tasks

To investigate the impact of different clinical stage tasks on survival risk prediction performance in a multi-task network, we designed ablation studies on different datasets (Table 4). In the gastric cancer dataset, the results showed that the performance of the network was poor when the network was used only for the primary task of survival prediction (c-index = 0.65, 95% CI: 0.58–0.72; HR = 2.06; 95% CI:1.06–4.00). When only one clinical stage task was added, the network prediction performance is improved (c-index: Task_{OS} vs. $\text{Task}_{\text{OS+T}}$ vs. $\text{Task}_{\text{OS+N}}$ = 0.65 vs. 0.68 vs. 0.66, HR: Task_{OS} vs. $\text{Task}_{\text{OS+T}}$ vs. $\text{Task}_{\text{OS+N}}$ = 2.06 vs. 1.68 vs. 2.02). When clinical T and N staging were both added into the backbone network, the multi-task network showed the best performance with a c-index of 0.74 (95% CI: 0.67–0.80) and HR of 3.39 (95% CI: 1.53–7.51).

In the datasets of lung cancer patients, the results demonstrate that single task (Task_{OS}) learning was poor for survival prediction. When the survival task was combined with clinical T stage prediction or clinical N stage prediction, the performance was improved for survival prediction (c-index: Task_{OS} vs. $\text{Task}_{\text{OS+T}}$ vs. $\text{Task}_{\text{OS+N}}$: 0.51 vs 0.63 vs 0.69; HR: 1.1 vs 1.94 vs 1.47). When the survival task was combined with both cTstage and cNstage, the multi-task network showed the best performance with c-index of 0.66 (95% CI: 0.60–0.71) and HR: 2.10 (95% CI: 1.37–3.21). The KM curves also indicated that knowledge-guided tasks can have an incremental contribution (Fig. 7).

3.4.4. Effects of multi-task scaling factor settings

To select the optimal value of scaling factor for survival risk prediction performance in a multi-task network, we set different values of scaling factor (k) on different human cancer datasets (Table 5). We experimentally found that the network is the most powerful with the scaling factor k of 0.2. The KM curves (Fig. 8) also present the best prognostic value of the KMAP-Net when the scaling factor is 0.2.

4. Discussion

We propose a collaborative knowledge-guided and task-oriented network of KMAP-Net with a tailored attention mechanism to improve the accuracy of OS prediction based on CT images. The main contributions include: (1) we propose an independent module equipped with attention mechanism to select and reinforce important features in both channel and spatial axes; (2) we propose a learning strategy of scale-adaptive inputs to capture related features within CT images. We experimentally demonstrated that the proposed network architecture could improve

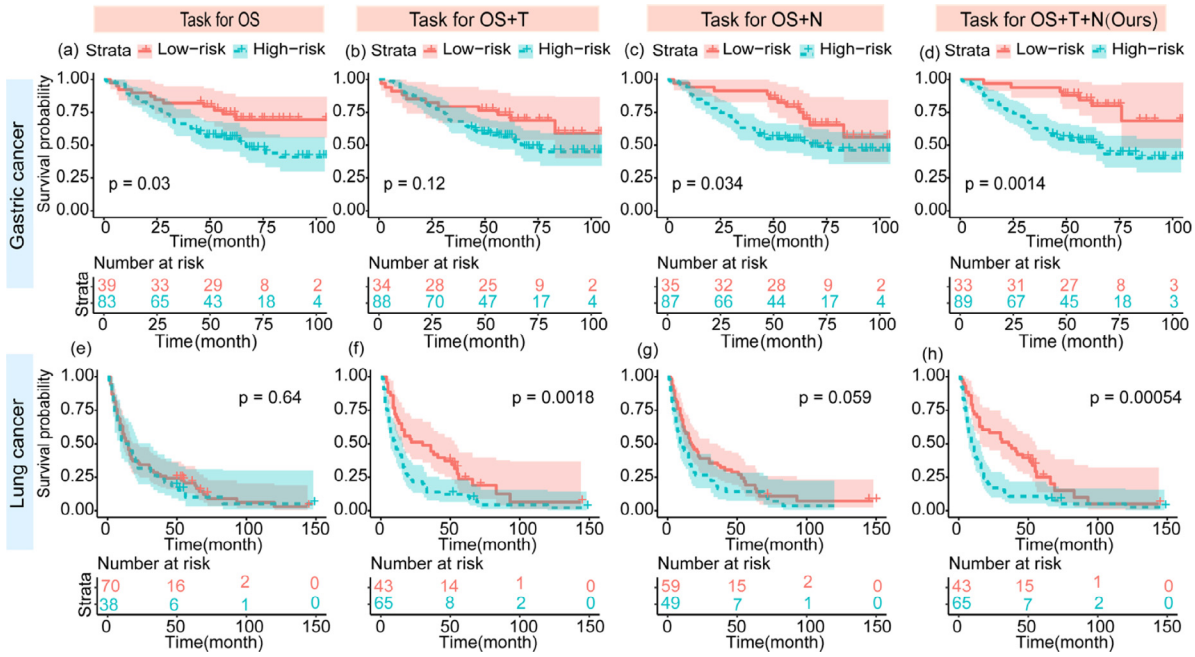

Fig. 7. Prognostic value evaluation using KM curves for performance contribution of each task in the KMAP-Net in the gastric and lung cancer datasets, respectively. The results showed that single task ($Task_{OS}$) learning was poor for OS prediction in lung cancer patients (no significance between high-risk and low-risk groups, log-rank p=0.64). When the survival task was combined with clinical T stage prediction or clinical N stage prediction, the performance was improved for survival prediction in lung cancer patients. Both knowledge-guided tasks of clinical T stage prediction and clinical N stage prediction had incremental contribution for OS prediction in gastric cancer and lung cancer patients.

Table 4
Performance contribution of each task in the KMAP-Net.

Method	Gastric cancer dataset				Lung public dataset			
	Primary task of OS		Secondary tasks		Primary task of OS		Secondary tasks	
	c-index	HR	Acc-T	Acc-N	c-index	HR	Acc-T	Acc-N
OS	0.65 (0.58–0.72)	2.06 (1.06–4.00)	–	–	0.51 (0.44–0.57)	1.10 (0.73–1.68)	–	–
OS+T	0.68 (0.60–0.76)	1.68 (0.86–3.25)	0.78	–	0.63 (0.56–0.69)	1.94 (1.27–2.96)	0.64	–
OS+N	0.66 (0.59–0.74)	2.02 (1.04–3.93)	–	0.66	0.59 (0.53–0.65)	1.47 (0.99–2.20)	–	0.60
OS+T+N	**0.74** (0.67–0.80)	**3.39** (1.53–7.51)	**0.81**	**0.72**	**0.66** (0.60–0.71)	**2.10** (1.37–3.21)	**0.65**	**0.61**

Table 5
Performance evaluation of the proposed KMAP-Net with different scaling factors.

Scaling factor k	Gastric cancer dataset				Lung public dataset			
	Primary task of OS		Secondary tasks		Primary task of OS		Secondary tasks	
	c-index	HR	Acc-T	Acc-N	c-index	HR	Acc-T	Acc-N
$k = 0$	0.70 (0.63–0.77)	2.21 (1.11–4.39)	0.81	0.71	0.61 (0.55–0.67)	1.57 (1.04–2.35)	0.64	0.57
$k = 0.1$	0.72 (0.65–0.79)	2.51 (1.22–5.14)	0.78	0.72	0.64 (0.58–0.70)	1.94 (1.28–2.96)	0.65	0.55
$k = 0.2$	0.74 (0.67–0.80)	3.39 (1.53–7.51)	0.81	0.72	0.66 (0.60–0.71)	2.10 (1.37–3.21)	0.65	0.61
$k = 0.3$	0.71 (0.64–0.77)	3.32 (1.32–8.34)	0.81	0.71	0.61 (0.55–0.66)	1.74 (1.15–2.62)	0.60	0.60

the accuracy of risk prediction. Meanwhile, our multi-task architecture enables gain incremental margins for the target of survival risk prediction, which indicates that the knowledge-guided task of cTNM staging prediction can further exploit valuable information of clinical TNM staging.

For selection of scaling factor k, when we experimentally set k value as 0.4, we found that the model performance evaluated in lung cancer dataset showed the poorest HR of 1.56 (1.03–2.35) compared with the result obtained using scaling factor k range from 0 to 0.3 (Table 5). Furthermore, the model performance evaluated in gastric cancer dataset also shows the lowest c-index of 0.69 (0.62–0.77). The results indicate that our proposed learning strategy based on scale-adaptive inputs can capture rich information from 2D CT images of intratumoral and peritumoral areas.

We experimentally found that the task-oriented network outperformed the S-net (Jiang et al., 2020), residual network (Zhang et al., 2020), and TNM staging manual (Li et al., 2019) evaluated in terms of metrics including c-index, HR, and KM curves. Our results also indicate that the multi-task network showed an incremental margin compared to the mono-task network for survival risk prediction, which demonstrates the superiority of the multi-task network compared with the mono-task network for OS prediction. Therefore, our results demonstrate that our tailored multi-task architecture jointly learned representative multi-level semantic features. Meanwhile, our work proposed a novel strategy using clinical stages as collaborative tasks to train the model. which demonstrates that multi-task architecture can further exploit valuable information of clinical TNM staging. The findings indicated that the knowledge-guided tasks of cTNM staging

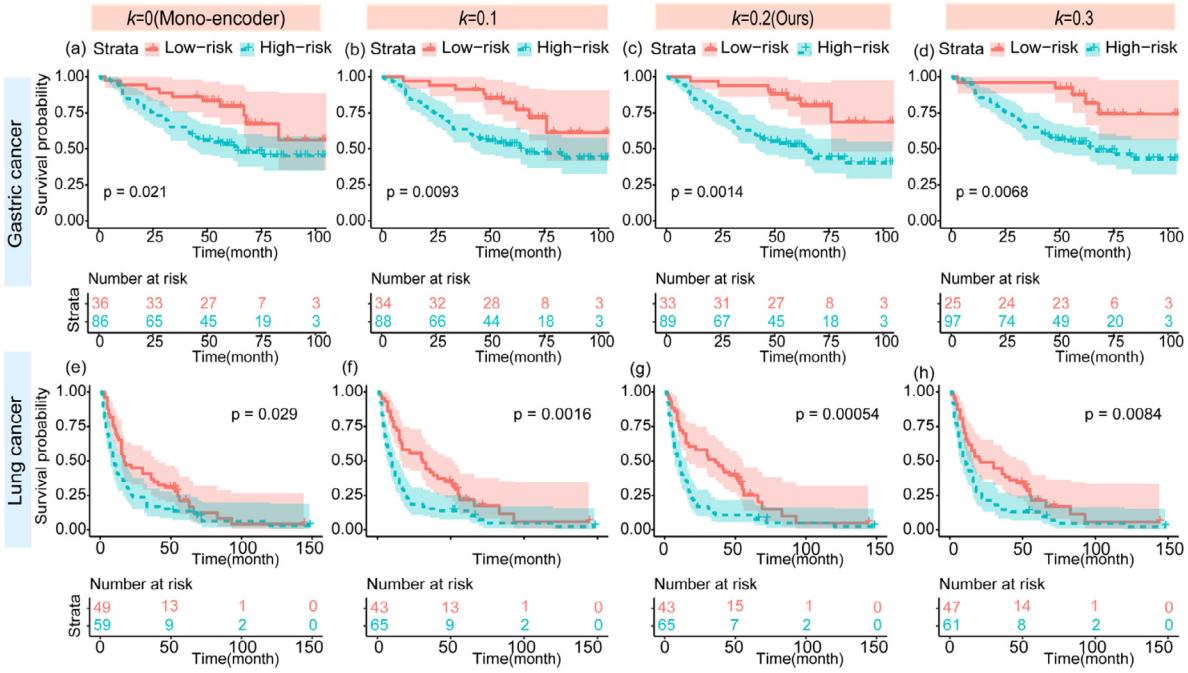

Fig. 8. Prognostic value evaluation using Kaplan–Meier (KM) curves for the effects of multi-task scaling factor settings for prediction in the gastric and lung cancer datasets, respectively. For each survival curve, a *p*-value is calculated by log-rank test, which shows the differences in prognosis between high-risk and low-risk groups. The result showed that when the input of scale-adaptive ROI with a certain peritumoral region can provide useful information for improvement of the model performance. The best results were obtained when a parameter of scaling factor k was set as 0.2.

prediction gain incremental margins for the target task of survival risk prediction. Note that our well-designed network can not only improve the accuracy of OS prediction for gastric cancers, but can also show prognostic value for lung cancer patients simultaneously.

Our results show that MTL is helpful in predicting prognosis than networks based only on CT images without any guidance from clinical information. Studies have shown that the intuitive plausibility for MTL has been proven in three respects. (1) It increases the sample size for training, (2) improves accuracy by learning new tasks with the guidance of knowledge acquired by learning related tasks, and (3) reduces the risk of overfitting (Vandenhende, Georgoulis, Proesmans, Dai, & Gool, 2020; Zhang & Yang, 2018). Tang et al. proposed a multi-task network for prediction of genomic biomarkers and survival time of glioblastoma patients, and the results showed that multi-task learning could improve the accuracy of OS prediction (Tang et al., 2020). Our results demonstrate that MTL is able to learn representative features due to its design by utilizing useful information contained in multiple learning tasks to help train a more accurate machine learning network model for prediction of clinical stages and survival risk of gastric and lung cancer patients. Besides, the results show that the model performance (c-index) is improved when the secondary tasks are integrated. However, single secondary task of N stage or T stage prediction not boost assessment criteria of hazard ratio (HR) for overall survival (OS) prognostic prediction in gastric cancer patients (HR : Task_{OS} vs $\text{Task}_{\text{OS+T}}$ vs $\text{Task}_{\text{OS+N}}$: 2.06 vs 1.68 vs 2.02). The results indicate that single different stage task may not provide stably incremental margin for OS prediction.

To improve the model performance for OS prediction, we proposed an attention module of SRAM to select important features and reinforce them both in channel and spatial axes, which could remove redundant information and reduce the computational cost of the model. In our ablation study, our designed attention mechanism SRAM was superior to the popular attention modules SENet and CBAM for the improvement of OS prediction (Hu et al., 2018; Woo et al., 2018). Although the attention modules of SENet and CBAM are significant for detection tasks, they fail to shrink all of the extracted features and select the representative information. The main reason is that none of the mentioned attention networks were equipped with the function of shrinking all of the features and selecting the representative information instead of emphasizing or suppressing features. Our proposed SRAM comprised of CAM and SAM. The modules are designed to sequentially focus on information in both the channel and spatial axes so that each convolutional block can find important channels and regions to emphasize.

Our experiments indicate that the significance of the three clinical stages in the prognosis analysis for model performance followed the order: $\text{Task}_{\text{OS+T+N}} > \text{Task}_{\text{OS+T}} > \text{Task}_{\text{OS+N}}$, which is consistent with the radiologists' consensus. We experimentally demonstrated that our proposed network could provide prognostic value for both lung cancer and gastric cancer patients. Accurate risk prediction of different cancer species using CT images has been of increasing interest in survival analysis (Jiang et al., 2020; Mukherjee et al., 2020). However, these studies only focused on a single cancer species for survival prediction. Few studies have focused on exploring a CNN network that can predict prognosis for both lung and gastric cancer patients using CT images. Currently, the clinical TNM staging manual is a widely used guideline for lung and gastric patients, which is based on valuable information obtained from CT images based on radiologists' experience. The results confirmed our hypothesis that the task for stage prediction would improve the accuracy of risk prediction for OS in lung and gastric cancer patients. Our results demonstrated that our network was able to learn prognostic features for both lung and gastric cancer patients with auxiliary

guidance for prediction of clinical stages. This may be a potential tool to aid radiologists in decision-making.

The results show that the model performance for OS prediction of gastric cancer patients is better than the performance for OS prediction of lung cancer patients, which is consistent with published work (Huang et al., 2016 and Mukherjee et al., 2020). Mukherjee et al. also found that the accuracy of clinical model based on clinical features of age, sex, histology and cancer stage prediction is also poor (c-indexes of clinical model are 0.69, 0.58, 0.55 and 0.52 in four cohorts, respectively). Huang et al. also showed that the AJCC staging system had poor C-index of 0.629. Actually, the label for clinical tumor stages and node stages of lung cancer patients are not accurate due to the subjective radiologists' evaluation. Therefore, the accuracy for prediction of tumor or node stage for lung cancer is poor. We should note that the prediction for tumor or node stage is the secondary task to improve the prediction of OS, although the poor accuracy in predicting the tumor or node stage for lung cancer may be not acceptable clinically.

Our study has some limitations. First, although our network is evaluated on two kinds of cancers, one of limitations of our work is that our model should be further evaluated in other cancer datasets, which is necessary to show the model robustness and generalization for potentially clinical application. Furthermore, our model also should be further tested in other modalities. Besides, the size of each dataset was limited due to the cost of survival data requisition. In contrast, the performance of the proposed method should be further validated for other human cancers. Meanwhile, although we trained our method with multi-center datasets, the samples in different clinical stages are unbalanced due to scarce survival data with clinical stage information, which is not analyzed on the performance of the proposed network models. We only investigated the importance of clinical stage as a task to improve the accuracy of risk prediction, and whether our network can show greater prognostic value with other tasks should be explored. Finally, we only investigate the applicability of our method for CT images, and more modalities should be explored.

5. Conclusion

We propose a knowledge-guided multi-task attention pyramid network (KMAP-Net) to improve the model performance of OS prediction based on CT images in lung and gastric cancer patients. Our proposed multi-task network equipped with the tailored attention module is a powerful model for improving the accuracy of risk prediction in lung and gastric patients. The multi-task architecture exploits its design by capturing and sharing information from other related learning tasks, enabling our proposed network to better generalize our original task and avoid overfitting owing to the limited sample size. The results indicate that our method is a potential assistive tool for decision-making in clinical practice. Besides, further work should be done to integrate automatic segmentation network combined with other modality as input to predict survival risk with different cancer dataset collections.

CRediT authorship contribution statement

Liwen Zhang: Conceptualization, Investigation, Methodology, Software, Writing – original draft. **Lianzhen Zhong:** Methodology, Software, Writing – original draft. **Cong Li:** Methodology, Software, Writing – review & editing. **Wenjuan Zhang:** Investigation, Data curation. **Chaoen Hu:** Methodology, Software. **Di Dong:** Conceptualization, Writing – review & editing. **Zaiyi Liu:** Conceptualization, Supervision, Data curation. **Junlin Zhou:** Validation, Resources, Data curation. **Jie Tian:** Conceptualization, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Ministry of Science and Technology of China under Grant No. 2017YFA0205200, National Key R&D Program of China (2017YFC1309100), National Natural Science Foundation of China (62027901, 82022036, 91959130, 81971776, 81771924, 6202790004, 81930053), Chinese Academy of Sciences under Grant No. GJJSTD20170004 and QYZDJ-SSW-JSC005, the Project of High-Level Talents Team Introduction in Zhuhai City (Zhuhai HLHPTP201703), Strategic Priority Research Program of Chinese Academy of Sciences (XDB38040200), Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application (No. 2022B1212010011), Special Foundation of State Key Laboratory of Complex Systems Management and Control (2022QN03) and the Youth Innovation Promotion Association CAS (2017175). The authors would like to acknowledge the instrumental and technical support of Multimodal Biomedical Imaging Experimental Platform, Institute of Automation, Chinese Academy of Sciences.

References

Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, *5*, 1–9.

Ajani, J. A., D'Amico, T. A., Almhanna, K., Bentrem, D. J., Chao, J., Das, P., et al. (2016). Gastric cancer, version 3.2016, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network*, *14*, 1286–1312.

Amin, M. B., & Edge, S. B. (2017). AJCC cancer staging manual: springer.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*, 394–424.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251–1258).

Dai, J., He, K., & Sun, J. (2016). Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3150–3158).

Dong, D., Fang, M.-J., Tang, L., Shan, X.-H., Gao, J.-B., Giganti, F., et al. (2020). Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multi-center study. *Annals of Oncology*, *31*, 912–920.

Dorado-Moreno, M., Navarin, N., Gutiérrez, P. A., Prieto, L., Sperduti, A., Salcedo-Sanz, S., et al. (2020). Multi-task learning for the prediction of wind power ramp events with deep neural networks. *Neural Networks*, *123*, 401–411.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama*, *247*, 2543–2546.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology (Cambridge, Mass)*, *21*, 13–15.

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran Jr., W. J., Wu, Y.-L., et al. (2017). Lung cancer: current therapies and new targeted treatments. *The Lancet*, *389*, 299–311.

Hong, T. S., Tomé, W. A., & Harari, P. M. (2012). Heterogeneity in head and neck IMRT target design and clinical practice. *Radiotherapy and Oncology*, *103*, 92–98.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132–7141).

Huang, Y., Liu, Z., He, L., Chen, X., Pan, D., Ma, Z., et al. (2016). Radiomics signature: A potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell Lung cancer. *Radiology*, *281*, 947–957.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).

Jiang, Y., Jin, C., Yu, H., Wu, J., Chen, C., Yuan, Q., et al. (2020). Development and validation of a deep learning CT signature to predict survival and chemotherapy benefit in gastric cancer: A multicenter, retrospective study. *Annals of Surgery*, *274*, e1153–e1161.

Kather, J. N., Pearson, A. T., Halama, N., Jäger, D., Krause, J., Loosen, S. H., et al. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature Medicine*, *25*, 1054–1056.

Kattan, M. W., Hess, K. R., Amin, M. B., Lu, Y., Moons, K. G., Gershenwald, J. E., et al. (2016). American joint committee on cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA: A Cancer Journal for Clinicians*, *66*, 370–374.

Kim, B.-C., Yoon, J. S., Choi, J.-S., & Suk, H.-I. (2019). Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection. *Neural Networks*, *115*, 1–10.

Kleinbaum, D. G., & Klein, M. (2012). Kaplan–Meier survival curves and the log-rank test. In *Survival analysis* (pp. 55–96). Springer.

Li, W., Zhang, L., Tian, C., Song, H., Fang, M., Hu, C., et al. (2019). Prognostic value of computed tomography radiomics features in patients with gastric cancer following curative resection. *European Radiology*, *29*, 3079–3089.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117–2125).

Liu, S., Johns, E., & Davison, A. J. (2019). End-to-end multi-task learning with attention. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1871–1880).

Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8759–8768).

Liu, F., Wang, K., Liu, D., Yang, X., & Tian, J. (2021). Deep pyramid local attention neural network for cardiac structure segmentation in two-dimensional echocardiography. *Medical Image Analysis*, *67*, Article 101873.

Lu, S., Wang, S.-H., & Zhang, Y.-D. (2020). Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm. *Neural Computing and Applications*, *33*, 10799–10811.

Lu, S., Zhu, Z., Gorriz, J. M., Wang, S. H., & Zhang, Y. D. (2022). NAGNN: Classification of COVID-19 based on neighboring aware representation from deep graph neural network. *International Journal of Intelligent Systems*, *37*, 1572–1598.

Mukherjee, P., Zhou, M., Lee, E., Schicht, A., Balagurunathan, Y., Napel, S., et al. (2020). A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional computed tomography image datasets. *Nature Machine Intelligence*, *2*, 274–282.

Pak, K. H., Jo, A., Choi, H. J., Choi, Y., Kim, H., & Cheong, J.-H. (2015). The different role of intratumoral and peritumoral lymphangiogenesis in gastric cancer progression and prognosis. *Bmc Cancer*, *15*(498).

Pang, S., Du, A., Orgun, M. A., Wang, Y., & Yu, Z. (2021). Tumor attention networks: Better feature selection, better tumor segmentation. *Neural Networks*, *140*, 203–222.

Raykar, V. C., Steck, H., Krishnapuram, B., Dehing-Oberije, C., & Lambin, P. (2007). On ranking in survival analysis: Bounds on the concordance index. In *Conference on advances in neural information processing systems*.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv:abs/1706.05098.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).

Tang, Z., Xu, Y., Jin, L., Aibaidula, A., Lu, J., Jiao, Z., et al. (2020). Deep learning of imaging Phenotype and genotype for predicting overall survival time of Glioblastoma patients. *IEEE Transactions on Medical Imaging*, *39*, 2100–2109.

Tegels, J. J., De Maat, M. F., Hulsewé, K. W., Hoofwijk, A. G., & Stoot, J. H. (2014). Improving the outcomes in gastric cancer surgery. *World Journal of Gastroenterology: WJG*, *20*(13692).

Vandenhende, S., Georgoulis, S., Proesmans, M., Dai, D., & Gool, L. V. (2020). Revisiting multi-task learning in the deep learning Era. arXiv:abs/2004.13379.

Wang, X.-X., Ding, Y., Wang, S.-W., Dong, D., Li, H.-L., Chen, J., et al. (2020). Intratumoral and peritumoral radiomics analysis for preoperative Lauren classification in gastric cancer. *Cancer Imaging*, *20*, 1–10.

Wang, S., Shi, J., Ye, Z., Dong, D., Yu, D., Zhou, M., et al. (2019). Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *European Respiratory Journal*, *53*, Article 1800986.

Warfield, S. K., Zou, K. H., & Wells, W. M. (2008). Validation of image segmentation by estimating rater bias and variance. *Philosophical Transactions of the Royal Society of London A (Mathematical and Physical Sciences)*, *366*, 2361–2375.

Woo, S., Park, J., Lee, J.-Y., & So Kweon, I. (2018). Cbam: Convolutional block attention module. In *Proceedings of the european conference on computer vision* (pp. 3–19).

Wu, X., Hui, H., Niu, M., Li, L., Wang, L., He, B., et al. (2020). Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study. *European Journal of Radiology*, Article 109041.

Zhang, L. W., Chen, B. J., Liu, X., Song, J. D., Fang, M. J., Hu, C. E., et al. (2018). Quantitative biomarkers for prediction of epidermal growth factor receptor mutation in non-small cell Lung cancer. *Translational Oncology*, *11*, 94–101.

Zhang, L., Dong, D., Zhang, W., Hao, X., Fang, M., Wang, S., et al. (2020). A deep learning risk prediction model for overall survival in patients with gastric cancer: A multicenter study. *Radiotherapy and Oncology*, *150*, 73–80.

Zhang, Y., & Yang, Q. (2018). An overview of multi-task learning. *National Science Review*, *5*, 30–43.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921–2929).

Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8697–8710).