Medical Image Analysis 73 (2021) 102150

Contents lists available at ScienceDirect

Medical Image Analysis

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

DeepPrognosis: Preoperative prediction of pancreatic cancer survival and surgical margin via comprehensive understanding of dynamic contrast-enhanced CT imaging and tumor-vascular contact parsing

Jiawen Yao[a,1,*], Yu Shi[b,1], Kai Cao[c,1], Le Lu[a], Jianping Lu[c], Qike Song[b], Gang Jin[d], Jing Xiao[e], Yang Hou[b,*], Ling Zhang[a]

[a] PAII Inc., Bethesda, MD 20817, USA
[b] Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, China
[c] Department of Radiology, Changhai Hospital, Shanghai, China
[d] Department of Surgery, Changhai Hospital, Shanghai, China
[e] Ping An Technology, Shenzhen, China

ARTICLE INFO

Article history:
Received 5 February 2021
Revised 8 May 2021
Accepted 24 June 2021
Available online 29 June 2021

Keywords:
Pancreatic ductal adenocarcinoma (PDAC)
3D contrast-enhanced convolutional LSTM (CE-ConvLSTM)
Preoperative survival prediction
Resection margin prediction

ABSTRACT

Pancreatic ductal adenocarcinoma (PDAC) is one of the most lethal cancers and carries a dismal prognosis of ~10% in five year survival rate. Surgery remains the best option of a potential cure for patients who are evaluated to be eligible for initial resection of PDAC. However, outcomes vary significantly even among the resected patients who were the same cancer stage and received similar treatments. Accurate quantitative preoperative prediction of primary resectable PDACs for personalized cancer treatment is thus highly desired. Nevertheless, there are a very few automated methods yet to fully exploit the contrast-enhanced computed tomography (CE-CT) imaging for PDAC prognosis assessment. CE-CT plays a critical role in PDAC staging and resectability evaluation. In this work, we propose a novel deep neural network model for the survival prediction of primary resectable PDAC patients, named as 3D Contrast-Enhanced Convolutional Long Short-Term Memory network (CE-ConvLSTM), which can derive the tumor attenuation signatures or patterns from patient CE-CT imaging studies. Tumor-vascular relationships, which might indicate the resection margin status, have also been proven to hold strong relationships with the overall survival of PDAC patients. To capture such relationships, we propose a self-learning approach for automated pancreas and peripancreatic anatomy segmentation without requiring any annotations on our PDAC datasets. We then employ a multi-task convolutional neural network (CNN) to accomplish both tasks of survival outcome and margin prediction where the network benefits from learning the resection margin related image features to improve the survival prediction. Our presented framework can improve overall survival prediction performances compared with existing state-of-the-art survival analysis approaches. The new staging biomarker integrating both the proposed risk signature and margin prediction has evidently added values to be combined with the current clinical staging system.

1. Introduction

Pancreatic cancer is the third most common cause of cancer deaths in the United States (Siegel et al., 2019). Pancreatic ductal adenocarcinoma (PDAC) is the most common (approximately 95%) pancreatic cancer and has the poorest prognosis among all solid malignancies with a 5-year overall survival (OS) rate of 10% (Grossberg et al., 2020). Surgical resection with a negative resection margin may be a potentially curative treatment option for patients with PDAC (Joo et al., 2019). Offering surgery to those who would most likely benefit (e.g., cured with a high chance of long-term survival) is thus very important for improving patients' life expectancy.

Recently, many machine learning and deep learning methods have been proposed for the preoperative prognosis of various human cancers using computed tomography (CT) or MRI imaging. Radiomics, an emerging technique that converts medical images into hand-crafted radiomic features, has successfully demonstrated its power in glioblastoma brain tumor (Bakas et al., 2017), lung cancer (Aerts et al., 2014) and Head & Neck cancer (Kwan et al., 2018), etc. The handcrafted radiomics approach usually involves manual

* Corresponding authors.
E-mail addresses: yaojiawen076@paii-labs.com (J. Yao), houyang1973@163.com (Y. Hou).
[1] Co-first author. Jiawen Yao, Yu Shi and Kai Cao contributed equally.

https://doi.org/10.1016/j.media.2021.102150
1361-8415/

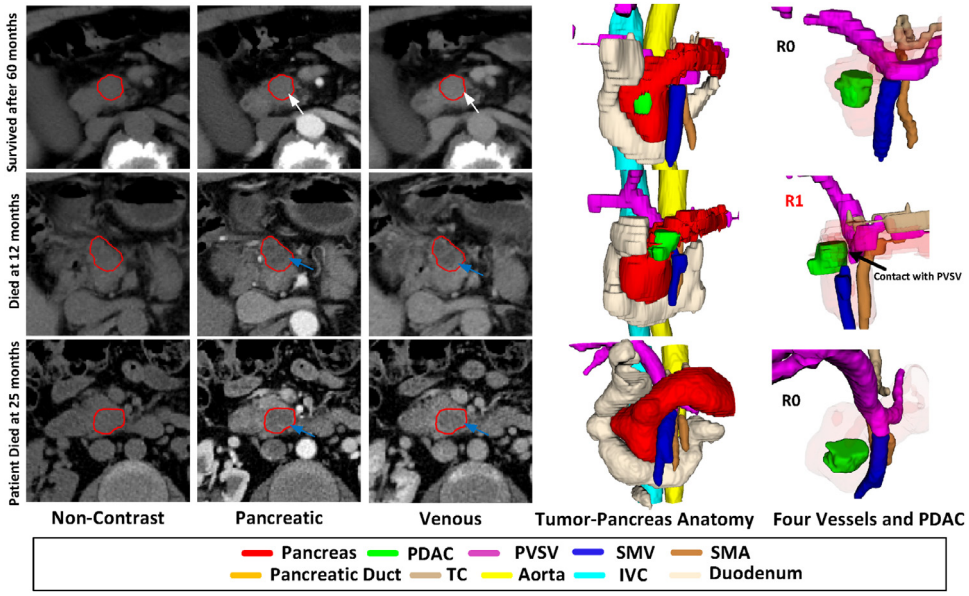

Fig. 1. Example of the multi-phase CE-CT images, anatomy segmentation and PDAC tumor enhancement patterns. The blue arrow depicts a hypo-attenuating tumor; white arrow indicates an iso-attenuating mass. The proposed pancreas and peripancreatic anatomy segmentation has 10-class labels. PVSV is short for "Portal Vein and Splenic Vein", "SMV" for "Superior Mesenteric Vein", "SMA" for "Superior Mesenteric Artery", "TC" for "Truncus Coeliacus", "IVC" for "Inferior Vena Cava". (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

segmentation of the region of interest (e.g., the tumor) on medical imaging, and extraction of hundreds of quantitative features from the ROI, which describe tumor geometry, intensity, and texture characteristics. Machine learning methods are used in the final step to identify or select the imaging features that are associated with a given clinical task. Though hand-crafted radiomics feature-based methods have shown promising results, there have been valid concerns of reproducibility and/or human bias introduced due to the reliance on human feature design, manual segmentation, and variations in imaging and pre-processing techniques for feature extraction (Traverso et al., 2018). Compared with conventional radiomics feature learning, deep learning methods allow for automated learning of clinically relevant radiographic image and geometric features and avoid the requirement on human interventions, thus attracting more and more interest in recent years. 3D convolutional neural network (CNN)-based prognosis models have shown good performances in outcome prediction of lung cancer (Lou et al., 2019; Xu et al., 2019) and gliomas (Nie et al., 2016; Liu et al., 2019). The success of 3D CNNs contributes to capturing the deep features in both the 3D gross tumor volume and peritumoral regions. However, such models may not generalize well for PDAC prognosis because important predictive information has not been effectively exploited from Dynamic Contrast-enhanced computed tomography (DCE-CT) imaging and pancreas and peripancreatic anatomies.

Dynamic Contrast-Enhanced Computed tomography (DCE-CT) remains the primary initial imaging modality of choice for the pancreatic cancer diagnosis. It plays a major role in depicting, staging, patient management (Zhao et al., 2021) and evaluating PDAC resectability (Dickinson et al., 2020). The preoperative multi-phase CE-CT pancreatic imaging used in this study have been scanned at three time points. After the non-contrast phase, average imaging time delays are 40–50 s for the pancreatic phase and 65–70 s for the portal venous phase. Fig. 1 shows three examples to illustrate different tumor attenuation patterns and resection margins of patients with PDAC. Tumor attenuation visual patterns in specific CT phases are very important characteristics to identify and detect the tumor. Each row in Fig. 1 represents one PDAC patient, and red boundaries are the tumor annotations. The blue arrow indicates a typical hypo-attenuating tumor, while the white arrow depicts an iso-attenuating tumor. Besides the channel of tumor attenuation, another very critical factor is the resection margin indicating the margin of apparently non-tumorous tissue around a tumor that has been surgically removed. More specifically, the resection margin is characterized as R0 (microscopically margin-negative) when no evidence of malignant glands was identified microscopically at the primary tumour site. R1 (margin-positive) resections have malignant glands infiltrating at least one of the resection margins on the permanent section (Konstantinidis et al., 2013). Although the margin status is only available via microscopic pathology after the surgery is conducted, there still have cues appearing in CT images to facilitate potential preoperative predictions of margin status. For example, when looking at the PDAC and surrounding vessels of the middle patient in Fig. 1, we can observe that the tumor has the contact with the portal vein and splenic vein (PVSV) and the final resection margin is positive (R1) for this patient. Nevertheless no obvious tumor-vascular contacts can be depicted in the first and the third patients that have R0 status. Examples of these patients show that it is possible or feasible to predict the post-operative resection margin from preoperative CT imaging.

Both tumor attenuation and resection margin are associated with patients' clinical outcomes. In previous studies, Kim et al. reported that visually isoattenuating PDACs are associated with better survival rates after surgery, as opposed to typical hypo-attenuating PDACs (Kim et al., 2010). Surgeries resulting in an R0 resection usually associate with relatively long-term survival. In contrast, an R1 resection may have a high cancer recurrence chance, and thus patients suffer worse clinical outcomes. Specifically, according to one recent study, R0 has a median OS of 22 months versus R1 of 15 months (Tummers et al., 2019). Hypo-attenuating mass can be clearly observed in both pancreatic and venous phases of the second and third patients, indicating low stromal fractions (worse clinical outcomes). The first patient in Fig. 1 reflects both isoattenuating in the pancreatic and venous phase compared with its adjacent pancreas regions, indicating high stromal fractions (better survival). Between the second and third patients (Fig. 1), though hypo-attenuating could be seen in both patients, the second patient undergoes a R1 resection, and we could also see the higher degree of tumor-vascular contact by ob-

serving the PDAC and surrounding anatomies (four vessels and pancreas). Survival information shows that this patient died at 12 months, which has a worse outcome than 25 months of the third patient. Fig. 1 illustrates that both tumor enhancement patterns across phases and tumor-vascular contact are very useful imaging features to reflect tumor's pathological heterogeneity, location, and vascular contact for building a more accurate prognosis model.

1.1. Main contributions

Though several studies can be found on survival prediction using deep learning, to the best of our knowledge, this is the first deep learning-based prognosis approach for directly and efficiently using dynamic multi-phase CT imaging for predicting OS and resection margin of patients with PDAC. Clinically, the deep learning output is analyzed in conjunction with established clinical prognostic factors (e.g., pTNM staging, pathological tumor size, CA19-9, and stromal fraction etc.) to confirm that it is an independent risk factor. The major contributions of our work can be summarized as follows.

- PDAC tumor attenuation across phases is associated with OS. To capture this cue, we propose a novel 3D Contrast-Enhanced Convolutional Long Short-Term Memory (CE-ConvLSTM) deep network to learn the enhancement dynamics of tumor attenuation from multi-phase CE-CT images. This model can capture the tumor's temporal changes across several phases more effectively than the early fusion of input images.
- Tumor-vascular invasion and tumor-pancreas image contrast are associated with resection margin and OS. To depict these channels, we exploit a self-learning approach for automated segmentation of pancreas and peripancreatic anatomies (including pancreas, pancreatic duct, PVSV, superior mesenteric vein [SMV], superior mesenteric artery [SMA], and truncus coeliacus [TC]), without requiring any manual annotations of our PDAC dataset.
- To make the prediction of OS and resection margin benefit from each other, we present a multi-task learning framework to conduct a joint prediction. The jointly learning of cancer risk and resectability in a multi-task setting can derive more effective and comprehensive prognosis related deep image features and subsequently improve the prediction accuracies for both tasks.
- Extensive evaluation and statistical analysis verify the effectiveness of our presented framework. The signature built from the proposed model remains statistically strong in the multivariable analysis adjusting for established clinical predictors, and has the potential be be combined with the established clinical factors for risk stratification and treatment decisions of patients with PDAC.

Our preliminary work that uses deep multi-task learning for the same prediction tasks has been reported in MICCAI 2020 (Yao et al., 2020a). The current study provides more sophisticated/elaborated methodology development and significantly more comprehensive evaluations. Specifically, we provide a new segmentation method of the pancreas and peripancreatic anatomies using a self-learning framework (Zhang et al., 2020). The full anatomy structures, for the first time, are incorporated into a more complete multi-task deep prognosis model, allowing a more comprehensive understanding of PDAC prognosis problem. For instance, PDAC tumor-vascular contact not only determines the resectability (Hong et al., 2018; Grossberg et al., 2020; Mizrahi et al., 2020) but also could be OS predictors (Dickinson et al., 2020). Image contrast between the tumor and surrounding pancreas parenchyma (excluding the pancreatic duct) reflects molecular and pathological heterogeneity of PDAC – it may be used to stratify patients into distinct subtypes, including longer-shorter survival time (Koay et al., 2018; Cai et al., 2020). Notably better experimental results have been achieved than the previous work (Yao et al., 2020a) and traditional radiomics approaches, which do not incorporate the full anatomy information. In addition, we provide a more thorough evaluation and statistical analysis in a larger patient cohort (n=296) by enrolling 91 more patients than our preliminary work (Yao et al., 2020a).

2. Related work

Computed tomography (CT) is the most commonly adopted imaging modality to detect, stage, and evaluate human cancers, including pancreatic cancers. During recent years, many methods have been proposed for survival prediction using CT images, and they can be categorized into two categories: traditional radiomics-based models and deep learning-based methods.

2.1. Radiomics approaches

Radiomics features have demonstrated the potential to describe underlying tumor biology and thus are used to associate with various diagnosis and prognosis tasks across many cancer types (Aerts et al., 2014; Gillies et al., 2016). Radiomics approaches convert imaging data into high-dimensional measurable and predefined features using texture analysis, which captures the spatial variations in pixel intensities within a tumor. Those features usually include signal intensity, shape, texture, and higher-order texture features. Signal intensity features are calculated based on histograms of individual voxel intensities. Shape features are designed from the 2D/3D geometry of the target (e.g., tumor). Texture features are computed in either 2D or 3D to consider the spatial relationships of attenuation of neighboring pixels or voxels. Higher-order texture features include statistics by adding a filtration process like wavelet filter before feature extraction (Lubner et al., 2017). The radiomics feature extraction process typically generates hundreds of hand-crafted features, followed by dimension reduction and feature selection to be performed to identify the most relevant features.

One of the most important clinical-relevant applications of radiomics is the ability to help predict patient survival of cancer patients in a preoperative setting, especially for patients with PDAC (Chu et al., 2020). Different from other solid tumor type cancers, surgical resection at present remains the only cure for patients with PDAC. However, pancreatic resection can cause major morbidity and poses a risk of surgical mortality. Valid and effective preoperative risk models would be very useful for patients who will benefit the most from pancreatic resection. In multiple research work, radiomics features were used to predict overall survival and disease-free survival of resectable PDACs (Cassinotto et al., 2017; Yun et al., 2018; Eilaghi et al., 2017; Attiyeh et al., 2018). Cassinotto et al. enrolled 99 PDAC patients with portal venous phase CT and extracted histogram features of the largest tumor slice (Cassinotto et al., 2017). They found tumor hypoattenuation is associated with higher tumor grade, greater lymph node invasion, and shorter disease-free survival. Another study extracted radiomics features from all tumor slices from portal venous phase CT and found heterogeneously hypo-attenuating tumors associated with poor overall survival (Attiyeh et al., 2018). Though some reports concluded that heterogeneous tumors are associated with worse survivals (Cassinotto et al., 2017; Attiyeh et al., 2018); others found that more homogeneous tumors are associated with poor survival (Yun et al., 2018) which contradicts the previous findings. The apparent difference may come from both the variability in patient selection and the human bias introduced into the radiomics process by using 2D manual selected tumor slices and human pre-defined features. Recently, traditional radiomics approaches are

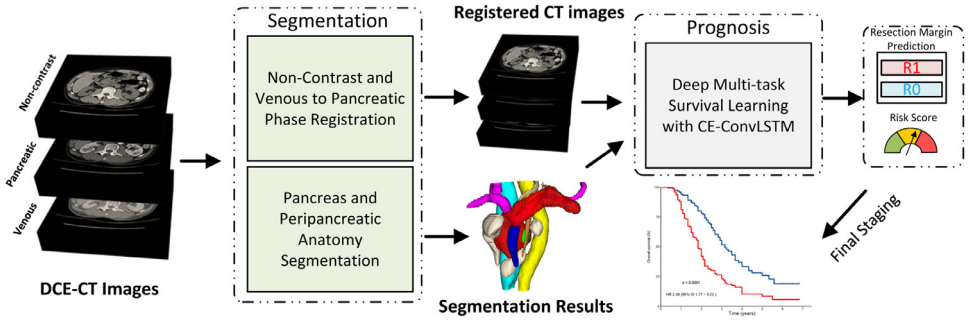

Fig. 2. The overall workflow of our DeepPrognosis system for PDAC patients prognosis.

criticized and challenged by their lack of reproducibility and interpretability as well as over-fitting on small datasets (Traverso et al., 2018).

2.2. Deep learning survival models

Recently, with the advance of deep neural networks, deep learning-based approaches have shown good performances not only in medical imaging diagnosis (Wang et al., 2017; 2019; Zhang et al., 2017), but also in deep survival models for seeking more powerful deep representations (Katzman et al., 2016) based on various imaging modalities, including pathology (Yao et al., 2017; 2019; Zhu et al., 2016; 2017; Wulczyn et al., 2021), radiographic images (Jiang et al., 2020; Xu et al., 2019) and PET imaging (Cheng et al., 2021). In computational digital pathology, many deep learning survival models with pathological slides (Yao et al., 2020b; Mobadersany et al., 2018; Skrede et al., 2020; Wulczyn et al., 2020), have been proven successful in addressing both various cancer types and the specific colorectal cancer from the real large-scale clinical environments (Skrede et al., 2020). Different from pathological slides that are only available through operative biopsy procedures, dynamic contrast-enhanced CT serves as the frontline imaging modality for staging and evaluating cancer. The prediction model with CT images may be more important and convenient for patients with PDAC as most of their pathological results can only be obtained after surgery.

Many studies have been proposed on various deep learning survival models in CT images for glioblastoma and lung cancer patients (Tang et al., 2020; Lou et al., 2019; Xu et al., 2019; Kim et al., 2020). Glioblastoma (GBM) is the most common malignant brain tumor, and many work designed models and reported their survival prediction of GBM patients. In (Jungo et al., 2017), radiomics features extracted from brain tumors using multimodal MRI images are fed to train an artificial neural network for OS (Overall Survival) prediction. To automatically learn OS-related deep MRI features, several deep learning-based OS prediction methods (Nie et al., 2016; 2019) are presented where a multi-channel CNN is proposed. A multi-task model which conducts a joint prediction of both tumor genotype and OS time is proposed for much improved OS prediction accuracy (Tang et al., 2020). A similar multi-task deep learning model can also be found for individualizing radiotherapy dose for lung cancer patients (Lou et al., 2019).

Though deep learning models are successfully applied for automated learning of relevant radiographic information without the need for manual definitions, recent models are still not suitable to handle pancreatic cancer imaging protocol where multi-phase dynamic contrast-enhanced CT is utilized.

2.3. Self-learning for medical image segmentation

Image segmentation is a fundamental problem in medical image analysis. UNet-based approaches have shown robust image segmentation accuracy in many medical applications (Isensee et al., 2021), when a training dataset with plenty quantities of pixel-level fully-annotated images is available. However, it is usually infeasible and inconvenient to construct a large, well-organized, and volumetric-annotated medical image dataset. Also, recent PDAC (Zhu et al., 2018; 2019) and pancreatic duct segmentation (Wang et al., 2020) used a fully annotated dataset to train the model. Self-learning assumes that a deep model (student) trained from noisy annotations (teacher) has the potential to surpass the teacher (Guan et al., 2018; Khoreva et al., 2017; Zhang et al., 2018). Recent work find new effective strategies to improve the student performance further, including adding regularization (Roth et al., 2019) or noises (Xie et al., 2020) to perturb the noisy annotations and generate noisy but informative annotations on a large unannotated external dataset (Xie et al., 2020). Most recently, we adapt the self-learning framework for improving tumor segmentation performance in the scenario of multi-institutional multi-phase partially-annotated CT scans being available (Zhang et al., 2020). In this paper, we adopt and extend a similar framework for the pancreas and peripancreatic anatomy segmentation.

A pre-trained segmentation model and deep learning survival model could be integrated into an end-to-end framework. However, this might be challenging in a self-learning setting, in which the lack of ground truth annotations cannot ensure reliable joint learning. Moreover, the optimal image intensity normalization methods might be different for the segmentation and prediction tasks. In clinical studies, the manual segmentation is mostly performed blind to the patient outcome information. Previous deep learning-based prognosis studies (Tang et al., 2020; Lou et al., 2019; Xu et al., 2019; Kim et al., 2020) are also conducted with a manual mask available.

3. Methodology

The overall workflow of our DeepPrognosis is illustrated in Fig. 2, which consists of (1) pancreas and peripancreatic anatomy segmentation, (2) deep multi-task model with contrast-enhanced (CE) ConvLSTM for OS and resection margin prediction, and (3) building the final PDAC staging from risk scores and margin predictions.

3.1. Pancreas and peripancreatic anatomy segmentation

Fig. 3 depicts our proposed self-learning framework for the pancreas and peripancreatic anatomy segmentation using multi-institutional multi-phase partially annotated CT scans. Given the self-collected PDAC multi-phase CT datasets from two (i.e., A and

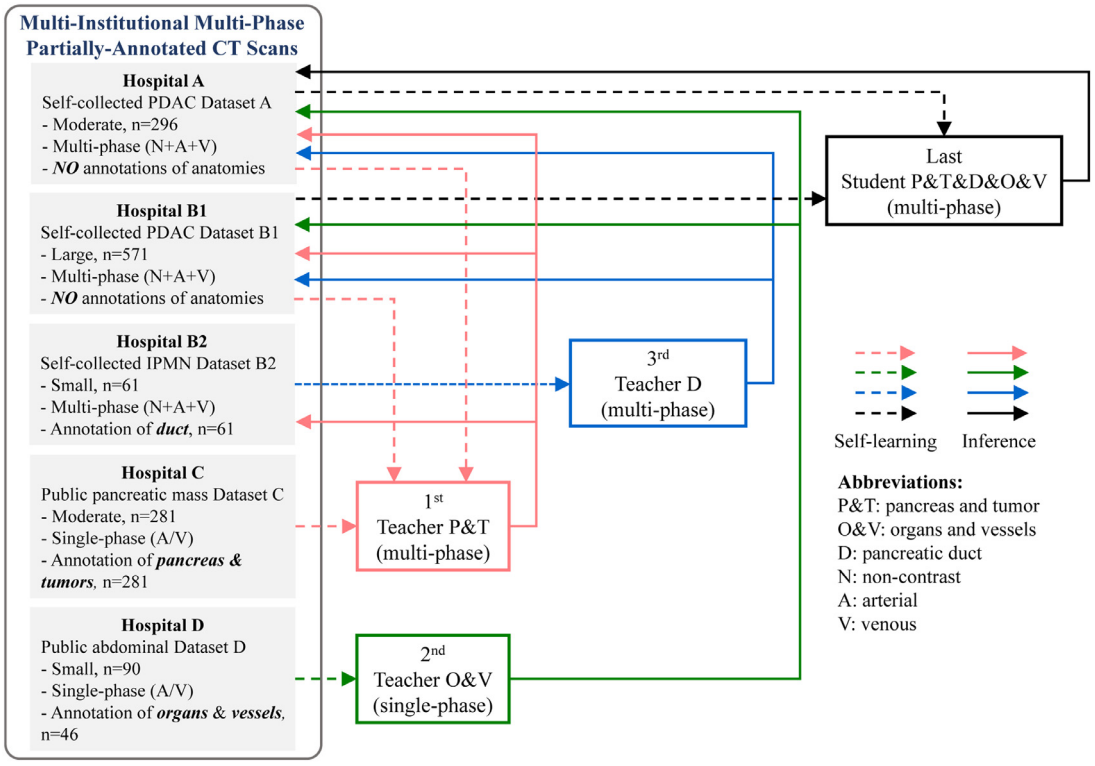

Fig. 3. The proposed self-learning framework for segmentation of pancreas and peripancreatic anatomies without requiring manual annotations of self-collected datasets.

B1) hospitals without any annotation of anatomies, our method can effectively incorporate and utilize other datasets (i.e., hospital B2, C, D) and annotations to segment pancreas and peripancreatic anatomies, described as follows.

In details, *firstly*, we train a multi-phase Pancreas and Tumor (denoted as Teacher P&T) segmentation model. Dataset A and B1 are two self-collected, large-scale (n =867), multi-phase CT PDAC Datasets without annotations of pancreas anatomies. Dataset C (Simpson et al., 2019) is a public venous phase CT dataset including manual pixel-level annotations of pancreas and tumor. It is used to train a (five-fold) ensemble segmentation model, which is used to generate pseudo annotations of the pancreas and tumor in the venous phase CT images in Dataset A and B1. Then teacher P&T is trained on the registered non-contrast, arterial and venous phases CT scans in Dataset A and B1 with those generated pseudo annotations.

Secondly, a single-phase Organ and Vessel (denoted as Teacher O&V) segmentation model is trained using the same process as in Zhang et al. (2020). Briefly, Dataset D is a public dataset (Gibson et al., 2018), including 90 patient abdominal CT volumes with partially annotated organs and vessels up to 14 classes. We complete annotations of 46 CTs for 17 classes, including three additional peripancreatic vessel classes (SMA, SMV, and TC) under the supervision of a board-certified radiologist. Then the self-learning method is used to train the Teacher O&V on all 90 CT scans with 17 classes of annotations.

Thirdly, we train a multi-phase pancreatic Duct (denoted as Teacher D) segmentation model. Dataset B2 is a self-collected, multi-phase CT, IPMN (intraductal papillary mucinous neoplasms) datasets with radiologist-annotated IPMN. IPMNs are mucin-producing cysts and may spatially involve the main duct, branch duct, or a combination of both (Dalal et al., 2020). Teacher D is trained on the registered multi-phase CT images with the pancreas and IPMN/duct annotations. Note that the pancreas annotations used in the model training are generated by Teacher P&T output segmentation on this dataset (B2).

Fourthly, the above three Teachers are respectively applied on multi-phase CT scans in Dataset A and B1 to infer corresponding pseudo annotations. More specifically, the pancreas&tumor, pancreatic duct, and organ&vessels pseudo annotations are generated by Teacher P&T, Teacher D, and Teacher O&V, respectively. Since Teacher O&V is learned to segment single-phase (either arterial or venous) CT, we first apply it to segment the arterial and venous phase CT images in Dataset A and B1. Then the two resulting vessel masks are combined by trusting the artery masks (i.g., TC, SMA, and aorta) in the arterial phase and vein masks (i.g., PVSV, SMV, and inferior vena cava) in the venous phase. The duodenum masks generated on the venous phase are used as the duodenum pseudo annotations, as the duodenum is better visible on the venous phase. Note that all teachers can segment the pancreas. We use the pancreas masks segmented by Teacher P&T since it is self-learned on the multi-phase CT images with the largest data size. In addition, vessel masks are used to overlay the pancreas masks. Because the former segmentation task is easier than the latter in general, Teacher P&T is not learned to distinguish the pancreas and vessels. As such, teachers with different specialties behave collaboratively like an ensemble when generating the pseudo annotations on the large datasets (n=867) that demonstrate more anatomy variations, allowing the student to take advantage of different teachers and learn beyond teachers effectively to be capable of segmenting more challenging images desirably.

Last, we train a Student model on both Datasets A and B1 with pseudo annotations (generated as above) of ten classes in total, i.e., pancreas, tumor, duct, aorta, inferior vena cava, duodenum, TC, SMA, PVS, and SMV. Learning multiple classes in a single model allows the student to implicitly capture the spatial context information, further improving segmentation accuracy for each class compared to the teachers. Note that some organs and vessels are excluded (e.g., liver, spleen, and esophagus, etc.) from the student's

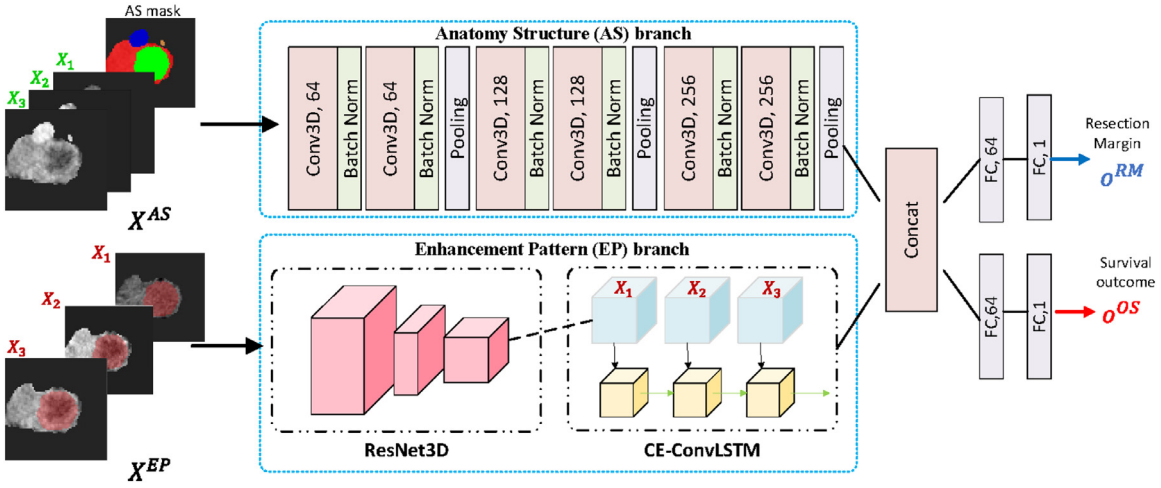

Fig. 4. An overview of the proposed multi-task model with CE-ConvLSTM.

learning process since they are not directly related to PDAC OS prediction. Finally, the trained Student model is applied to Dataset A to generate masks used in the prognosis model (Section 3.2). Our overall self-learning training framework utilizing five datasets with different levels of annotations and patient distributions is shown in Fig. 3.

For the multi-phase CT imaging registration as preprocessing, we use DEEDS (Heinrich et al., 2013), which performs the best in a recent evaluation of abdominal CT imaging registration algorithms (Xu et al., 2016). For the training of our segmentation models mentioned above, we use the nnUNet backbone (Isensee et al., 2021) due to its high accuracy on several medical image segmentation tasks, such as abdominal organs, vessels, and tumors (Isensee et al., 2021). 3D UNet working with full image resolution is used as the network architecture. A combination of Dice and cross-entropy loss is utilized. We train the models to optimize the loss of all given classes. The model that produces the best Dice score on the validation set is selected as the best segmentation model.

3.2. Deep multi-task model with CE-ConvLSTM

We use multiple time points of 1, 2, 3 to represent non-contrast, pancreatic, and venous phases, respectively. The dataset is prepared for every tumor volume from each phase scan, to build the 4D CE-CT data format for anatomy structure (AS) and enhancement pattern (EP) branch (as $X_t^b, t \in \{1,2,3\}$, b $\in$ {AS, EP}). After self-learn segmentation, we now have pancreas anatomy segmentation models and results. A multi-phase sequence of image sub-volumes of $64 \times 64 \times 64$ pixels^3 centered at the tumor 3D centroid is cropped to cover the entire tumor and its surrounding vessels and pancreas. Our joint learning network architecture is shown in Fig. 4. This network is designed for predicting both the resection margins and survival outcomes.

The branch of Anatomy Structure (AS) uses one 3D-CNN model with six convolutional layers equipped with Batch Normalization and ReLu. Similar 3D architecture has shown good prediction performance for lung cancer (Lou et al., 2019). To consider the tumor-vascular involvement, we include segmentation results of four vessels (i.e., PVSV, SMV, SMA, and TC), pancreas, and PDAC. The input of this branch is the concatenation of CT volumes at different time points and the corresponding tumor, pancreas, and vessels masks: $X_t^{AS} \in \mathbf{R}^{4\times 64^3}$. This branch will attempt to learn the CT intensity attenuation variations and the relationships between tumor and surrounding pancreas regions and vessels, which help classify the tumor into different resection status. Note that R0/R1 can only be obtained after the surgery and pathology. Our model can be applied preoperatively in real scenarios to offer patients with PDAC the appropriate advice regarding surgical decisions.

To consider tumor attenuation, the branch of Enhancement Pattern (EP) uses CT volume at each phase (each phase is CT-*M* two-channel input, $X_t^{EP} \in \mathbf{R}^{2\times 64^3}$). We crop 3D Volume of Interest (VOI) using PDAC, pancreas, and pancreatic duct mask. Then PDAC mask is used as the one addition channel of inputs in this branch (PDAC mask shown on original CT images X^{EP} in Fig. 4). Tumor attenuation usually means the contrast differences between the tumor and its surrounding pancreas tissues so that this branch can help capture the tumor attenuation patterns across phases. The core part of this branch is a recurrence module that allows the network to retain what it has seen and to update the memory when it observes a new phase image. A naive approach is to use a vanilla LSTM or ConvLSTM network. Conventional ConvLSTM is capable of modeling 2D spatio-temporal image sequences by explicitly encoding the 2D spatial structures into the temporal domain (Chen et al., 2016). A more recent ST-ConvLSTM simultaneously learns both the spatial consistency among successive image slices and the temporal dynamics across different time points for the tumor growth prediction (Zhang et al., 2019). Instead of using adjacent 2D CT slices and motivated by 3D object reconstruction (Choy et al., 2016), we propose to use a ResNet3D-based Contrast-Enhanced 3D Convolutional LSTM (CE-ConvLSTM) network to capture the temporally-enhanced imaging patterns from CE-CT sequences. CE-ConvLSTM can model 4D spatio-temporal CE-CT sequences by explicitly encoding and projecting their 3D spatial structures into the temporal domain.

The main equations of ConvLSTM are as follows:

$$
\begin{aligned}
f_t &= \sigma\left(W_f^X * X_t^{EP} + W_f^H * H_{t-1} + b_f\right) \\
i_t &= \sigma\left(W_i^X * X_t^{EP} + W_i^H * H_{t-1} + b_i\right) \\
o_t &= \sigma\left(W_o^X * X_t^{EP} + W_o^H * H_{t-1} + b_o\right) \\
C_t &= f_t \odot C_{t-1} + i_t \odot tanh\left(W_C^X * X_t^{EP} + W_C^H * H_{t-1} + b_C\right) \\
H_t &= o_t \odot tanh(C_t)
\end{aligned}
\tag{1}
$$

where X_t^{EP} is the CE-CT sequences at time t, $*$ denotes the convolution operation, and $\odot$ denotes the Hadamard product. All the gates f, i, o, memory cell C, hidden state H are 4D tensors. We use $3 \times 3 \times 3$ convolutional kernels and 128 as the channel dimension of hidden states for the LSTM unit. We employ 3D-ResNet18 (Hara et al., 2018; Chen et al., 2019) as the encoder to encode each three-channel input to the lower-dimensional feature maps for CE-ConvLSTM.

After the concatenation of feature maps from both tasks, the channel number of this common representation is 256. Then two separate fully-connected networks will use the common representation for each prediction task. In the training phase, labels of the resection status **y** and patient overall survival information (OS time t_{OS} and censoring status δ) are known for each input CE-CT sequence. Suppose the output from each fully connected network for resection margin and overall survival prediction task is denoted as o^{RM} and o^{OS}, respectively. Since most patients have R0 resection margins, and the imbalanced class exists in practices. Therefore, the weighted binary cross-entropy (BCE) loss is applied to the resection margin prediction task,

$$L_{RM} = -\frac{1}{N}\sum_{i}^{N} w_0 y_i \cdot \log o_i^{RM} + w_1(1-y_i)\cdot log(1-o_i^{RM}), \tag{2}$$

The negative log partial likelihood ((Cheng et al., 2021)(Katzman, Shaham, Cloninger, Bates, Jiang, Kluger)) which allows the usage of information from censored data is employed as the survival loss to predict the survival outcome o^{OS} of this patient. This survival loss is summarized as

$$L_{OS} = \sum_{i} \delta_i(-o_i^{OS} + \log \sum_{j:t_j>=t_i} \exp(o_j^{OS})). \tag{3}$$

where j is from the set whose survival time is equal or larger than t_i ($t_j \geq t_i$) and δ is 1 for death happened while 0 for censored. The final training loss is $L = L_{OS} + \lambda L_{RM}$ and λ is set to 0.5.

3.3. Final PDAC staging

The model is valid to provide continuous risk scores to indicate patients' OS and predict their resection margin status. For clinical use and patients risk stratification needed in clinical practice, we then build a cancer staging system based on risk scores and predicted resection margin. When applying the trained models on the testing set, we could get the testing risk score and resection margin prediction for each patient. Let m_i^{tr} be the median score on training patients, we denote risk staging by using the median training score as the cutoff. For the i-th patient in testing set, its risk staging is defined as $S_i^R = 1$ if $o_i^{OS} \geq m_i^{tr}$ otherwise $S_i^R = 0$. As we know that patients with margin positive (R1) resection usually are associated with poor clinical outcomes, we can identify and allocate the predicted R1 patients into the high-risk group. Therefore, we then build the final staging by considering both risk staging and resection margin staging as below

$$S_i^{Final} = \begin{cases} 0 & \text{if } S_i^R = 0 \text{ and } S_i^{RM} = 0, \\ 1, & \text{if } S_i^R = 1 \text{ or } S_i^{RM} = 1 \end{cases} \tag{4}$$

This staging system can be used to not only assess individual patient's risk but also may help guide the treatment decisions for personalized medicine.

4. Experiments

4.1. Dataset description

Datasets from four hospitals (1,209 patients with pancreatic tumor/cyst) are used in this work. **Dataset A**, including 296 patients with pathologically confirmed PDACs (the median tumor size is 2.5 cm), is collected from Shengjing Hospital of China Medical University (SHCMU) with non-contrast, pancreatic (late-arterial), and venous phases of CT scans. Such a multi-phase CT imaging setting is the standardized protocol for depiction, staging, and resectability evaluations of PDAC, specified in the National Comprehensive Cancer Network (NCCN) guidelines (NCCN-PDAC, 2020). The median imaging spacing is 0.70 × 0.70 × 3mm in [X,Y,Z].

Table 1
Demographic and tumor characteristics in Data A. Median [interquartile range, 25th–75th percentile] values are reported for continuous variables.

Characteristics	Dataset A for prognosis (n=296)
Sex, n (%)	
Female/Male	131(44%)/165(56%)
Age at diagnosis, yrs	60(53–66)
pT Stage, n (%)	
pT1/pT2	62(21%)/199(67%)
pT3/pT4	18(6%)/17(6%)
pN Stage, n (%)	
pN0/pN1/pN2	199(67%)/78(26%)/19(7%)
pTNM Stage, n (%)	
I/II/III	179(61%)/84(28%)/33(11%)
Tumor Size, cm	2.5 (2.1–3.0)
Tumor location, n (%)	
head	134(45%)
uncinate	141(48%)
body or tail	21(7%)
Resection Margin, n (%)	
R0/R1	258(87%)/38(13%)
CA 19-9 (U/mL)	
≤ 210/> 210/Missing	155(52%)/136(46%)/5(2%)
Adjuvant therapy, n (%)	
No adjuvant	22(7%)
Chemotherapy	117(40%)
Chemoradiotherapy	157(53%)

PDAC tumors are manually traced and annotated on the pancreatic phase slice by slice by a radiologist (YS) with 16 years of experience in pancreatic imaging. **Dataset B1** and **Dataset B2**, including 571 patients with pathologically confirmed PDACs (no manual annotations) and 61 patients with pathologically confirmed IPMNs (manual annotations are performed by a board-certified radiologist (KC) with 14 years of specialized experience in pancreatic imaging), respectively, are collected from Changhai Hospital with non-contrast, early-arterial/pancreatic, and venous phase CTs used in this work. The median imaging spacing is 0.68 × 0.68 × 3mm. **Dataset C**, including 281 patients with pancreatic tumor annotations, is a public dataset provided by Memorial Sloan Kettering Cancer Center (Simpson et al., 2019). The median imaging spacing is 0.80 × 0.80 × 2.5mm. **Dataset D** is a combination of two public datasets (described in (Gibson et al., 2018)), including abdominal CT scans of 90 patients with 17 classes of pixel-level organ and vessel annotations. More details about the annotation process are described in our previous work (Zhang et al., 2020).

Among these datasets, only **Dataset A** (n=296) has complete data to perform OS prediction, including PDAC baseline CT imaging, OS time and status, preoperative clinical as well as postoperative pathology (e.g., resection margin status) information. Surgical procedures were performed by experienced surgeons at SHCMU – a high-volume pancreatic cancer institution where extended amounts of pancreatic resections are commonly performed. Dataset A has been conducted the quality control to exclude patients with stents which could cause imaging artifacts to influence annotations. Therefore, Dataset A is used in the prediction and prognosis experiments. All information including adjuvant therapy of Data A can be seen below in Table 1. All five datasets are used to train the pancreas and peripancreatic anatomy segmentation model.

4.2. Evaluating of segmentation performance

4.2.1. Implementation details

For the nnUNet training (Isensee et al., 2021), most parameters are set by default. The original 3D scans are resampled to the median spacing of the training data. The network inputs are 3D subvolumes. Multi-phase registered CT images are directly concatenated as input channels to feed the network. The input-level

Table 2
Performance comparison of teacher and student models for vessel segmentation. Bold indicates better results. TC: truncus coeliacus; SMA: superior mesenteric artery; SMV: superior mesenteric vein; PVSV: portal vein and splenic vein. O&V: organs and vessels.

Vessels	Metrics	Teacher O&V	Student
TC	Dice	0.77±0.13	**0.81±0.12**
	ASD	0.70±0.85	0.66±1.16
	HD	9.38±8.88	**9.21±7.28**
SMA	Dice	0.71±0.19	**0.74±0.17**
	ASD	2.16±4.62	**1.72±3.04**
	HD	11.74±13.23	**11.66±12.09**
SMV	Dice	0.81±0.09	**0.84±0.05**
	ASD	0.85±0.51	**0.60±0.30**
	HD	7.83±3.34	**6.62±2.93**
PVSV	Dice	0.77±0.08	**0.81±0.06**
	ASD	1.56±1.28	**1.10±1.01**
	HD	20.74±17.29	**15.83±15.32**

fusion is widely adopted in the multi-modality tumor segmentation tasks (Zhou et al., 2019). The Teacher models O&V and P&T and Student model are trained with 200 epochs (250 batches per epoch). Teacher D is trained with 1000 epochs. The training process is taken on a NVIDIA Titan RTX-6000 GPU. For training these models, a random training-validation (80%-20%) splitting is used.

4.2.2. Evaluation methods and results

The pancreas segmentation in CT images is a relatively well-studied problem. nnUNet produces leading performance on this task (Simpson et al., 2019). Therefore, we report the segmentation performance of IPMN/duct and four peripancreatic vessels (i.e., TC, SMA, SMV, and PVSV). On the validation set in Dataset B2 with manual annotations as ground truth, the Dice score of IPMN/duct segmentation is 0.73. As references, the Dice scores of pancreatic duct segmentation are 0.62-0.64 in recent studies (Wang et al., 2020; Xia et al., 2020). However, we acknowledge that these numbers are not directly comparable since different datasets are used – IPMN cases in ours and PDAC cases in the reference work.

To assess the vessel segmentation performance, we invite a medical student (QS) in pancreatic imaging to perform manual segmentation under the supervision of a radiologist (YS). Nineteen cases with 9 R0 and 10 R1 (four pT1, eight pT2, three pT3, and four pT4) are randomly selected for the assessment. The manual segmentation of TC and SMA is performed on the pancreatic phase, while SMV and PVSV on the venous phase. Dice score (ranging between 0 and 1), average surface distance (ASD, mm), and Hausdorff distance (HD, mm) are used as metrics. The Teacher O&V (Zhang et al., 2020) which is trained on the public dataset (Gibson et al., 2018) is used for comparison. Note that the evaluation only considers the vessels near the pancreas region (defined as 5 mm larger than the pancreas bounding-box in each direction in this work), as they are meaningful and clinically relevant to PDAC prognosis. Results are shown in Table 2. The student model outperforms its teacher for all four vessel classes. An illustrative vessel segmentation example is shown in Fig. 5.

4.3. Evaluation of prognosis performance

4.3.1. Implementation details and metrics

We adopt the nested 4-fold cross-validation (with training, validation, and testing sets in each fold) to evaluate our prognosis model and other competing methods. All CT images were resampled to an isotropic 1 mm^3 resolution. During training, we did the following data augmentations, (1) rotating the volumetric tumors in the axial direction around the tumor center with the step size of 90° to get the corresponding 3D CT image patches and their mirrored patches, (2) selecting the cropped regions with random shifts for each iteration during the training process, 3) applying random Gaussian noise on 3D CT image patches. The data augmentation can improve the network's ability to locate the desired translational invariants. The batch sizes of our method and other models are set the same, which are 8. For training, we use Adam optimization with a weight decay 5×10^{-4}. The learning rate is set to 10^{-4}. The training process monitors the loss on the validation set, and it will early stop if the loss goes increased noticeably. The maximum iteration is set to be 500 epochs, and the model with the best performance on the validation set during training is selected for testing.

For the prognosis evaluation, we choose the Harrell's concordance index (C-index) (Harrell et al., 1982). The C-index quantifies the ranking quality of rankings whether the predicted survival times are ranked in the same order as their true survival time. It is calculated as follows

$$c = \frac{1}{n} \sum_{i \in \{1 \ldots N | \delta_i = 1\}} \sum_{t_j > t_i} I[f_i > f_j] \tag{5}$$

where n is the number of comparable pairs and $I[.]$ is the indicator function. $t.$ is the actual time observation. $f.$ denotes the corresponding risk. The value of C-index ranges from 0 to 1 where the larger the value is, the better the model predicts.

In the ablation study, we first validate deep survival prediction performance under different imaging modality protocols, including the pancreatic phase only, venous phase only, and all three phases together with early fusion. ResNet3D-18 with the pretrained weights (Chen et al., 2019) is used as an advanced model compared to the conventional 3D ConvNets. Radiomics signature from single-phase and radiomic nomogram for multi-phase CT images are also evaluated for ease of comparison.

4.3.2. Other competing methods

Multi-task models. To evaluate the effectiveness of the multi-task model, we compare the proposed model with other deep multi-task models. The first one is our preliminary work (Yao et al., 2020a) that does not have the pancreas anatomy segmentation, and neither considers the tumor-vascular involvement. Tang et al. propose using separate branches of 3D CNNs to predict both OS time and tumor genotype for glioblastoma (GBM) patients (Tang et al., 2020). Lou et al. present a multi-task training model on the shared hidden representations from a single model (Lou et al., 2019). We replace the RMSE loss from the original implementation of Tang et al. (2020) with the negative log partial likelihood loss because it can handle alive patients (whereas authors discarded some patients who are still alive in Tang et al. (2020)). For other baselines (Tang et al., 2020; Lou et al., 2019), inputs include not only PDAC but also surrounding vessels and pancreas tissues to let the model have tumor-vascular information.

Radiomics signature and nomogram. To compare with traditional radiomics methods, we built radiomics signature of each CT phase and then created radiomic nomogram from a multiphase signature. For each phase (non-contrast, pancreatic, and venous), we constructed the signature in the following. First, we extracted features using an open-source Python package, Pyradiomics (Van Griethuysen et al., 2017)[2], from 3D tumor regions. There are 482 radiomics features in total, which can be divided into four groups: 1) intensity, 2) geometry, 3) texture, and 4) wavelet features.

- The intensity features quantified the first-order statistical distribution of the voxel intensities within the volumes of interest (tumor or lymph nodes). The statistical measurements include

[2] https://pyradiomics.readthedocs.io/

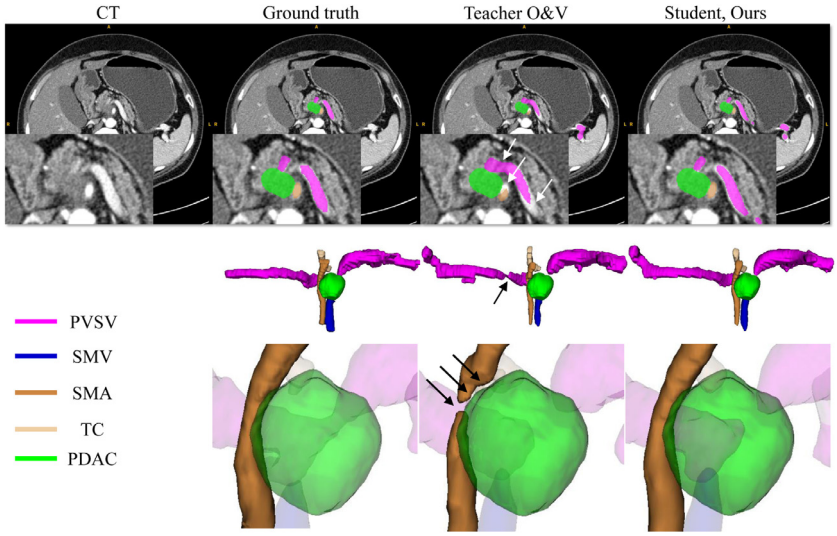

Fig. 5. A qualitative examples of vessel segmentation by the teacher and student models. Upper panel: The student model further improves the teacher's accuracy by identifying/delineating more accurate PVSV and SMA boundaries (white arrowheads). Middle panel: an inaccurate PVSV narrowing (black arrowhead) identified by the teacher model can be correct by the student model. Lower panel: an incorrect tumor-SMA separate (black arrowheads) identified by the teacher model can be correct by the student model. The histological examination confirmed the diagnosis of stage T4 PDAC, i.e., tumor involving the SMA. Note that the same manual PDAC mask is used in this example.

Energy, Entropy, etc. This group has 18 features in total (1st–18th).
- The texture features measured the spatial distribution of the voxel intensities, thereby quantifying the intra-tumoral heterogeneity. There are 34 texture features in total (19th–52nd).
- The geometry feature group contains features that quantified 3D shape characteristics of the tumor or lymph nodes. It is composed of 14 features in total (53rd–66th).
- Wavelet features were calculated by applying wavelet transformations to the original input images. There are 416 features in total (67th–482nd).

Subsequently, a coarse-to-fine feature selection strategy was used to select the most important imaging features and reduce the risk of bias and potential over-fitting. Univariable analysis was completed for each feature in the training set. Features with p-value < 0.1 were considered to be potentially associated with OS and were selected into the next process. The least absolute shrinkage and selection operator (LASSO) Cox regression (Tibshirani et al., 1997) method was then used to obtain the most statistically informative prognostic features from candidate features and construct the radiomic signature. The number of final selected features for each phase is different across folds, ranging from 6 to 20. To consider the power of multiphase CT imaging, we built a radiomic nomogram (Huang et al., 2016) that combines the radiomic signatures from 3 phases.

4.3.3. Results: effects of using single phase and multi-phase imaging protocols

In this section, we investigate the prognosis performances using single pancreatic and venous phase as well as all three phases from the DCE-CT imaging protocol (**Dataset A**). We build deep survival learning baseline models using ResNet3D as the backbone and then test them on a single contrast-enhanced phase (pancreatic or venous) and all three phases as inputs. Radiomics signature on single-phase and radiomics nomogram on all phases are reported accordingly. Table 3 presents the performances of different predicted models. The pancreatic phase CT is one specific imaging phase for pancreatic cancer staging and prognosis in the clinical environment, which provides the best contrast for observing the pancreas and tumor (NCCN-PDAC, 2020). From Table 3, we can first observe that deep survival models achieve better predictions than radiomics model using the same CT phases. Then, we can find the best performances of single-phase are from models using the pancreatic phase (0.623 vs 0.595 using CE-ResNet3D; 0.601 vs 0.575 using Radiomic signature). Better results can be observed when fusing all three phases together in both deep learning and radiomics models. ResNet3D with CE-ConvLSTM has further improved performance versus CE-ResNet3D with early fusion (0.645 vs 0.635). Similar trends could also be found when evaluating using survival AUC values at 1-year and 2-year clinical interested time-points.

To achieve statistical analysis, we follow one recent study (Lou et al., 2019) and apply a bootstrap method. For each fold, the test set is randomly resampled for calculating the C-index. This is repeated 100 times for each fold. The Wilcoxon signed rank test is then used to compare the C-index distributions. Results are presented in Table 4 and the bootstrapped C-index of Res-CE-ConvLSTM is 0.645 (95% CI 0.552–0.723) and for CE-ResNet3D with all three phases is 0.638 (95%CI 0.538–0.744). Two-sided p values after Bonferroni correction are reported. Results in this table illustrate that the dynamic enhancement CT imaging patterns learned and captured by CE-ConvLSTM can help achieve significant prediction improvements compared against early fusion CNNs.

4.3.4. Results: comparisons of multi-task deep learning models

We first investigate correlations between patients with "vessel contact" and R0/R1 resection margin status. We could study such correlations using automatically self-learned segmentation. We consider PDAC contact information with the surrounding four important vessels, which are consistent with our Section 4.2, including PVSV, SMV, SMA, and TC. We calculate the 3D Euclidean distance from each voxel of PDAC to the nearest voxels of those target vessels. If the distance of one PDAC voxel to the nearest voxel of one target $\leq$ 1mm, we consider such a voxel has contacted with the target vessel (considering the auto-segmentation error). Then we could calculate Spearman's correlation between contact areas (the number of contacted voxels) and resection margin status (R0/R1). Table 5 shows correlations of such correlations. We

Table 3
C-index, 1-year, and 2-year overall survival AUC values of different models across 4 folds. The average and standard deviation values are reported. N: non-contrast, P: pancreatic, V: venous phase.

Method	CT phases	C-index	1-yr AUC	2-yr AUC
Res-CE-ConvLSTM	N+P+V	0.645±0.013	0.684±0.057	0.689±0.032
CE-	N+P+V	0.635±0.030	0.616±0.117	0.682±0.040
ResNet3D	P	0.623±0.023	0.664±0.041	0.673±0.058
	V	0.595±0.030	0.605±0.032	0.619±0.027
Radiomics	N+P+V	0.608±0.050	0.600±0.046	0.662±0.090
Signature	P	0.601±0.059	0.585±0.066	0.637±0.096
	V	0.575±0.069	0.540±0.095	0.608±0.112

Table 4
Prediction performance of different deep learning models of outcome prediction using bootstrapped samples.

Methods	CT phases	C-index	95% CI	Adjusted *p*-value
Res-CE-ConvLSTM	N+P+V	0.645	0.552-0.723	ref
CE-	N+P+V	0.638	0.538-0.744	0.036
ResNet3D	P	0.625	0.525-0.718	<0.0001
	V	0.598	0.488-0.694	<0.0001
Radiomics	N+P+V	0.610	0.486-0.716	<0.0001
Signature	P	0.603	0.470-0.722	<0.0001
	V	0.571	0.396-0.700	<0.0001

Table 5
Spearman's correlations between tumor-vessel contacted area and resection margin status (R0/R1).

	Spearman's correlation	
Target vessel	ρ	p value
PVSV	0.11	0.056
SMV	0.29	<0.0001
SMA	-0.06 (0.67-1.27)	0.296
TC	-0.03 (0.97-1.96)	0.575

can see that the PVSV- and SMV-tumor contacts are significantly correlated with R0/R1 while SMA- and TC- contacts are not.

Our previous model (Yao et al., 2020a) did not include the pancreas anatomy segmentation. We then investigate if simply adding "contact with anatomy" could bring benefits. We encode 4-element by considering the PDAC contact areas with PVSV, SMV, SMA and TC, as these four vessels are believed to be more relevant to the resection margin and even OS. Then we concatenate these variables in the concat layer of our previous model (Yao et al., 2020a). Table 6 shows performances of each deep learning model. It could be seen that for R0/R1 prediction, the modified MICCAI 2020 model (Yao et al., 2020a) with contact feature could have improved sensitivity but decreased specificity than the original model. Similarly, for survival prediction, the modified model has a slightly lower c-index. The result could illustrate that simply adding confounding variables into the model might not provide significant performance gain.

To further evaluate the performance of multi-task baseline models, we report the results in comparison to recent multi-task deep prediction methods (Yao et al., 2020a; Tang et al., 2020; Lou et al., 2019), as shown in Table 7. Our previous model (Yao et al., 2020a) did not include the pancreas anatomy segmentation while other two models (Tang et al., 2020; Lou et al., 2019) cannot capture the tumor temporal enhancement changes because they only adopt 3D CNN networks on early-fused CT images. Classification performances are evaluated by the metrics of Balanced-Accuracy, Sensitivity, and Specificity while survival predictions are compared using c-index. In Table 7, single classification task uses CE-ResNet3D and CE-3DCNN can be found in the last two rows. It is shown that RM predictions from multi-task model achieve better balanced-ACC results, which demonstrates that the joint learning by adding survival branch can benefit classification task. We can see the proposed framework and our preliminary work (Yao et al., 2020a) achieve better results than the baseline deep multi-task models with early fusion (Tang et al., 2020; Lou et al., 2019). *This shows the effectiveness of capturing tumor dynamic changes across phases using CE-ConvLSTM.* By incorporating both the tumor attenuation and tumor-vascular involvement from pancreas and peripancreatic anatomy, the proposed model further improves performances in both survival and resection margin task over the previous model without using it (Yao et al., 2020a).

We assess the prognostic relevance of final staging systems from different multi-task models for predicting overall survival. Because staging in one fold is actually independent to the results in other folds, we can conduct analysis on the entire Dataset A by combining all four testing folds together. Univariate Cox analysis is first performed in this study and the importance of individual covariates is assessed by computing the Wald χ^2 statistic, as shown in Table 8. We can see that the proposed final staging as a time-dependent covariate yields an HR of 2.39 (95%CI 1.77-3.22) with a χ^2 value of 34.81 which provides improvements from single "risk staging" and "RM prediction". It can also be observed that the different final staging models achieve higher χ^2 and HR values than corresponding single staging systems, except for one model (Tang et al., 2020). Incorporating the prediction of R0/R1 with OS prediction in the multi-task model could bring performance gain from the model with only OS prediction. The jointly learning of cancer risk and resectability in a multi-task setting can derive more effective and comprehensive prognosis related deep image features and subsequently improve the prediction accuracies for both tasks.

Table 9 presents how different final staging results perform in the multivariable Cox model. Further multivariable analysis demonstrates the greatest contribution from the proposed model to the OS among all multi-task methods. Specifically, the highest χ^2 value of the proposed model indicatse its more than 50% contribution to the overall model χ^2. Results suggest that staging from the proposed model could be a (statistically) very strong prognostic marker compared with other staging markers.

4.4. Added clinical values

To validate the potential clinical values of our new imaging marker and staging system from the proposed model, we conduct extensive quantitative experiments and analysis to compare with both preoperative and postoperative important prognostic factors for pancreatic cancer. Preoperative factors are variables that capture before surgery, including imaging markers from baseline CT images, markers from the blood test (e.g., CA19-9), and Demographics (e.g., age, sex). Post-operative factors for PDAC are from evaluations according to NCCN/AJCC guidelines after surgery, which usually include standard TNM staging, tumor size, and resection margin measurement.

We first plot ROC curves of OS prediction in Fig. 6 of the proposed imaging marker with other factors including pTNM and tu-

Table 6
Ablation study of adding "contact with anatomy". Averaged results over four folds are reported.

	Task	Survival	Resection Margin: R0/R1		
		C-index	B-ACC	Sensitivity	Specificity
Yao et al., 2020a+contact feat	Mul	0.640	0.639	**0.563**	0.715
Yao et al., 2020a	Mul	**0.651**	**0.640**	0.489	**0.791**

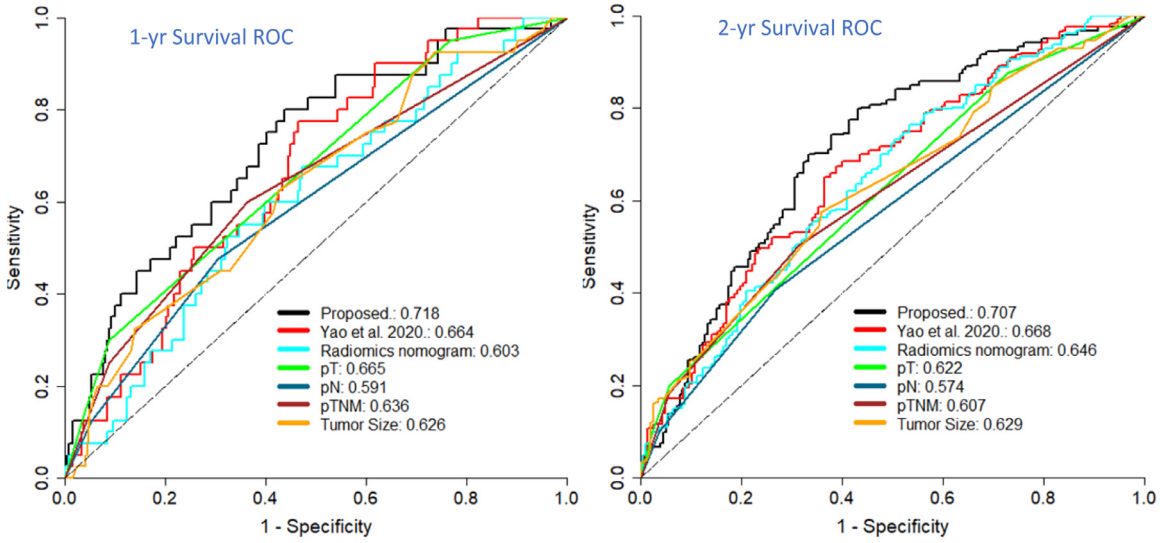

Fig. 6. 1-year and 2-year OS ROC curves of different markers.

Table 7
Average results of different methods over four folds. Mul: multi-task; cls: single classification.

	Task	Survival	Resection Margin: R0/R1		
		C-index	B-ACC	Sensitivity	Specificity
Proposed	Mul	**0.667**	**0.671**	**0.598**	0.743
Yao et al., 2020a	Mul	0.651	0.640	0.489	**0.791**
Tang et al., 2020	Mul	0.610	0.638	0.522	0.755
Lou et al., 2019	Mul	0.622	0.628	0.492	0.764
CE-ResNet3D	cls	-	0.609	0.473	0.745
CE-3DCNN	cls	-	0.611	0.508	0.714

Table 9
Multivariable Cox proportional hazards regression model for overall survival in the dataset by combing testing folds.

	Final Staging		
	HR (95%CI)	χ^2	p value
Proposed	1.80 (1.27–2.55)	11.17	0.00083
Yao et al., 2020a	1.51 (1.07–2.12)	5.74	0.017
Tang et al., 2020	0.92 (0.67–1.27)	0.26	0.61
Lou et al., 2019	1.38 (0.97–1.96)	3.23	0.07

mor size. The ROC curves of 1-year and 2-year OS are presented, respectively. It is interesting to see that the preoperative imaging marker from the proposed model can obtain statistically better results than pT stage, pN stage, and tumor size on pathological examinations to identify if PDAC patients could survive more than 1 year or 2 years. Our signature also performs better than other deep imaging signature (Yao et al., 2020a) and radiomics nomogram (Huang et al., 2016).

In Table 10, univariate and multivariate cox proportional-hazards models are used to evaluate the Hazard Ratio (HR) and log-rank test p-value for each factor, including final staging from different models, radiomics nomogram staging, and other clinicopathologic factors (preoperative CA19-9 and post-operative tumor size measurement). From the statistic analysis (in Table 10), the proposed signature is a strong prognostic factor in the univariate analysis just behind the post-operative resection margin in terms of HR values. The proposed signature remains representatively powerful in multivariable analysis (HR=1.809, p=0.0019) ad-

Table 8
Evaluation of different staging strategies by univariable analysis. *** indicates $p < .0001$, ** is $p < .001$, * is $p < .01$. *NS* is no significance. RM short for Resection Margin.

Method	Risk Staging		RM Prediction		Final Staging	
	HR (95%CI)	χ^2	HR (95%CI)	χ^2	HR(95% CI)	χ^2
Proposed	2.32 (1.74-3.10)	33.90***	1.53 (1.13-2.07)	7.27*	2.39 (1.77-3.22)	34.81***
Yao et al., 2020a	2.09 (1.57-2.78)	26.18***	1.79 (1.31-2.43)	12.39**	2.13 (1.59-2.85)	26.91***
Tang et al., 2020	1.84 (1.38-2.45)	17.59***	1.25 (0.92-1.71)	1.97^{NS}	1.50 (1.13-1.99)	7.99*
Lou et al., 2019	2.07 (1.55-2.77)	25.28***	1.55 (1.15-2.10)	7.59*	2.10 (1.57-2.81)	25.71***

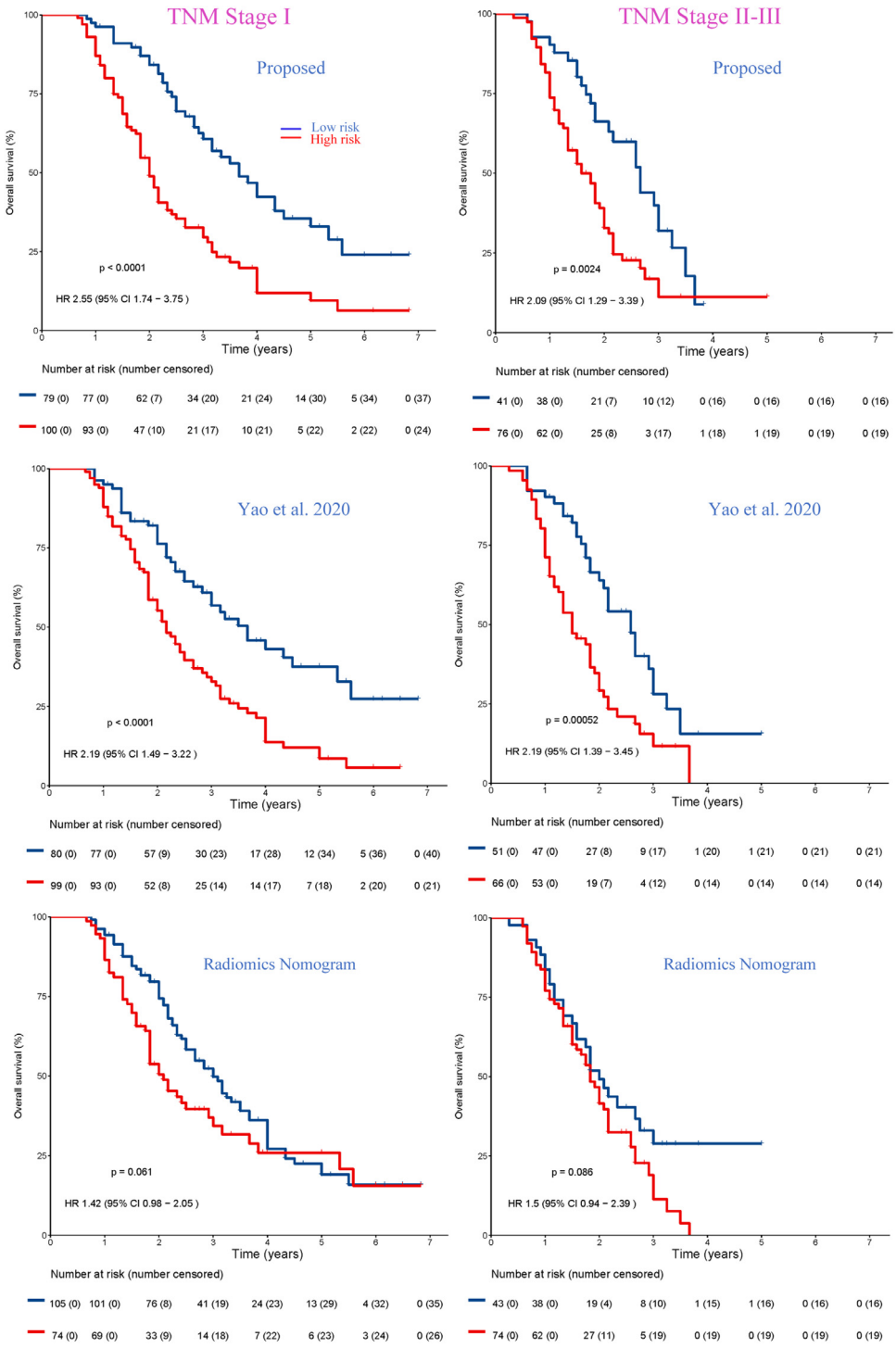

Fig. 7. Kaplan–Meier analyses of overall survival according to the proposed staging (1st row), our previous staging (2nd row) and radiomics nomogram (3rd row) in patients within two different subgroups (TNM I vs. II-III stage). The proposed staging significantly stratifies all subgroups.

justing for established clinicopathologic prognostic markers, for example, stromal fractions (HR=0.164, p=0.0013), resection margins (HR=4.323, $p < .0001$). Contributions of each factor could be found in Table 11 and our model provides the strongest factor among all preoperative factors. The proposed marker is also stronger than any other CT-derived imaging signature using radiomics nomogram (Huang et al., 2016) or our preliminary model (Yao et al., 2020a).

To demonstrate the added clinical value of our proposed marker, we plot Kaplan–Meier survival curves in Figs. 7 and 8 for patients with further stratification after grouping by TNM and tumor size, respectively, which are two well-established stratification criteria for making treatment decisions or recommendations. The log rank test is conducted to test the difference of two curves. *P*-values < 0.05 are considered as statistically significant. TNM stage and pathological tumor size are two clinically-established predictors. We study the added values of our proposed marker to subgroups of patients stratified by: (1) TNM staging I vs. II-III, (2) tumor size $<$ 3.0cm vs. $\geq$ 3.0 cm; where 3.0 cm is the clinically meaningful tumor size cutoff that has been used in recent clini-

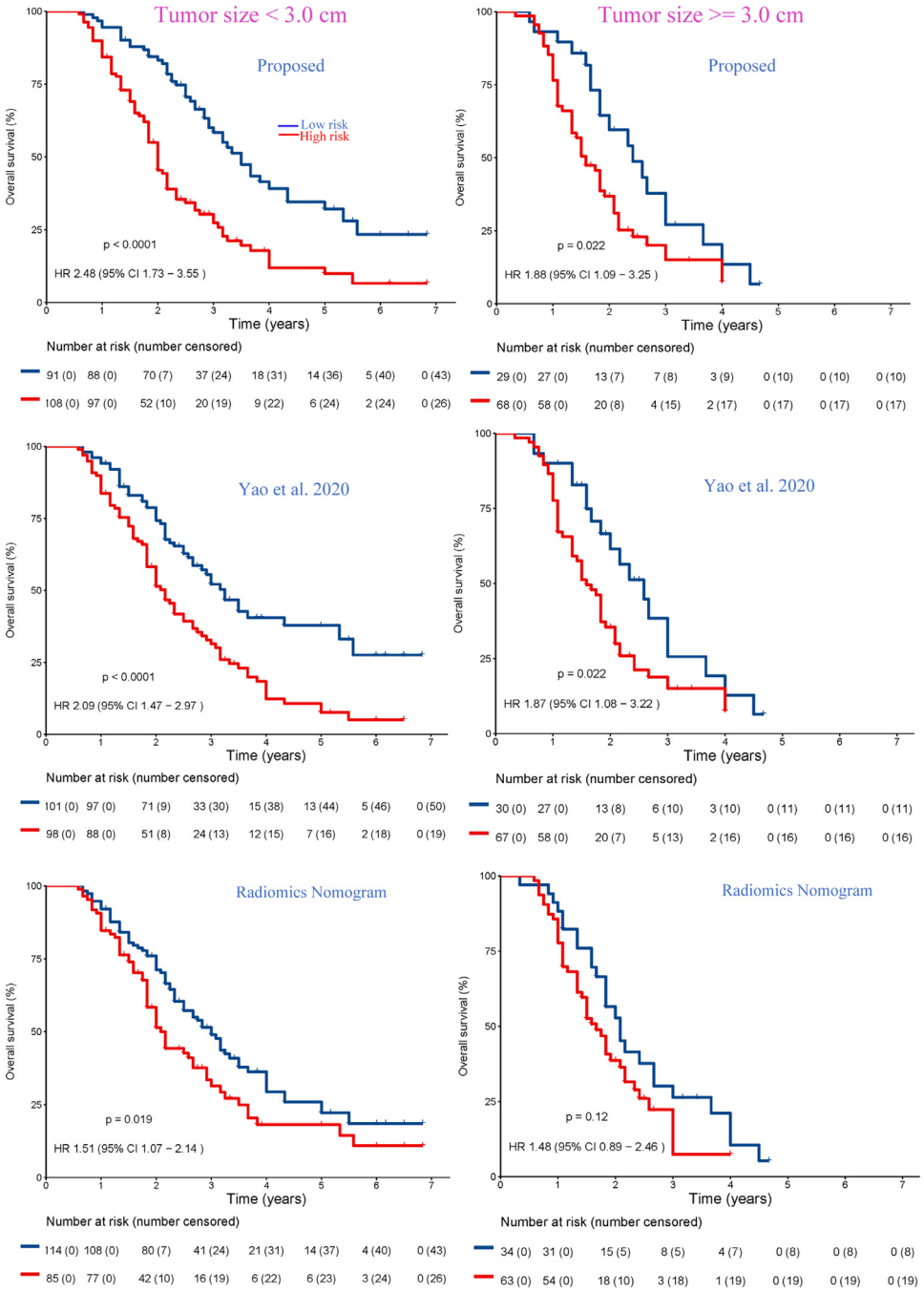

Fig. 8. Kaplan–Meier analyses of overall survival according to the proposed staging (1st row), our previous staging (2nd row) and radiomics nomogram (3rd row) in patients within two different subgroups (tumor size $<$ 3.0 cm vs. $\geq$ 3.0 cm). The proposed signature significantly stratifies all subgroups.

Table 10
Univariate and Multivariate Cox regression analysis. HR: hazard ratio; CI: confidence interval.

	Univariate Analysis		Multivariate Cox	
Factors	HR (95% CI)	*p*-value	HR (95% CI)	*p*-value
Stromal Fraction	0.131(0.047–0.367)	$<$ 0.0001	0.164(0.055–0.495)	0.0013
Tumor size	1.531(1.324–1.771)	$<$ 0.0001	0.959(0.667–1.379)	0.82
Resection margin	3.804(2.573–5.624)	$<$ 0.0001	4.323(2.665–7.014)	$<$ 0.0001
pT stage	1.911(1.569–2.328)	$<$ 0.0001	1.533(1.131–2.077)	0.0059
pN stage	1.545(1.226–1.947)	$<$ 0.0001	1.319(0.970–1.790)	0.078
CA19-9	1.001(1.000–1.001)	0.48	1.000(0.999–1.001)	0.098
Radiomics nomogram	1.610(1.215–2.133)	$<$ 0.0001	1.495(1.081–2.069)	0.015
Yao et al., 2020a	2.126(1.587–2.849)	$<$ 0.0001	1.428(0.979–2.084)	0.064
Proposed	2.390(1.773–3.223)	$<$ 0.0001	1.809(1.243–2.633)	0.0019

Table 11
Contribution of each variable to the overall model.

Variable	χ^2	Variable	χ^2
Resection Margin	27.23	pN Stage	5.73
Stromal fractions	10.04	Yao et al., 2020a	3.48
Proposed	9.77	CA19-9	2.63
pT Stage	6.36	Tumor Size	0.05
Radiomics Nomogram	5.80		

cal studies (Groot et al., 2019). From Fig. 7, it is shown that the proposed marker remains the most significant log-rank test outcome in the majority TNM subgroup – stage I. Radiomics marker does not reach the statistical significance within the two patient subgroups. Fig. 8 shows that the proposed staging achieves much better risk stratification performance (higher HR values) in smaller tumor size subgroup ($<$ 3.0 cm) than our previous final staging model (Yao et al., 2020a) and performs slighter better in $\geq$ 3.0 cm subgroup. After using the current clinicopathologic TNM staging system or tumor size, our proposed multi-phase CT imaging based staging can indeed further provide the risk stratification with significant evidence. This novel deep learning-based predictive prediction system (as described) may be combined with the established clinicopathological criteria to refine the risk stratification and guide the individualized treatment of patients with PDAC.

In one study (Bilimoria et al., 2007), the authors reported C-index 0.613 for patients underwent pancreatectomy (the same setting as our study); 0.63 for all patients including both resection and non-resection. Our method reaches a notably higher c-index of 0.667 vs 0.613 (AJCC 6th edition). Similarly, on our data in Fig. 6, our marker has higher AUCs of 0.718–0.707 than pT stage of 0.665–0.622. It's also worth mentioning that our marker is obtained pre-operatively while the AJCC staging for patients underwent resection is post-operative. In other words, our marker can be used to guide PDAC treatment selection before surgery while likely has a higher prediction accuracy than the post-operative staging system. Regarding the reported potentially higher c-index values in some clinical radiomics papers, we think it may be caused by several reasons, such as different patient cohorts, different disease (e.g., pT stage) distributions, smaller data size, and different evaluation design etc.

5. Discussion and conclusion

Our method could be improved in the following aspects. First, a quality check of the segmentation by expert radiologist is needed, especially for the tumor-vessel contact regions. Second, a more sophisticated representation of the tumor-vessel relationship can be designed to better describe the important clinical prior knowledge. Third, a CT-based lymph node metastasis detection might help understand the disease more comprehensively. This work mainly aims to develop a comprehensive deep learning-based imaging biomarker and investigates its prognostic value when adjusted by other established clinical and pathological variables. The evaluation of the added value of this biomarker to clinical variables and staging system such as CA 19-9 and AJCC is the next step of this development study.

In this paper, we propose a new multi-task CNN framework for cancer survival prediction by simultaneously predicting the tumor resection margins for PDAC patients. The use of CE-ConvLSTM and self-learning segmentation to consider and encode the dynamic tumor attenuation patterns and pancreas anatomies boosts the whole framework, to significantly outperform the early fusion deep learning models and conventional radiomics-based survival models. Our results also validate that final staging from both survival and resection margin predictions can serve as a strong prognostic and predictive quantitative biomarker for subgroup patients after staging by the well-established pathological TNM or tumor size criteria. This makes our model very promising in future clinical usage to refine the risk stratification and guide the surgery treatment recommendations of patients with primary resectable PDAC tumors.

Declaration of Competing Interest

We have no conflicts of interest to disclose.

CRediT authorship contribution statement

Jiawen Yao: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft. **Yu Shi:** Conceptualization, Funding acquisition, Data curation, Formal analysis, Methodology, Writing – review & editing. **Kai Cao:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. **Le Lu:** Supervision, Writing - review & editing. **Jianping Lu:** Supervision, Writing - review & editing. **Qike Song:** Funding acquisition, Data curation, Methodology, Writing - review & editing. **Gang Jin:** Supervision, Writing - review & editing. **Jing Xiao:** Supervision, Writing - review & editing. **Yang Hou:** Supervision, Writing - review & editing. **Ling Zhang:** Conceptualization, Methodology, Formal analysis, Software, Writing - original draft, Supervision, Writing - review & editing.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (82071885, 81771802, 81771893) and the National Youth Talent Support Program (2019).

References

Aerts, H.J., Velazquez, E.R., Leijenaar, R.T., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., et al., 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat. Commun. 5 (1), 1–9.

Attiyeh, M.A., Chakraborty, J., Doussot, A., Langdon-Embry, L., Mainarich, S., Gönen, M., Balachandran, V.P., D'Angelica, M.I., DeMatteo, R.P., Jarnagin, W.R., et al., 2018. Survival prediction in pancreatic ductal adenocarcinoma by quantitative computed tomography image analysis. Ann. Surg. Oncol. 25 (4), 1034–1042.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C., 2017. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data 4 (1), 1–13.

Bilimoria, K.Y., Bentrem, D.J., Ko, C.Y., Ritchey, J., Stewart, A.K., Winchester, D.P., Talamonti, M.S., 2007. Validation of the 6th edition AJCC pancreatic cancer staging system: report from the national cancer database. Cancer Interdiscip. Int. J. Am. Cancer Soc. 110 (4), 738–744.

Cai, X., Gao, F., Qi, Y., Lan, G., Zhang, X., Ji, R., Xu, Y., Liu, C., Shi, Y., 2020. Pancreatic adenocarcinoma: quantitative CT features are correlated with fibrous stromal fraction and help predict outcome after resection. Eur. Radiol. 30 (9), 5158–5169.

Cassinotto, C., Chong, J., Zogopoulos, G., Reinhold, C., Chiche, L., Lafourcade, J.-P., Cuggia, A., Terrebonne, E., Dohan, A., Gallix, B., 2017. Resectable pancreatic adenocarcinoma: role of ct quantitative imaging biomarkers for predicting pathology and patient outcomes. Eur. J. Radiol. 90, 152–158.

Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D.Z., 2016. Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 3036–3044.

Chen, S., Ma, K., Zheng, Y., 2019. Med3d: transfer learning for 3D medical image analysis. arXiv preprint arXiv:1904.00625.

Cheng, N.-M., Yao, J., Cai, J., Ye, X., Zhao, S., Zhao, K., Zhou, W., Nogues, I., Huo, Y., Liao, C.-T., Wang, H.-M., Lin, C.-Y., Lee, L.-Y., Xiao, J., Lu, L., Zhang, L., Yen, T.-C., 2021. Deep Learning for Fully Automated Prediction of Overall Survival in Patients with Oropharyngeal Cancer Using FDG-PET Imaging. Clin. Cancer Res. 27 (14), 3948–3959. doi:10.1158/1078-0432.CCR-20-4935, In press.

Choy, C.B., Xu, D., Gwak, J., Chen, K., Savarese, S., 2016. 3D-R2N2: a unified approach for single and multi-view 3d object reconstruction. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Proceedings of the ECCV 2016. Springer, Cham, pp. 628–644.

Chu, L.C., Park, S., Kawamoto, S., Yuille, A.L., Hruban, R.H., Fishman, E.K., 2020. Pancreatic cancer imaging: a new look at an old problem. Curr. Probl. Diagn. Radiol. 50 (4), 540–550. doi:10.1067/j.cpradiol.2020.08.002.

Dalal, V., Carmicheal, J., Dhaliwal, A., Jain, M., Kaur, S., Batra, S.K., 2020. Radiomics in stratification of pancreatic cystic lesions: machine learning in action. Cancer Lett. 469, 228–237.
Dickinson, S.M., McIntyre, C.A., Schilsky, J.B., Harrington, K.A., Gerst, S.R., Flynn, J.R., Gonen, M., Capanu, M., Wong, W., Lawrence, S., et al., 2020. Preoperative ct predictors of survival in patients with pancreatic ductal adenocarcinoma undergoing curative intent surgery. Abdom. Radiol. 1–11.
Eilaghi, A., Baig, S., Zhang, Y., Zhang, J., Karanicolas, P., Gallinger, S., Khalvati, F., Haider, M.A., 2017. Ct texture features are associated with overall survival in pancreatic ductal adenocarcinoma–a quantitative analysis. BMC Med. Imaging 17 (1), 1–7.
Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., Davidson, B., Pereira, S.P., Clarkson, M.J., Barratt, D.C., 2018. Automatic multi-organ segmentation on abdominal CT with dense v-networks. TMI 37 (8), 1822–1834.
Gillies, R.J., Kinahan, P.E., Hricak, H., 2016. Radiomics: images are more than pictures, they are data. Radiology 278 (2), 563–577.
Groot, V.P., Gemenetzis, G., Blair, A.B., Rivero-Soto, R.J., Yu, J., Javed, A.A., Burkhart, R.A., Rinkes, I.H.B., Molenaar, I.Q., Cameron, J.L., et al., 2019. Defining and predicting early recurrence in 957 patients with resected pancreatic ductal adenocarcinoma. Ann. Surg. 269 (6), 1154–1162. doi:10.1097/SLA.0000000000002734.
Grossberg, A.J., Chu, L.C., Deig, C.R., Fishman, E.K., Hwang, W.L., Maitra, A., Marks, D.L., Mehta, A., Nabavizadeh, N., Simeone, D.M., et al., 2020. Multidisciplinary standards of care and recent progress in pancreatic ductal adenocarcinoma. CA Cancer J. Clin. 70 (5), 375–403.
Guan, M., Gulshan, V., Dai, A., Hinton, G., 2018. Who said what: Modeling individual labelers improves classification. In: Proceedings of the AAAI Conference on Artificial Intelligence, 32.
Hara, K., Kataoka, H., Satoh, Y., 2018. Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and imagenet? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6546–6555.
Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L., Rosati, R.A., 1982. Evaluating the yield of medical tests. JAMA 247 (18), 2543–2546.
Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, S.M., Schnabel, J.A., 2013. Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.), Proceedings of the MICCAI 2013. Springer, Berlin, Heidelberg, pp. 187–194.
Hong, S.B., Lee, S.S., Kim, J.H., Kim, H.J., Byun, J.H., Hong, S.M., Song, K.-B., Kim, S.C., 2018. Pancreatic cancer CT: prediction of resectability according to NCCN criteria. Radiology 289 (3), 710–718.
Huang, Y.-q., Liang, C.-h., He, L., Tian, J., Liang, C.-s., Chen, X., Ma, Z.-l., Liu, Z.-y., 2016. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J. Clin. Oncol. 34 (18), 2157–2164.
Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods 18, 203–211.
Jiang, Y., Jin, C., Yu, H., Wu, J., Chen, C., Yuan, Q., Huang, W., Hu, Y., Xu, Y., Zhou, Z., 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.. Ann. Surg. doi:10.1097/SLA.0000000000003778.
Joo, I., Lee, J.M., Lee, E.S., Son, J.-Y., Lee, D.H., Ahn, S.J., Chang, W., Lee, S.M., Kang, H.-J., Yang, H.K., 2019. Preoperative ct classification of the resectability of pancreatic cancer: interobserver agreement. Radiology 293 (2), 343–349.
Jungo, A., McKinley, R., Meier, R., Knecht, U., Vera, L., Pérez-Beteta, J., Molina-García, D., Pérez-García, V.M., Wiest, R., Reyes, M., 2017. Towards uncertainty-assisted brain tumor segmentation and survival prediction. In: Proceedings of the International MICCAI Brainlesion Workshop. Springer, pp. 474–485.
Katzman, J., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y., 2016. Deep survival: a deep cox proportional hazards network. arXiv preprint arXiv:1606.00931.
Khoreva, A., Benenson, R., Hosang, J., Hein, M., Schiele, B., 2017. Simple does it: Weakly supervised instance and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 876–885.
Kim, H., Goo, J.M., Lee, K.H., Kim, Y.T., Park, C.M., 2020. Preoperative ct-based deep learning model for predicting disease-free survival in patients with lung adenocarcinomas. Radiology 296 (1), 216–224.
Kim, J.H., Park, S.H., Yu, E.S., Kim, M.-H., Kim, J., Byun, J.H., Lee, S.S., Hwang, H.J., Hwang, J.-Y., Lee, S.S., et al., 2010. Visually isoattenuating pancreatic adenocarcinoma at dynamic-enhanced CT: frequency, clinical and pathologic characteristics, and diagnosis at imaging examinations. Radiology 257 (1), 87–96.
Koay, E.J., Lee, Y., Cristini, V., Lowengrub, J.S., Kang, Y., San Lucas, F.A., Hobbs, B.P., Ye, R., Elganainy, D., Almahariq, M., et al., 2018. A visually apparent and quantifiable ct imaging feature identifies biophysical subtypes of pancreatic ductal adenocarcinoma. Clin. Cancer Res. 24 (23), 5883–5894.
Konstantinidis, I.T., Warshaw, A.L., Allen, J.N., Blaszkowsky, L.S., Fernandez-del Castillo, C., Deshpande, V., Hong, T.S., Kwak, E.L., Lauwers, G.Y., Ryan, D.P., et al., 2013. Pancreatic ductal adenocarcinoma: is there a survival difference for R1 resections versus locally advanced unresectable tumors? What is a ætrueg R0 resection? Ann. Surg. 257 (4), 731–736.
Kwan, J.Y.Y., Su, J., Huang, S.H., Ghoraie, L.S., Xu, W., Chan, B., Yip, K.W., Giuliani, M., Bayley, A., Kim, J., et al., 2018. Radiomic biomarkers to refine risk models for distant metastasis in hpv-related oropharyngeal carcinoma. Int. J. Radiat. Oncol.* Biol.* Phys. 102 (4), 1107–1116.
Liu, Z., Sun, Q., Bai, H., Liang, C., Chen, Y., Li, Z.-C., 2019. 3d deep attention network for survival prediction from magnetic resonance images in glioblastoma. In: Proceedings of the IEEE International Conference on Image Processing (ICIP). IEEE, pp. 1381–1384.
Lou, B., Doken, S., Zhuang, T., Wingerter, D., Gidwani, M., Mistry, N., Ladic, L., Kamen, A., Abazeed, M.E., 2019. An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. Lancet Digit. Health 1 (3), e136–e147.
Lubner, M.G., Smith, A.D., Sandrasegaran, K., Sahani, D.V., Pickhardt, P.J., 2017. Ct texture analysis: definitions, applications, biologic correlates, and challenges. Radiographics 37 (5), 1483–1503.
Mizrahi, J.D., Surana, R., Valle, J.W., Shroff, R.T., 2020. Pancreatic cancer. Lancet 395 (10242), 2008–2020.
Mobadersany, P., Yousefi, S., Amgad, M., Gutman, D.A., Barnholtz-Sloan, J.S., Vega, J.E.V., Brat, D.J., Cooper, L.A., 2018. Predicting cancer outcomes from histology and genomics using convolutional networks. Proc. Natl. Acad. Sci. 115 (13), E2970–E2979.
NCCN-PDAC, 2020. NCCN clinical practice guidelines in oncology (NCCN Guidelinesë) pancreatic adenocarcinoma.
Nie, D., Lu, J., Zhang, H., Adeli, E., Wang, J., Yu, Z., Liu, L., Wang, Q., Wu, J., Shen, D., 2019. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. Sci. Rep. 9 (1), 1–14.
Nie, D., Zhang, H., Adeli, E., Liu, L., Shen, D., 2016. 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 212–220.
Roth, H., Zhang, L., Yang, D., Milletari, F., Xu, Z., Wang, X., Xu, D., 2019. Weakly supervised segmentation from extreme points. In: Large-Scale Annotation of Biomedical Data and Expert Label Synthesis and Hardware Aware Learning for Medical Imaging and Computer Assisted Intervention. Springer International Publishing, Cham, pp. 42–50.
Siegel, R.L., Miller, K.D., Jemal, A., 2019. Cancer statistics, 2019. CA Cancer J. Clin. 69 (1), 7–34.
Simpson, A. L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., et al., 2019. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063.
Skrede, O.-J., De Raedt, S., Kleppe, A., Hveem, T.S., Liestøl, K., Maddison, J., Askautrud, H.A., Pradhan, M., Nesheim, J.A., Albregtsen, F., et al., 2020. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet 395 (10221), 350–360.
Tang, Z., Xu, Y., Jin, L., Aibaidula, A., Lu, J., Jiao, Z., Wu, J., Zhang, H., Shen, D., 2020. Deep learning of imaging phenotype and genotype for predicting overall survival time of glioblastoma patients. IEEE Trans. Med. Imaging 39 (6), 2100–2109.
Tibshirani, R., et al., 1997. The lasso method for variable selection in the cox model. Stat. Med. 16 (4), 385–395.
Traverso, A., Wee, L., Dekker, A., Gillies, R., 2018. Repeatability and reproducibility of radiomic features: a systematic review. Int. J. Radiat. Oncol.* Biol.* Phys. 102 (4), 1143–1158.
Tummers, W., Groen, J., Mulder, B.S., Farina-Sarasqueta, A., Morreau, J., Putter, H., van de Velde, C., Vahrmeijer, A., Bonsing, B., Mieog, J., et al., 2019. Impact of resection margin status on recurrence and survival in pancreatic cancer surgery. Br. J. Surg. 106 (8), 1055.
Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.-C., Pieper, S., Aerts, H.J., 2017. Computational radiomics system to decode the radiographic phenotype. Cancer Res. 77 (21), e104–e107.
Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M., 2017. Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2097–2106.
Wang, Y., Lu, L., Cheng, C.-T., Jin, D., Harrison, A.P., Xiao, J., Liao, C.-H., Miao, S., 2019. Weakly supervised universal fracture detection in pelvic x-rays. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.-T., Khan, A. (Eds.), Proceedings of the MICCAI 2019. Springer, Cham, pp. 459–467.
Wang, Y., Wei, X., Liu, F., Chen, J., Zhou, Y., Shen, W., Fishman, E.K., Yuille, A.L., 2020. Deep distance transform for tubular structure segmentation in CT scans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3833–3842.
Wulczyn, E., Steiner, D.F., Moran, M., Plass, M., Reihs, R., Tan, F., Flament-Auvigne, I., Brown, T., Regitnig, P., Chen, P.-H.C., et al., 2021. Interpretable survival prediction for colorectal cancer using deep learning. npj Digit. Med. 4 (1), 1–13.
Wulczyn, E., Steiner, D.F., Xu, Z., Sadhwani, A., Wang, H., Flament-Auvigne, I., Mermel, C.H., Chen, P.-H.C., Liu, Y., Stumpe, M.C., 2020. Deep learning-based survival prediction for multiple cancer types using histopathology images. PLoS ONE 15 (6), e0233678.
Xia, Y., Yu, Q., Shen, W., Zhou, Y., Fishman, E.K., Yuille, A.L., 2020. Detecting pancreatic ductal adenocarcinoma in multi-phase ct scans via alignment ensemble. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 285–295.
Xie, Q., Luong, M.-T., Hovy, E., Le, Q.V., 2020. Self-training with noisy student improves imagenet classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10687–10698.
Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, I., Mak, R.H., Aerts, H.J., 2019. Deep learning predicts lung cancer treatment response from serial medical imaging. Clin. Cancer Res. 25 (11), 3266–3275. doi:10.1158/1078-0432.CCR-18-2495.
Xu, Z., Lee, C.P., Heinrich, M.P., Modat, M., Rueckert, D., Ourselin, S., Abramson, R.G.,

Landman, B.A., 2016. Evaluation of six registration methods for the human abdomen on clinically acquired CT. IEEE Trans. Biomed. Eng. 63 (8), 1563–1572.
Yao, J., Shi, Y., Lu, L., Xiao, J., Zhang, L., 2020a. Deepprognosis: preoperative prediction of pancreatic cancer survival and surgical margin via contrast-enhanced CT imaging. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 272–282.
Yao, J., Zhu, X., Huang, J., 2019. Deep multi-instance learning for survival prediction from whole slide images. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 496–504.
Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N., Huang, J., 2020b. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. Med. Image Anal. 65, 101789.
Yao, J., Zhu, X., Zhu, F., Huang, J., 2017. Deep correlational learning for survival prediction from multi-modality data. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 406–414.
Yun, G., Kim, Y.H., Lee, Y.J., Kim, B., Hwang, J.-H., Choi, D.J., 2018. Tumor heterogeneity of pancreas head cancer assessed by ct texture analysis: association with survival outcomes after curative resection. Sci. Rep. 8 (1), 1–10.
Zhang, L., Gopalakrishnan, V., Lu, L., Summers, R.M., Moss, J., Yao, J., 2018. Self-learning to detect and segment cysts in lung CT images without manual annotation. In: Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 1100–1103.
Zhang, L., Lu, L., Nogues, I., Summers, R.M., Liu, S., Yao, J., 2017. Deeppap: deep convolutional networks for cervical cell classification. IEEE J. Biomed. Health Inform. 21 (6), 1633–1643.
Zhang, L., Lu, L., Wang, X., Zhu, R.M., Bagheri, M., Summers, R.M., Yao, J., 2019. Spatio-temporal convolutional LSTMs for tumor growth prediction by learning 4D longitudinal patient data. IEEE Trans. Med Imaging 39 (4), 1114–1126.
Zhang, L., Shi, Y., Yao, J., Bian, Y., Cao, K., Jin, D., Xiao, J., Lu, L., 2020. Robust pancreatic ductal adenocarcinoma segmentation with multi-institutional multi-phase partially-annotated CT scans. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 491–500.
Zhao, T., Cao, K., Yao, J., Nogues, I., Lu, L., Huang, L., Xiao, J., Yin, Z., Zhang, L., 2021. 3D graph anatomy geometry-integrated network for pancreatic mass segmentation, diagnosis, and quantitative patient management. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13743–13752.
Zhou, T., Ruan, S., Canu, S., 2019. A review: Deep learning for medical image segmentation using multi-modality fusion. Array 100004.
Zhu, X., Yao, J., Huang, J., 2016. Deep convolutional neural network for survival analysis with pathological images. In: Proceedings of the BIBM. IEEE, pp. 544–547.
Zhu, X., Yao, J., Zhu, F., Huang, J., 2017. WSISA: making survival prediction from whole slide histopathological images. In: Proceedings of the CVPR, pp. 7234–7242.
Zhu, Z., Xia, Y., Shen, W., Fishman, E., Yuille, A., 2018. A 3D coarse-to-fine framework for volumetric medical image segmentation. In: Proceedings of the International Conference on 3D Vision (3DV). IEEE, pp. 682–690.
Zhu, Z., Xia, Y., Xie, L., Fishman, E.K., Yuille, A.L., 2019. Multi-scale coarse-to-fine segmentation for screening pancreatic ductal adenocarcinoma. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 3–12.