International Journal of Computer Vision
https://doi.org/10.1007/s11263-024-02314-1

Deep Attention Learning for Pre-operative Lymph Node Metastasis Prediction in Pancreatic Cancer via Multi-object Relationship Modeling

Zhilin Zheng[1] · Xu Fang[2] · Jiawen Yao[1] · Mengmeng Zhu[2] · Le Lu[1] · Yu Shi[3] · Hong Lu[4] · Jianping Lu[2] · Ling Zhang[1] · Chengwei Shao[2] · Yun Bian[2]

Received: 6 February 2024 / Accepted: 26 November 2024

Abstract

Lymph node (LN) metastasis status is one of the most critical prognostic and cancer staging clinical factors for patients with resectable pancreatic ductal adenocarcinoma (PDAC, generally for any types of solid malignant tumors). Pre-operative prediction of LN metastasis from non-invasive CT imaging is highly desired, as it might be directly and conveniently used to guide the follow-up neoadjuvant treatment decision and surgical planning. Most previous studies only use the tumor characteristics in CT imaging alone to implicitly infer LN metastasis. To the best of our knowledge, this is the first work to propose a fully-automated LN segmentation and identification network to directly facilitate the LN metastasis status prediction task for patients with PDAC. Specially, (1) we explore the anatomical spatial context priors of pancreatic LN locations by generating a guiding attention map from related organs and vessels to assist segmentation and infer LN status. As such, LN segmentation is impelled to focus on regions that are anatomically adjacent or plausible with respect to the specific organs and vessels. (2) The metastasized LN identification network is trained to classify the segmented LN instances into positives or negatives by reusing the segmentation network as a pre-trained backbone and padding a new classification head. (3) Importantly, we develop a LN metastasis status prediction network that combines and aggregates the holistic patient-wise diagnosis information of both LN segmentation/identification and deep imaging characteristics by the PDAC tumor region. Extensive quantitative nested five-fold cross-validation is conducted on a discovery dataset of 749 patients with PDAC. External multi-center clinical evaluation is further performed on two other hospitals of 191 total patients. Our multi-staged LN metastasis status prediction network statistically significantly outperforms strong baselines of nnUNet and several other compared methods, including CT-reported LN status, radiomics, and deep learning models.

Keywords Pancreatic ductal adenocarcinoma (PDAC) · Lymph node metastasis · Lymph node segmentation · Contrast-enhanced computed tomography

Communicated by Ziyue Xu.

Xu Fang, Jiawen Yao have contributed equally to this work.

✉ Le Lu
le.lu@alibaba-inc.com

✉ Chengwei Shao
cwshao@sina.com

✉ Yun Bian
bianyun2012@foxmail.com

Zhilin Zheng
zhengzhilin.zzl@alibaba-inc.com

Xu Fang
fx0412@foxmail.com

Jiawen Yao
yjiaweneecs@gmail.com

[1] DAMO Academy, Alibaba Group, Beijing, China

[2] Department of Radiology, Changhai Hospital, Shanghai, China

[3] Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, China

[4] Department of Radiology, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China

Published online: 20 December 2024

1 Introduction

Pancreatic cancer is the third leading cause of overall cancer death in the United States (Siegel et al., 2021), of which approximately 95% is pancreatic ductal adenocarcinoma (PDAC) (Grossberg et al., 2020). With the poorest prognosis (i.e., 5-year overall survival (OS) of 10% approximately) among all types of solid cancerous malignancies, surgical resection is the most effective way to achieve long-term survival for patients with PDAC (Grossberg et al., 2020). However, not all patients can benefit from the margin-negative (R0) resection. Comprehensive treatment protocols are established for pancreatic cancer. For example, selecting the proper cancer treatment for each patient must consider the condition whether peripancreatic lymph nodes (LNs) have been metastasized so that the options of adjuvant radiotherapy (RT) or chemotherapy become first applicable. It is found that neoadjuvant therapy before surgery was associated with improved survival and time-to-recurrence in patients with LN metastasis, since neoadjuvant therapy can not only treat lymphovascular invasion but also benefit tumor downstaging (Roland et al., 2015; Kanda et al., 2011). Accurate pre-operative localization and prediction of LN metastasis becomes vital and would aid critically in the treatment planning and management.

Contrast-enhanced CT is used as the typical imaging protocol for identifying the presence of peripancreatic metastatic disease to LNs, but it is a very challenging task for radiologists to determine whether a patient has LN metastasis when using only CT scans. To this end, low diagnostic accuracy of CT with a pooled sensitivity of 25% and positive predictive value (PPV) of 28% was reported in a meta-analysis (Tseng et al., 2014) on assessing extra-regional LN metastasis in pancreatic and peri-ampullary cancer. On the other hand, recently several radiomics based approaches have been proposed to tackle the LN metastasis differentiation problem from various cancer types (Ji et al., 2019; Wang et al., 2020; Li et al., 2020; Bian et al., 2019; Yang et al., 2020; Gao et al., 2020; Meng et al., 2020). However, these methods require hand-crafted feature design which can bring concerns of reproducibility and human bias is often introduced since manual selection of 2D representative tumor slices and regions of interest is a prerequisite by radiomics methods. Additionally, there are several previous deep learning work that report promising performance on predicting LN metastasis status in gastric cancer (Jin et al., 2021; Dong et al., 2020). Those models assume that the risk estimate of metastases is fundamentally driven by the primary tumor. They rely on LN CT report information for the integration model without exploiting any explicit process of LNs detection or segmentation. Thus if both tumor morphology and lymphatic anatomy can be taken into account simultaneously, it could be of great clinical utility. This information processing pipeline is actually more similar with the diagnostic processes performed by human radiologist readers. Last, PET/CT is another imaging modality worth exploring. PET/CT based approaches (Kim et al., 2019; Asagi et al., 2013; Dahmarde et al., 2020) generally use the maximum standardized uptake value (SUV_{max}) of manually-drawn LN RoIs as the prediction element, but it comes with challenges of numerous false positives from inflammatory LNs and false negatives from small-sized metastatic LNs (Tseng et al., 2021; Jung et al., 2017). Also, PET/CT is not as common as CT, more expensive, less affordable and accessible. Hence we opt for multiphase CT imaging to tackle this problem.

In this paper, we tackle the LN metastasis status prediction problem for patients with PDAC by first segmenting and identifying instances of LNs and then holistically classifying the patients into *metastasis-positive* or *-negative* group. LNs are tiny anatomical structures surrounding organs and vessels. Their locations are mapped into 18 stations with respect to the primary pancreatic cancer tumor by their relative positions against adjacent anatomical structures, as defined by Japan Pancreas Society (JPS) (Kanehara, 2017) (see Table 6 in the Appendix A for details). Examples of their spatial distribution are shown in Fig. 1. Response Evaluation Criteria in Solid Tumors (RECIST) criteria (Eisenhauer et al., 2009) defines the criteria for LN metastasis suspicion, i.e., nodes with short axis greater than 10*mm*, heterogeneity and central necrosis. However, these criteria are not pathognomonic since there exist false negatives associated with small node micrometastases and false positives with inflammatory nodes larger than 10*mm* in short axis. Finding LNs in CT images is quite time-consuming and can be inconsistent based on radiologists' subjective experiences. It is usually ambiguous for radiologists to identify nodal positivity from CT alone without referring to pathology reports. The current gold standard for determination of metastasis is based on post-operative pathological evaluation of pancreatectomy specimens. Automated and reliable pre-operative LN segmentation and identification are highly desirable and still considered as a critical unmet clinical need.

LN segmentation is especially challenging due to two reasons: (1) tiny to small sizes of LNs cause extreme foreground-background class imbalance problem; (2) LNs have CT attenuation values that are very similar to that of vessels and other soft-tissue structures. Existing work (Oda et al., 2018; Bouget et al., 2019; Guo et al., 2021) mainly adopt U-Net based segmentation networks (Ronneberger et al., 2015; Çiçek et al., 2016; Isensee et al., 2021) as strong backbones. The spatial relationship between lymphatic anatomy and adjacent anatomical structures is not well explored. To exploit inter-object relationships, previous methods either apply attention mechanism to highlight probable object cues (Sang et al., 2022; Min et al., 2022), or incorporate topological constraints (Gupta et al., 2022).

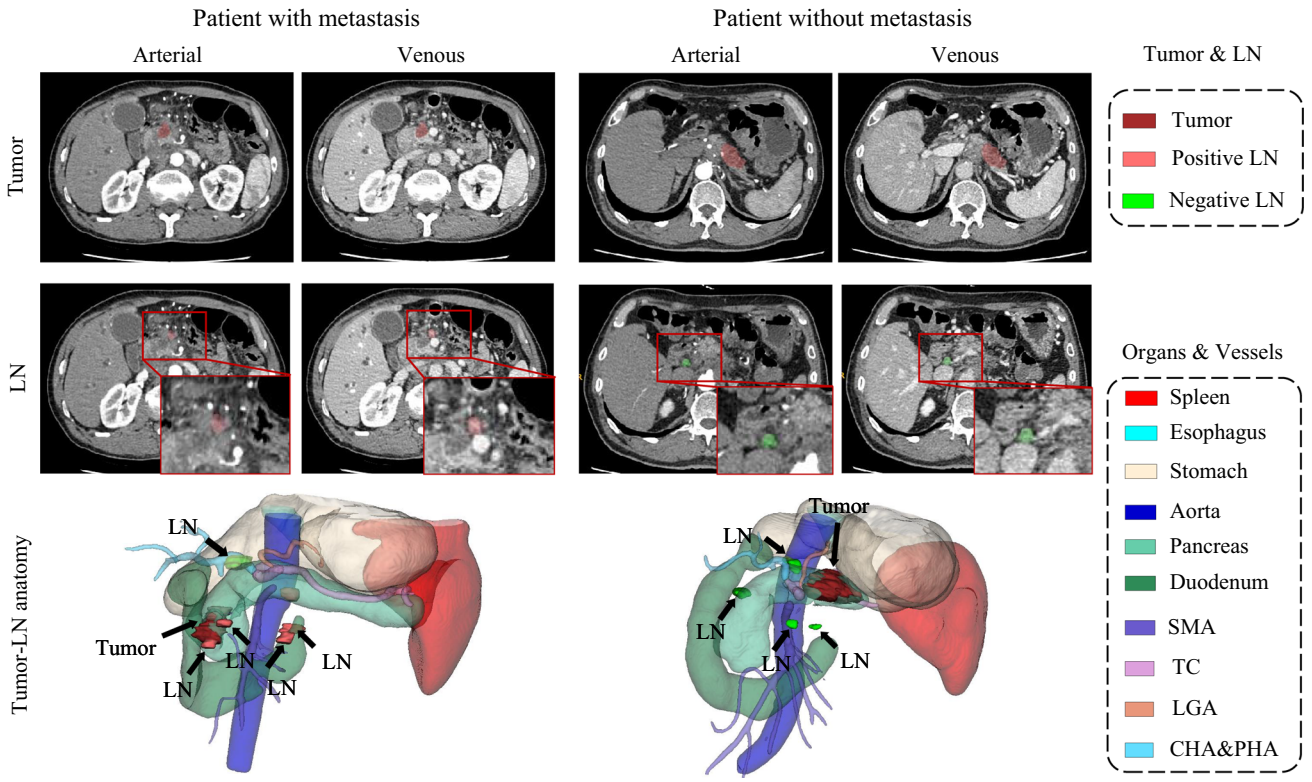


Fig. 1 A visualization of pancreatic tumor (in dark red) and LNs (in pink red for positives or green for negatives) in multi-phase CT images and their spatial distributions corresponding to key anatomical structures as follows. SMA: superior mesenteric artery. TC&SA: truncus coeliacus and splenic artery. LGA: left gastric artery; CHA&PHA: common hepatic artery and proper hepatic artery (Color figure online)

Sang et al. (2022) quantifies the relationship between small and large objects using attention on different feature levels. Min et al. (2022) associates foreground-related contexts by generating an attention map defining the relation between the foreground and scene. While these work implicitly encode object relationships, they are limited in their capacity to deal with explicit topological interactions that naturally arise in anatomical structures. Gupta et al. (2022) considers two common types of object interactions (i.e. containment and exclusion) and enforces topological constraints by directly identifying locations where the constraints are violated. We address this issue by introducing a distance-guided attention map to fully utilize the anatomically-constrained spatial priors. In our work, distance maps of relevant organs/vessels are generated using distance transform (Rosenfeld and Pfaltz, 1968), a powerful tool widely used in medical image segmentation (e.g., Sironi et al. (2014); Wang et al. (2020)). LN attention map is obtained via a pre-defined mapping from distance maps and then integrated into the segmentation network to control its spatial focus. It improves sample selection strategy that filters out non-informative negative image samples ("informative negative selection") to tackle the class imbalance problem. For training, the segmented LNs are annotated as positive/negative using radiologist's judgment that combines information from both pathological reports and CT intensities. A LN classification network is trained by using the LN segmentation network as a pre-trained backbone where the classification task benefits from the dense structure prediction in the segmentation loss formulation. To predict LN metastasis in patients with PDAC, we employ a modified ResNet classification model He et al. (2016). Tumor characteristics are proven to be complementary yet important cues (Li et al., 2020; Bian et al., 2019; Gao et al., 2020), so we integrate both tumor and LN deep imaging characteristics by taking as inputs the image patches of tumor and the patient-wise aggregated LNs.

Our main contribution is summarized as follows. (1) To the best of our knowledge, this work is the first to directly exploit and incorporate LN segmentation and identification for explicitly assisting the patient-level pre-operative LN metastasis status prediction. (2) An attention-based LN segmentation network is proposed by enforcing the constraints of distances to surrounding anatomical structures (as exploiting the spatial priors) and addressing the foreground-background imbalance issue. (3) We design a holistic patient-level LN metastasis status prediction network by combining both tumor and positive LN imaging characteristics, revealing the promising performance effects of

joint tumor-LN guided cues. (4) Extensive quantitative experiments are conducted to evidently validate the effectiveness of our proposed deep local attention integration framework in both tasks of instance-level LN segmentation and patient-level LN metastasis status prediction. External multi-center clinical evaluations are performed to further demonstrate our method's generalization ability. Without loss of generality, our proposed method is applicable on calculating the pre-operative LN metastasis status of other solid cancer malignancies, such as liver or gastric cancers. (5) Last but not least, comparing with the current literature in pre-operative lymph node metastasis status assessment, which is one of core unsolved problems in cancer imaging to support more precision clinical decision making, the new ideas we proposed in this paper should inspire others to consider and address this problem from a new technical perspective.

2 Related Work

2.1 Lymph Node Segmentation

Automated LN segmentation in CT images is an essential yet very challenging task in medical image analysis. Traditional methods tackle this problem by the means of atlas based search space restriction (Feuerstein et al., 2012), spatial prior features combination (Liu et al., 2016, 2014), supervoxel clustering (Oda et al., 2017), etc. Recently, U-Net based deep networks have shown remarkable performance in various organ or tumor segmentation tasks (Seo et al., 2019; Huang et al., 2019; Kazemifar et al., 2018; Oktay et al., 2018; Gerard and Reinhardt, 2019). nnUNet (Isensee et al., 2021) is a self-configuring approach, with automatic configurations including preprocessing, network architecture, training and post-processing, and achieves robust performance and applicability. To address the strong class imbalance issues in LN segmentation, four other anatomical structures are included as training targets (Oda et al., 2018) using 3D U-Net (Çiçek et al., 2016) framework. Bouget et al. (2019) utilizes parallel networks of 2D U-Net (Ronneberger et al., 2015) and Mask R-CNN (He et al., 2017) with the supervision of all considered anatomical structures and LNs. Another strategy to incorporate anatomical context is to take organ segmentation masks as additional channels of the input. Bouget et al. (2021) proposes an ensemble approach for a slab-wise and a downsampled full volume based LN segmentation, taking the concatenation of CT image and segmented anatomical structure masks as input. DeepStationing (Guo et al., 2021) presents a key referencing organ auto-search strategy and combines selected organs into the network via input concatenation for LN station parsing. All above methods implicitly exploit spatial priors of LNs by injecting the anatomical structure masks either as inputs or supervisions. Furthermore, there are no studies on how LN segmentation could be used for predicting the patient-level LN metastasis.

2.2 Lymph Node Metastasis Prediction

Radiomics Methods. Radiomics is a powerful technique for extracting quantitative image features with the purpose of clinical decision support, being widely used in cancer research (Kumar et al., 2012; Gillies et al., 2016; Lambin et al., 2012; Gao et al., 2020; Meng et al., 2020; Wei et al., 2023). It converts imaging data into different types of hand-crafted features, including shape, intensity, texture and filter-based (e.g., wavelet, Laplacian of Gaussian) features. Applications of radiomics on predicting LN metastasis from primary tumor have been explored in many previous work (Ji et al., 2019; Wang et al., 2020; Li et al., 2020; Bian et al., 2019; Yang et al., 2020). Radiomics features are first extracted from manually delineated tumor regions in any contrast-enhanced CT images. Feature selection and classification model construction using logistic regression, random forest are then performed to give LN metastasis prediction for various cancers, such as gastric cancer (Wang et al., 2020; Meng et al., 2020), biliary tract cancer (Ji et al., 2019) and PDAC (Li et al., 2020; Bian et al., 2019; Gao et al., 2020). Relying on the primary tumor imaging radiomics without considering LNs may limit the prediction accuracy, thus (Yang et al., 2020) uses manual annotations of the largest LN visible in the gastric region and combines LN radiomics into the prediction model for gastric cancer. However, the main problem remains since it only involves the largest depicted LN without identifying the nodal positivity.

Deep Learning based Methods. Recent advances in deep learning have made it a mainstream approach of formulating the entire workflow of cancer diagnosis and treatment on medical imaging, such as oropharyngeal cancer (Cheng et al., 2021), lung cancer (Xu et al., 2019), as well as pancreatic cancer (Xia et al., 2021; Zhao et al., 2021; Yao et al., 2020). Deep neural networks are applied to solve the problem of LN metastasis in Zheng et al. (2020); Dong et al. (2020); Harmon et al. (2020); Jin et al. (2021). In Zheng et al. (2020), deep features are extracted from tumor ROIs in bimodal image (i.e., US and SWE) using ResNet (He et al., 2016), and then fed into a SVM model for predicting axillary LN status in breast cancer. For gastric cancer, Dong et al. (2020) combines DenseNet Huang et al. (2017) features with some hand-crafted features, extracted from the 2D tumor ROI with the largest area in multi-phase CT images. To estimate the metastasis status of individual LN stations for gastric cancer, Jin et al. (2021) develops a system of multiple independent ResNets with tumor ROIs and corresponding annotation masks as inputs where each ResNet is responsible to predict metastasis at one specific nodal station. Most existing studies capture only tumor characteristics for LN

metastasis prediction, while (Yang et al., 2020) leverages LN radiomics by requiring manual delineation and simply only the LN with the largest size is considered. Automated and accurate LN segmentation and nodal positivity identification is hence of high importance for assisting the patient-level metastasis prediction.

3 Method

The overall framework is illustrated in Fig. 2, which is composed of a distance-guided attention-based LN segmentation and identification network (a), and a tumor and LN combined metastasis status prediction network (b).

3.1 Distance-guided Attention Lymph Node Segmentation and Identification Network

We perform LN detection from any input CT scan by a two-stage strategy: segmenting the image into two classes of LN and background voxels, followed by identifying segmented LN instances as positive or negative.

3.1.1 Class-agnostic Lymph Node Segmentation

Based on the spatial prior that LN stations are geometrically distributed or constrained around certain anatomical structures, we propose an attention based LN segmentation network by taking the distances to nearby organs/vessels into account. Our LN segmentation network differs from the strong baseline (i.e., nnUNet (Isensee et al., 2021)) in that attention mechanism is applied to guide possible LN locations, with the advantage of reducing false positive predictions outside those locations. The intuition behind the attention module is that the attention map can cover regions adjacently constrained to those organs and vessels.

Attention Map Generation. To explicitly capture and model the lymphatic anatomy, attention computation is implemented as a pre-defined geometric mapping function from organ&vessel distance maps. An example of attention map generation process is shown in Fig. 3. Specifically, given a multi-phase input CT volume $X \in \mathbb{R}^{N\times W\times H\times D}$, we first obtain organ&vessel segmentation mask using nnUNet (Isensee et al., 2021) model trained with 19 classes of annotations. Ten classes among them involved with 17 LN stations are used (see Table 6 in the Appendix A for the definition of LN stations), i.e., spleen, esophagus, stomach, aorta, pancreas, duodenum, superior mesenteric artery (SMA), truncus coeliacus and splenic artery (TC&SA), left gastric artery (LGA), common hepatic artery and proper hepatic artery (CHA&PHA). Note that station 15# (LNs along middle colic artery) is left aside here since it is related to distant metastasis that rarely happens in our patient population. A signed distance transform (SDT) is applied to each class of the segmentation mask $M \in \{0, 1, 2, ..., 10\}^{W\times H\times D}$, generating a total of 10 organ/vessel distance maps D^i where $i \in \{1, 2, ..., 10\}$ is the index of organ/vessel class. D^i has positive values at the voxels outside the i-th organ/vessel and negative scores inside it. Intuitively, LNs are likely to appear within a certain range of distance to each organ/vessel, which requires paying attention to. To obtain the distance-guided attention maps, D^i is passed to an isosceles trapezium-shaped non-linear mapping function (see Fig. 3), formulated as

$$f^i(d) = \begin{cases} 1, & d^i_{\min} \le d \le d^i_{\max} \\ -\dfrac{(d - d^i_{\max} - 3)}{3}, & d^i_{\max} < d < d^i_{\max} + 3 \\ \dfrac{(d - d^i_{\min} + 3)}{3}, & d^i_{\min} - 3 < d < d^i_{\min} \\ 0, & Otherwise \end{cases}. \tag{1}$$

where d is the individual element in D^i; $d^i_{\min}$ and $d^i_{\max}$ determine the distance range in mm; the smooth border $3mm$ is chosen empirically. This mapping function converts the distance maps to the attention scores ranging from 0 to 1, with 1 indicating possible locations of LNs, 0 indicating impossible locations, and decimals lying in between. The i-th attention map A^i is obtained by $A^i = f^i(D^i)$.

The final attention map A^{all} is produced by integrating all of the organ/vessel-specific attention maps, thus A^{all} can cover the whole areas that need attending to. In specific, A^{all} takes the element-wise maximum of all A^i except for the voxels inside an organ/vessel, illustrated as

$$a^{all}_{\mathrm{v}} = \begin{cases} \max\limits_{i=1,2,...,10} a^i_{\mathrm{v}}, & m_{\mathrm{v}} = 0 \\ a^i_{\mathrm{v}}, & m_{\mathrm{v}} = i \end{cases}. \tag{2}$$

where a^*_{v} and m_{v} are the values of A^* and M at the voxel v.

Attention based Lymph Node Segmentation. After obtaining the distance-guided attention map, we incorporate it to the segmentation network with 3D nnUNet (Isensee et al., 2021) as the backbone. Specifically, given the multi-phase input images $X \in \mathbb{R}^{N\times W\times H\times D}$, the feature map at the penultimate layer of the network is extracted (denoted as $Z \in \mathbb{R}^{C\times W\times H\times D}$, where C is the channel number of the feature). This feature is then spatially scaled (or multiplied) element-wisely with the attention map $A^{all} \in \mathbb{R}^{W\times H\times D}$. The multiplication operation emphasizes the regions that are likely to contain LNs based on the spatial priors from the attention map. The formula for scaling the features can be written as:

$$Z_S = Z \odot A^{all} \tag{3}$$

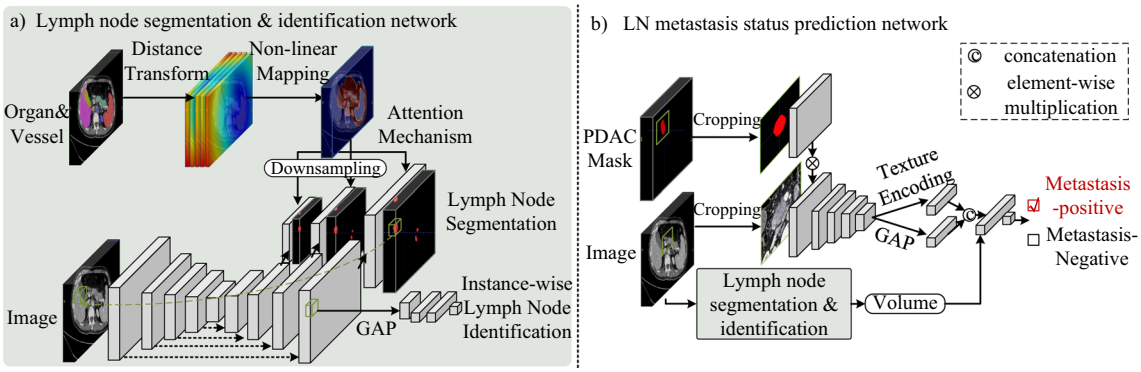
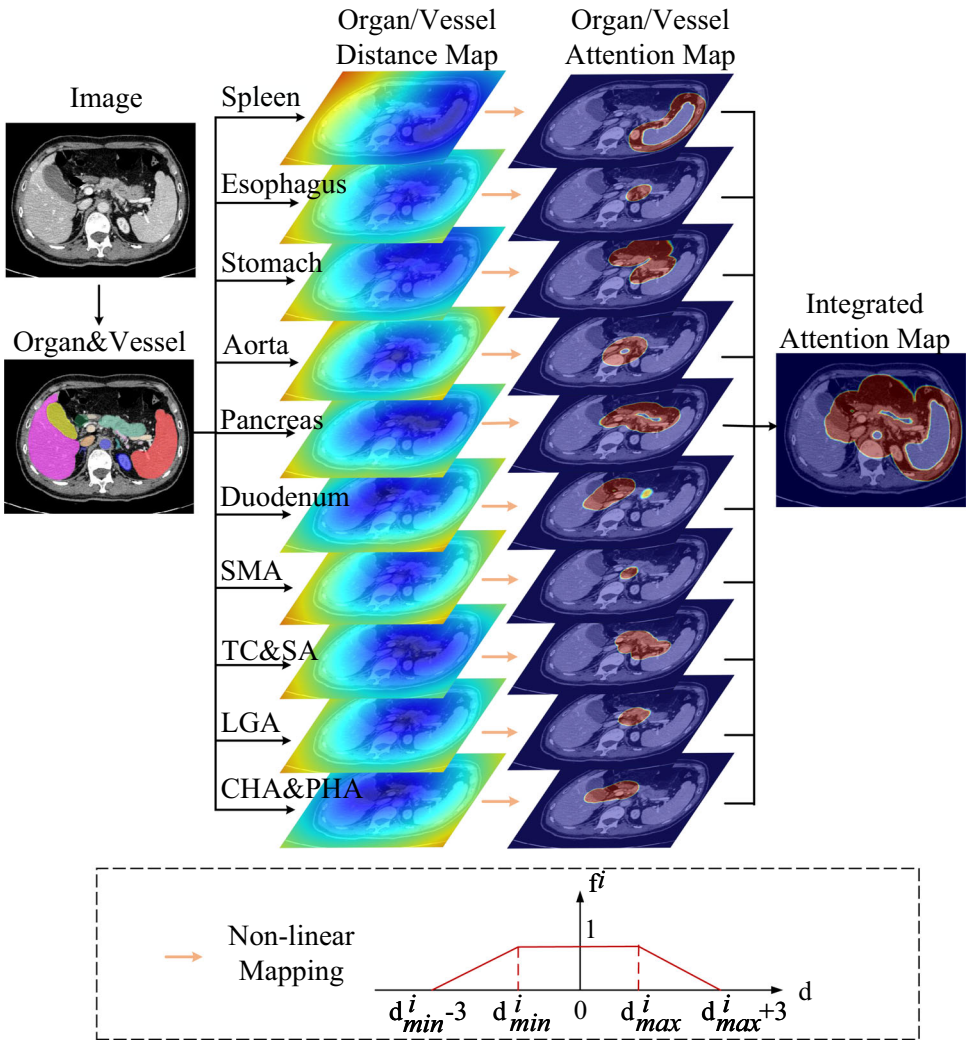


Fig. 2 The proposed framework for **a** two-stage LN segmentation and identification and **b** LN metastasis status prediction. GAP: global average pooling

Fig. 3 An illustration of attention map generation process. SMA: superior mesenteric artery; TC&SA: truncus coeliacus and splenic artery; LGA: left gastric artery; CHA&PHA: common hepatic artery and proper hepatic artery



where $\odot$ represents element-wise multiplication. The scaled feature $Z_S \in \mathbb{R}^{C\times W\times H\times D}$ is finally passed through a convolution block with a softmax layer to produce the segmentation output $P \in \mathbb{R}^{2\times W\times H\times D}$.

Due to GPU memory limitation, patch-based training strategy is employed. nnUNet randomly samples 3D image patches from the whole CT scan and enforces that more than a third of samples in a batch contain at least one foreground class to control the foreground-to-background ratio. Considering the extreme class imbalance problem caused by the small LN targets, we improve it with the "informative negative selection (INS)" scheme. Note that our proposed attention mechanism helps block out features at the voxels with a certain distance to or inside all organs and vessels, resulting in lots of non-informative negative patches filled with 0 by applying zero attention scores. Thus we can naturally throw out those non-informative patches, and select patches containing at least one attention score > 0 (called informative patches) for training. This sampling strategy further boosts the network's concentration on targeted regions of interest (ROIs) surrounding organs/vessels.

For training objectives to better balance precision and recall, we modify the Dice loss in nnUNet with Tversky loss (Salehi et al., 2017):

$$L_T = -\frac{2}{|V|} \times \frac{\sum_{\mathrm{v}} p_{1,\mathrm{v}} y_{1,\mathrm{v}}}{2\sum_{\mathrm{v}} p_{1,\mathrm{v}} y_{1,\mathrm{v}} + \alpha \sum_{\mathrm{v}} p_{1,\mathrm{v}} y_{0,\mathrm{v}} + \beta \sum_{\mathrm{v}} p_{0,\mathrm{v}} y_{1,\mathrm{v}}} \tag{4}$$

where $|V|$ is the number of voxels. $p_{1,\mathrm{v}}$ is the probability of voxel v being a LN, and $p_{0,\mathrm{v}}$ is the probability being a non-LN. Also, $y_{1,\mathrm{v}}$ is 1 for a LN voxel and 0 for a non-LN voxel, and vice versa for the $y_{0,\mathrm{v}}$. In practice, we set $\alpha = 0.5$ and $\beta = 1.5$ to emphasis on false negatives and boost recall. The whole network is trained by the combination of cross entropy loss $\mathcal{L}_{\mathrm{CE}}$ and Tversky loss $\mathcal{L}_{\mathrm{T}}$ with equal weights as in nnUNet.

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{|V|}\sum_{\mathrm{v}}\sum_{\mathrm{k}=0,1} y_{k,\mathrm{v}} \log(p_{k,\mathrm{v}}) \tag{5}$$

$$\mathcal{L}_{\mathrm{SEG}} = \mathcal{L}_{\mathrm{CE}} + \mathcal{L}_{\mathrm{T}} \tag{6}$$

Following nnUNet, the network is trained with deep supervision, i.e., losses are calculated over multi-resolution outputs given by final and intermediate layers of the decoder, and the corresponding downsampled ground-truth (GT) masks are used as targets. Here attention mechanism is applied in a multi-scale manner. That is, the attention map, after downsampled to match the resolution, is injected to the intermediate decoder feature for each deep supervision output.

3.1.2 Instance-wise Lymph Node Identification

After segmenting LN instances from the whole CT image, we then classify them into either positive or negative class. In this context, a positive LN refers to a lymph node that is considered to be malignant, meaning it contains metastasized cancer cells. Conversely, a negative LN refers to a lymph node that is considered to be benign, which does not contain metastasized cancer cells. To benefit from the already trained dense segmentation network of stage 1, the task of LN instance identification reuses 3D nnUNet backbone and is initialized using the trained segmentation parameters, with a new classification head added upon it. Cross entropy loss is adopted to finetune the whole network for classifying the instance as positive/negative. To generate LN instances, we crop patches centered at the connected components of the segmentation mask. GT LN instances are cropped and employed in the training phase. While at inference time, we can apply the classification network to identify each segmented LN of stage 1, and obtain a class-aware LN segmentation mask.

3.2 Patient-wise LN Metastasis Status Prediction Network via Aggregating Tumor and Lymph Node Information

Besides LNs, imaging characteristics in the primary tumor may also play an important role in predicting the LN metastasis status. To further boost the diagnosis performance, we propose a combined classification network, integrating the imaging information from both PDAC and LNs. In contrast to previous work that only consider tumor characteristics (Li et al., 2020; Bian et al., 2019; Gao et al., 2020), our method benefits from directly observing LN instances using automated LN segmentation and identification.

Given a CT image and the corresponding PDAC mask, 2D slices with the top three largest PDAC areas in each of axial, sagittal, and coronal planes are cropped, resulting in nine image patches in total. Each image patch is fed into a ResNet (He et al., 2016) pre-trained on ImageNet (Deng et al., 2009) for metastasis prediction. Inspired by Eppel (2018), a side branch with the PDAC mask as input is added and integrated into the backbone by element-wise multiplication to encourage the network to concentrate on the PDAC region. Our initial experiment empirically shows that such incorporation produces better performance than direct input-level fusion, as the convolution in the side branch learns which region to focus on in each channel of the feature (e.g., regions inside the mask, around the mask border or outside the mask; see Appendix F for feature map visualization). Before classification, we additionally employ a Texture Encoding Layer (TEL) (Zhang et al., 2017) on top of the "Layer4" feature $\mathcal{F}_{L4}$ to extract respective texture representation. The original deep feature after global average pooling (GAP) is concatenated

with the texture feature to form an enhanced representation $\mathcal{F}$.

The LN-related cues are further integrated into the network given the LN segmentation and identification results described in Sect. 3.1. A patient is considered as metastasis-positive if there exists at least one positive LN, thus it is very sensitive to the false positives in LN identification. Therefore, we employ the volume of positive LN as the feature instead of its binary status of presence/absence, based on the fact that positive LNs tend to have larger volume than negative ones. The volume of the largest positive LN in each patient $\mathcal{V}_{\text{pLN}}^{\max} = \max_{ln \in \{positive\ LNs\}} \mathcal{V}_{ln}$ (in mm^3) is mapped to a vector-shaped feature, and fused with $\mathcal{F}$ by element-wise addition, formulated as follows:

$$\mathcal{F}_{\text{comb}} = \text{FC}(\text{BN}(\mathcal{V}_{\text{pLN}}^{\max})) + \mathcal{F} \tag{7}$$

where FC and BN denote the full-connected layer and batch normalization layer, respectively. Other LN features, such as the average or total volume of positive LNs, are also evaluated, with the current setting giving the best result. Finally, the classification probabilities generated from nine image patches are averaged to given an ensembled prediction for a patient.

4 Experiments

In this section, we first describe the multi-center training and evaluation datasets (i.e., the discovery dataset and two external datasets) and implementation details, and elaborate the strategy we use to generate PDAC segmentation masks. All patient LN metastasis status ground truth annotations are curated and created by confirming against their pathology reports. Then we present results on the discovery dataset per each step of our method, including 1) organ&vessel segmentation, 2) attention map generation, 3) LN segmentation and identification, and 4) patient-level LN metastasis status prediction. Last, external validations are conducted to evaluate the generalization performance of our LN metastasis status prediction model.

4.1 Experimental Settings

4.1.1 Datasets

We conduct a multi-center study on three independent datasets with a total of 940 patients collected from Changhai Hospital in Shanghai, Shengjing Hospital in Liaoning Province, and Tianjin Cancer Hospital in Tianjin, China. All patients had a pathologically confirmed diagnosis of PDAC. For each patient, contrast-enhanced CT scans of arterial (A) and venous (V) phases acquired before treatment were included in this study. We treated the dataset from Changhai Hospital as Discovery dataset, and developed our models using nested cross-validation (CV). The rest two datasets from Shengjing Hospital and Tianjin Cancer Hospital (denoted as Ext-validation dataset 1 and Ext-validation dataset 2, respectively) were used as external validation sets where pathologically diagnosed LN metastasis status are provided. This study was reviewed and approved by the Biomedical Research Ethics Committee of the institution (No. CHEC2021164), and was performed in accordance with the ethical standards of the 1964 Declaration of Helsinki. The requirement for patient informed consent was waived by the Institutional Review Board due to the retrospective nature of the study and because all procedures performed were part of routine care.

Discovery dataset contains CT scans of 749 patients, among which there are 351 positive (patients with LN metastasis) and 398 negative samples (patients without LN metastasis). The annotation of LNs was performed by two board-certified radiologists (XF with 7 and MZ with 5 years of experiences specialized in pancreatic imaging) with referring to pathology reports and under the supervision of a senior radiologist (YB) with 17 years of experiences in pancreatic imaging. There are total 2,467 labeled LNs, of which 476 are positive and the rest are negative. Specifically, 351 metastasis-positive patients contain 476 positively and 322 negatively labeled LNs; and 398 metastasis-negative patients have the rest 1,669 negative LNs. This dataset was split using nested five-fold CV, with 64%, 16% and 20% as training, validation and testing sets in each CV round. As for the primary tumor, 163 patients among the whole dataset were annotated with 3D tumor masks by two radiologists (XF and MZ) (see Appendix C for tumor annotation details.). We use these 163 patients as the testing set and the remaining unlabeled 586 patients as the training set for an annotation-efficient PDAC segmentation. While for LN metastasis status prediction, the evaluation was conducted on all 749 cases from the discovery dataset, using the predicted PDAC masks. Additionally, we generate pseudo annotations of 17 classes of organs and vessels using the self-learning segmentation model described in our previous work (Yao et al., 2021), and manually annotate two additional classes of vessels (LGA, CHA&PHA) and extend other two vessels (SMA and TC&SA) under the supervision of a radiologist (XF) for 50 patients randomly sampled from our dataset. 40/10 split of these patients are used as training and validation sets respectively for organ&vessel segmentation evaluation.

Ext-validation dataset 1 contains CT scans of 132 patients with 39 positive and 93 negative patients; **Ext-validation dataset 2** obtains CT scans of 59 patients with 37 positive and 22 negative patients. More detailed information on three datasets can be seen in Table 1.

Table 1 Demographic distributions and tumor characteristics in the three datasets (**Discovery dataset**, **Ext-validation dataset 1** and **Ext-validation dataset 2**). Median [interquartile range, 25th-75th percentile] values are reported for continuous variables

Characteristics	Discovery dataset (n=749)	Ext-validation dataset 1 (n=132)	Ext-validation dataset 2 (n=59)
Gender, n (%)			
Female	282 (38%)	60 (45%)	28 (47%)
Male	467 (62%)	72 (55 %)	31 (53%)
Age at Diagnosis, yrs	63 [56–69]	60 [53–65]	58 [51–62]
pT Stage, n (%)			
pT1 / pT2	92 (12%) / 314 (42%)	24 (18%)/ 80 (61%)	10 (17%) / 31 (53%)
pT3 / pT4	316 (42%) / 13 (2%)	15 (11%)/ 13 (10%)	5 (8%)/ 13 (22%)
Missing	14 (2%)	0 (0 %)	0 (0 %)
pN Stage, n (%)			
pN0	398 (53%)	93 (70%)	22 (37%)
pN1	242 (32%)	32 (24%)	22 (37%)
pN2	109 (15%)	7 (5%)	15 (25%)
Tumor Size, cm	3.0 [2.5−4.1]	2.7 [2.2−3.0]	2.9 [2.2−3.4]
Tumor Location, n (%)			
Head / Uncinate	475 (63%)	56 (42%)/ 52 (39%)	35 (59%)/ 22 (37%)
Body / Tail	274 (37%)	2 (2%)/ 22 (17%)	2 (3%)/ 0 (0%)
Positive LN Volume, mm^3	665[210–804]	–	–
Negative LN Volume, mm^3	300[106–377]	–	–

4.1.2 Implementation Details

In our experiments, CT images of arterial phase are registered to venous phase using DEEDS (Heinrich et al., 2013), all being resampled to a median spacing of 0.68 × 0.68 × 0.80 mm. For LN segmentation and organ&vessel segmentation, sub-volumes of 160 × 192 × 80 voxels are randomly cropped as training patches. In calculating the non-linear mapping from distance maps to attention maps for LN segmentation, the parameters of this mapping function are determined by grouping GT LN voxels according to which organ/vessel is closest to, and calculating the minimum and maximum distances to the organ/vessel boundaries in each group. Parameters are listed in Table 7 in the Appendix B.

For instance-wise LN identification, 3D image training samples are generated by cropping a 96 × 96 × 80 sub-volume centered per each GT LN. SGD optimizer with Nesterov momentum (μ = 0.95) is adopted to train the network; the initial learning rate and weight decay are 5×10^{-4} and 1×10^{-4}, respectively. Furthermore, the final LN metastasis status prediction model takes 2D inputs of 224 × 224 voxels centered at PDAC as well, and is trained using the same optimizer as above. Details of the network architectures are presented in the Appendix D.

4.2 Main Results on Tumor and Lymph Node Segmentation and Identification

PDAC Segmentation Mask Acquisition and Evaluation. We employ an annotation-efficient strategy to generate 3D masks of tumors for the labor cost reduction purpose. Specifically, we start with the PDAC segmentation model trained with arterial-late phase described in our previous work (Zhao et al., 2021) to generate pseudo annotations. Note that the patients used in Zhao et al. (2021) have no overlap with the patients used in this study, and they are two independent cohorts.. Next, this model is fine-tuned under the supervision of pseudo annotations and then applied to produce segmentation masks on our dataset. To obtain the PDAC segmentation model on venous phase, those segmentation masks are registered to venous phase and are then used to train a nnUNet model from scratch to generate the final 3D masks of tumors. We evaluate the obtained PDAC segmentation model on the labeled testing set. Median Dice score, average surface distance (ASD, *mm*), and Hausdorff distance (HD, *mm*) are 0.683, 2.186, and 12.805 respectively.

Evaluation of Organ&Vessel Segmentation and Attention Maps. To evaluate the performance of organ&vessel segmentation, a testing set of 19 randomly selected CT volumes with ten classes of organ/vessel is manually annotated by a radiologist (XF). Dice scores of spleen, esophagus, stomach, aorta, pancreas, duodenum, SMA, TC&SA, LGA,

Table 2 Average instance-wise LN classification performance across 5 folds. The results are reported on GT instances

Metric	Performance
AUC	0.854
Accuracy	0.789
Balanced accuracy	0.771
Sensitivity	0.742
Specificity	0.800

and CHA&PHA are 0.959, 0.745, 0.905, 0.920, 0.847, 0.764, 0.746, 0.734, 0.651, and 0.715, respectively (see Appendix E for evaluation details). Qualitative evaluation of organ&vessel segmentation examples and their corresponding attention maps are visualized in Fig. 4b.

Evaluation of Lymph Node Instance Segmentation and Identification LNs are first detected by the class-agnostic segmentation model, and identified as positive/negative by applying the classification model on the cropped instances. For positive/negative LN identification, our classification model is trained with Ground-Truth (GT) LNs, yielding an average AUC of 0.854 across 5 folds (in Table 2). For inference, the automatically segmented LNs are cropped and identified by the classification model. To evaluate the segmentation performance before and after identification, we compare our method with three baseline models: (1) a strong baseline, vanilla nnUNet (Isensee et al., 2021), (2) nnUNet with a postprocessing step that keeps only segmented LNs where the attention map is nonzero, denoted as $\text{nnUNet}_{\text{pp}}$, and (3) nnUNet with the organ/vessel mask concatenated to the input, denoted as $\text{nnUNet}_{\text{concat}}$. To evaluate the independent effectiveness of the proposed informative negative selection (INS) scheme, we perform an ablation study that omits INS, referred to as ours w/o INS. The segmentation accuracy is measured by voxel-wise metrics (i.e., Dice, Recall, Precision) and instance-wise metrics (F-measure, Recall, Precision). For statistical analysis, we apply 1,000 iterations of Wilcoxon signed rank test to voxel-wise Dice and instance-wise F-measure. Results are provided in Table 3. An instance is considered successfully detected if its (intersect-over-union) IoU score between the segmentation mask and GT mask is ≥ 30 %.

Before identification, our segmentation model significantly outperforms nnUNet, $\text{nnUNet}_{\text{pp}}$ and $\text{nnUNet}_{\text{concat}}$ on both voxel-wise and instance-wise metrics, with the voxel-wise Dice increases of 1.8, 0.4 and 0.5% and the instance-wise F-measure increases of 4.5, 3.0 and 3.3% respectively, as shown in Table 3. Our model also yields superior performance after identification compared to nnUNet, $\text{nnUNet}_{\text{pp}}$ and $\text{nnUNet}_{\text{concat}}$, achieving 1.8, 1.7 and 1.3% higher voxel-wise Dice, and 1.8, 1.6 and 1.5% higher instance-wise F-measure in terms of positive LNs. For negative LNs, our model shows improvements of 0.2, 0.0% (comparable performance) and 0.5% in voxel-wise Dice, and 1.2, 1.0 and 6.6% in instance-wise F-measure. In total five out of six comparisons, our method is statistically significantly better or more accurate (i.e., with p-value < 0.05) in LN segmentation than the vanilla nnUNet baseline (without the attention maps) as well as $\text{nnUNet}_{\text{pp}}$ (with postprocessing) and $\text{nnUNet}_{\text{concat}}$ (with the organ/vessel mask concatenation). Compared with the variant without INS, our model demonstrates a comparable voxel-wise Dice score while improving the instance-wise F-measure by 1.7% before identification. After identification, our model outperforms ours w/o INS in both voxel-wise Dice and instance-wise F-measure for positive LNs. For negative LNs, it shows a superior instance-wise F-measure, despite a lower voxel-wise Dice.

Qualitative Evaluation. Examples of LN segmentation and identification results are shown in Fig. 4a for qualitative comparison. Our segmentation model leverages prior knowledge on LNs' spatial position distribution via incorporating the attention mechanism to remove false positives that are far from anatomically plausible LN areas. In Fig. 4a, we can observe that nnUNet tends to falsely detect an instance inside some organs or located very far, while our method provides noticeably less false positives.

4.3 Evaluation of Patient-wise Lymph Node Metastasis Status Prediction

Metrics. In this section, we evaluate various performance metrics of patient-level LN metastasis status prediction. For this binary classification problem, AUC, accuracy, balanced accuracy, sensitivity and specificity are adopted as evaluation metrics and the average results across 5 folds are reported. Statistical analysis is also carried out to verify the significance of performance improvement or comparison. We collect the predictions of all 5 folds, repeat 1,000 times of bootstrapping for calculating balanced accuracy, and apply Wilcoxon signed rank test to balanced accuracy distributions to compare our method with several other configurations. For comparing ROC curves, DeLong test is performed. P-values < 0.05 are considered as statistically significant. To compute 95% CI, the 2.5th percentile and 97.5th percentile are estimated after 1,000 times of bootstrapping.

Ablation Study. We first investigate the impact of each component in our framework. To evaluate the metastasis prediction performance of LN segmentation and identification, the results can be aggregated into patient-level prediction, based on the definition that a patient with at least one positive LN is metastasis-positive. However, due to a large number of false positives produced by segmentation (LN segmentation in CT images is challenging after all), it will lead to a poor performance if predicting metastasis simply based on the presence of any positive LN in the segmentation results. We

Table 3 Performance comparison on LN segmentation before and after instance-wise identification (denoted as *Class-agnostic Seg* and *Class-aware Seg*). Pos and Neg denote positive and negative LNs. Results are averaged across 5 folds. Wilcoxon signed rank test is conducted on voxel-wise Dice and instance-wise F-measure. * indicates p-value < 0.05. NS indicates no significance. INS indicates the informative negative selection scheme

Stage	Class	Method	Voxel-wise metrics (%)			Instance-wise metrics (%)		
			Dice	Recall	Precision	F-measure	Recall	Precision
	–	nnUNet	45.9*	75.4	36.2	36.1*	81.0	25.3
Class-agnostic Seg		$\text{nnUNet}_{\text{pp}}$	47.3*	75.4	37.8	37.6*	81.0	26.7
(before identification)		$\text{nnUNet}_{\text{concat}}$	47.2*	73.1	37.8	37.3*	78.6	26.4
		Ours w/o INS	**47.7**^{NS}	77.0	**38.2**	38.9*	**82.7**	27.9
		Ours	**47.7**^{ref}	**77.7**	37.5	**40.6**^{ref}	80.9	**29.9**
	Pos	nnUNet	10.2*	32.3	11.3	11.7*	36.1	12.0
		$\text{nnUNet}_{\text{pp}}$	10.3*	32.3	11.5	11.9*	36.1	12.3
		$\text{nnUNet}_{\text{concat}}$	10.7*	38.0	10.8	12.0*	**44.5**	10.7
		Ours w/o INS	11.0*	33.8	**12.6**	12.8*	39.3	**13.3**
Class-aware Seg		Ours	**12.0**^{ref}	**38.9**	11.7	**13.5**^{ref}	41.5	**13.3**
(after identification)	Neg	nnUNet	27.5^{NS}	51.1	25.4	27.7*	60.0	22.9
		$\text{nnUNet}_{\text{pp}}$	27.7^{NS}	51.1	25.9	27.9*	60.0	23.0
		$\text{nnUNet}_{\text{concat}}$	27.2^{NS}	44.5	**27.7**	22.3*	53.5	16.9
		Ours w/o INS	**28.4***	**53.0**	26.3	28.2*	**60.3**	23.5
		Ours	27.7^{ref}	49.0	27.0	**28.9**^{ref}	56.2	**25.8**

Bold values represent the best results

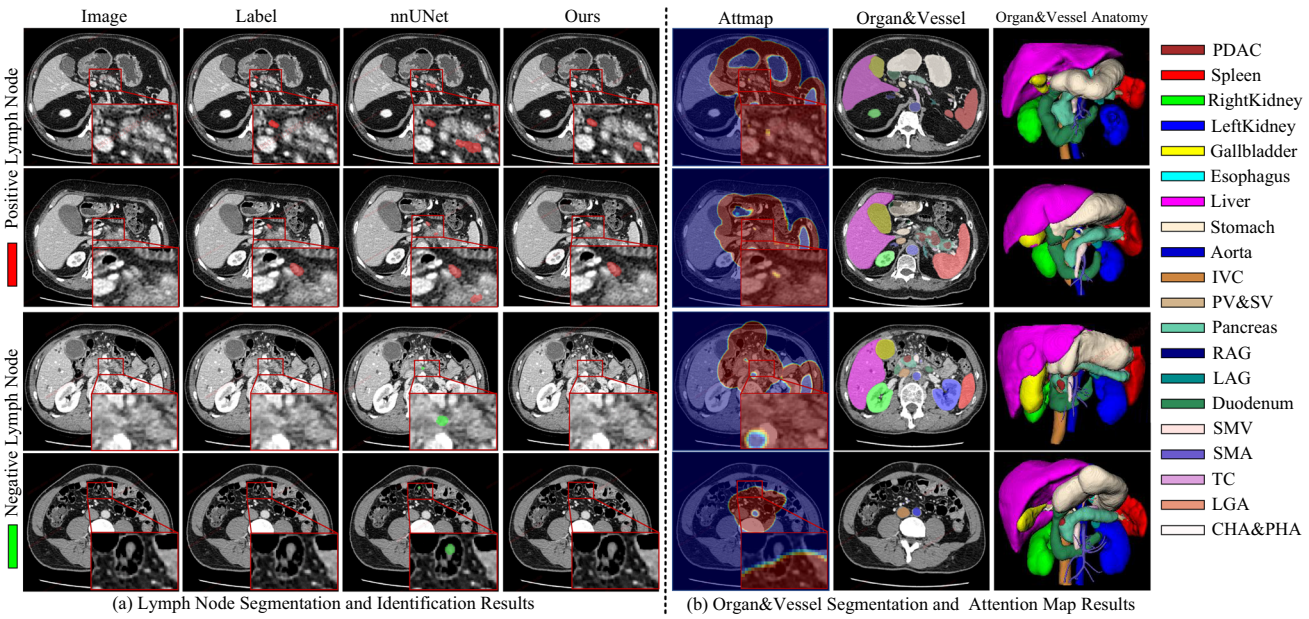


Fig. 4 Examples of **a** LN segmentation and identification results, and **b** Organ&Vessel segmentation and attention map results. PDAC: pancreatic ductal adenocarcinoma; IVC: inferior vena cava; PV&SV: portal vein and splenic vein; RAG: right adrenal gland; LAG: left adrenal gland; SMV: superior mesenteric vein; SMA: superior mesenteric artery; TC: truncus coeliacus; LGA: left gastric artery; CHA&PHA: common hepatic artery and proper hepatic artery

Table 4 Performance comparison and ablation study on Patient-level LN metastasis status prediction of Discovery dataset. Results are averaged across 5 folds. Wilcoxon signed rank test is conducted on balanced accuracy. * indicates p-value < 0.05. NS indicates no significance

Method	Balanced Accuracy [95% CI]	AUC [95% CI]	Accuracy [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]
CT-reported LN status	0.599*	-	0.599	0.588	0.609
	[0.564−0.634]		[0.565−0.634]	[0.538−0.635]	[0.558−0.657]
Radiomics	0.597*	0.648	0.603	0.508	0.686
	[0.563−0.633]	[0.598−0.681]	[0.569−0.637]	[0.456−0.561]	[0.638−0.734]
Radiomics +	0.604*	0.654	0.610	0.524	0.684
CT-reported LN status	[0.572−0.641]	[0.612−0.692]	[0.575−0.644]	[0.470−0.581]	[0.641−0.731]
ResNet3D	0.562*	0.609	0.554	0.599	0.524
	[0.521−0.593]	[0.550−0.631]	[0.519−0.587]	[0.538−0.644]	[0.475−0.568]
ResNet2D	0.571*	0.631	0.574	0.568	0.574
	[0.540−0.609]	[0.590−0.667]	[0.537−0.607]	[0.519−0.624]	[0.530−0.628]
DeepTEN	0.588*	0.634	0.593	0.609	0.566
	[0.560−0.628]	[0.599−0.679]	[0.559−0.628]	[0.564−0.667]	[0.520−0.621]
ClsfromPDAC	0.599*	0.654	0.597	0.600	0.597
	[0.558−0.634]	[0.608−0.685]	[0.561−0.633]	[0.547−0.647]	[0.550−0.646]
ClsbyLNSeg w/o Attn	0.545*	0.590	0.566	0.433	0.657
	[0.525−0.593]	[0.548−0.625]	[0.534−0.601]	[0.393−0.499]	[0.623−0.716]
ClsbyLNSeg w/ Attn	0.563*	0.603	0.572	0.351	**0.775**
	[0.530−0.594]	[0.564−0.642]	[0.539−0.605]	[0.299−0.396]	[0.731−0.814]
LNSegRadiomics	0.544*	0.547	0.547	0.535	0.554
	[0.509−0.584]	[0.504−0.590]	[0.509−0.583]	[0.483−0.589]	[0.504−0.604]
ClsfromPDAC +	0.603*	0.662	0.601	**0.638**	0.568
LNSegRadiomics	[0.557−0.650]	[0.602−0.709]	[0.553−0.649]	[0.571−0.712]	[0.504−0.631]
ClsfromPDAC +	0.616*	0.666	0.619	0.555	0.678
ClsbyLNSeg w/o Attn	[0.585−0.653]	[0.624−0.707]	[0.584−0.650]	[0.502−0.609]	[0.638−0.726]
Ours (ClsfromPDAC +	**0.633**^{ref}	**0.682**	**0.635**	0.618	0.649
ClsbyLNSeg w/ Attn)	[0.599−0.669]	[0.640−0.717]	[0.601−0.669]	[0.567−0.664]	[0.603–696]

Bold values represent the best results

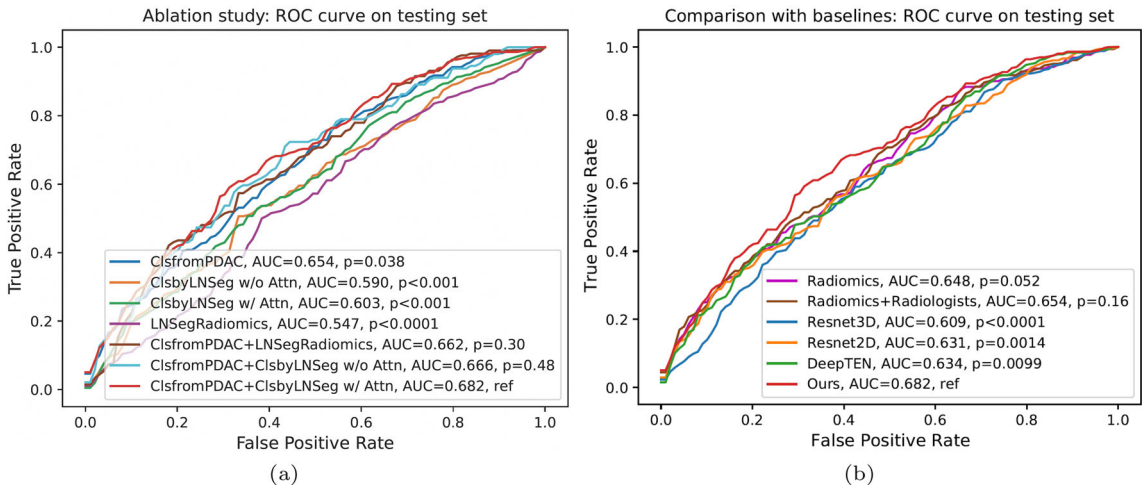


Fig. 5 ROC curve comparison of **a** ablation study and **b** baseline models and our method using nested five-fold cross-validation in Discovery dataset

instead conduct ROC analysis on the volume of the largest positive LN in each case, and find an optimal threshold with the best balanced accuracy in the validation set. Then this threshold is applied to the testing set. A patient with positive LNs larger than the threshold is classified into metastasis-positive; otherwise, it is classified as metastasis-negative. The five ablation models for comparison are listed below:

- ClsfromPDAC (1): The straightforward strategy combining ResNet2D (He et al., 2016) feature and DeepTEN (Zhang et al., 2017) feature, extracted from PDAC slices, in the input of the classification layer.
- ClsbyLNSeg w/o Attn (2): Patient-level metastasis aggregation from the results of LN segmentation without attention (i.e. nnUNet).
- ClsbyLNSeg w/ Attn (3): Patient-level metastasis aggregation from the results of our proposed LN segmentation with attention.
- LNSegRadiomics (4): Radiomics model using radiomics features of the largest positive LN generated by our proposed LN segmentation with attention.
- ClsfromPDAC + LNSegRadiomics (5): Combined model incorporating radiomics features of the largest segmented positive LN into the classification layer of ClsfromPDAC.
- ClsfromPDAC + ClsbyLNSeg w/o Attn (6): Combined model incorporating the volume of the largest positive LN given by ClsbyLNSeg w/o Attn into the classification layer of ClsfromPDAC.
- ClsfromPDAC + ClsbyLNSeg w/ Attn (7): Combined model incorporating the volume of the largest positive LN given by ClsbyLNSeg w/ Attn into the classification layer of ClsfromPDAC.

The results of the ablation experiments are summarized in Table 4, and ROC analysis is illustrated in Fig. 5a. By using only information about LNs, ClsbyLNSeg w/ Attn gives better aggregation results compared with ClsbyLNSeg w/o Attn and LNSegRadiomics (balanced accuracy 0.563 versus 0.545 and 0.544). Integrating tumor characteristics with LN-related information consistently improves the performance, with the tumor-combined models of LNSegRadiomics, ClsbyLNSeg w/o Attn and ClsbyLNSeg w/ Attn outperforming their standalone counterparts by margins of 0.059, 0.071, and 0.070 respectively *Notably, our final model (ClsfromPDAC + ClsbyLNSeg w/ Attn) significantly outperforms the other two tumor-combined models (ClsfromPDAC + LNSegRadiomics and ClsfromPDAC + ClsbyLNSeg w/o Attn) with a balanced accuracy of* **0.633** *(p-value* < 0.05*) versus* **0.603** *and* **0.616**. *These findings highlight the efficacy of integrating both tumor and LNs imaging information for patient-level metastasis status prediction.*

Comparison with Baselines. To validate the effectiveness of our method, radiomics model (Li et al., 2020) and 2D/3D deep classification models are taken for comparison. To build the radiomics model, 1688 radiomics features of PDAC for each CT phase are extracted using Pyradiomics package (Van Griethuysen et al., 2017),[1] and shrunk using the least absolute shrinkage and selection operator (LASSO) method. Then a logistic regression model is applied to the selected features. The combined model of radiomics and CT-reported LN status is implemented with a logistic regression model on radiomics signature and radiologists' diagnosis. For 2D deep networks, ResNet2D (He et al., 2016) and DeepTEN (Zhang et al., 2017), we use ResNet-18 backbone pre-trained on ImageNet (Deng et al., 2009); while for 3D deep networks, we adopt 3D-ResNet-18 (Hara et al., 2018) backbone pre-trained on Kinetics-700 (Kay et al., 2017) and Moments in Time (Monfort et al., 2019). In all 2D/3D deep networks, a side branch with the PDAC mask as input is added to the backbone, as we implemented in our method, for fair comparison. Table 4 and Fig. 5b present the quantitative results of different baseline models. *Importantly, our method (ClsfromPDAC + ClsbyLNSeg w/ Attn) yields the best balanced accuracy* (**0.633**) *and AUC* (**0.682**) *among all compared models, and is significantly better than the radiomics method* (**0.597**, *or* **0.648**) *and all of 2D/3D deep networks* (**0.562**∼**0.588**, *or* **0.609**∼**0.634**).

4.4 External Validation

In this section, we demonstrate the generalization ability of our LN metastasis status prediction in two external multi-center datasets (Ext-validation dataset 1 and Ext-validation dataset 2). After training the model on Discovery dataset using nested 5-fold cross validation, we apply the obtained models to external datasets for inference. For each patient, the ensemble prediction of five models is generated by averaging the model predictions from five folds. We first evaluate the performance of ablation variants, then compare our method with baseline models. Metrics are used the same as Sect. 4.3.

Ablation Study. We conduct ablation study on two external datasets, and results are shown in Table 5. With respect to LN metastasis status prediction using only LN-related information, our method (ClsbyLNSeg w/ Attn) outperforms ClsbyLNSeg w/o Attn and LNSegRadiomic on both two datasets: balanced accuracy 0.589 versus 0.579 and 0.564 on Ext-validation dataset 1, 0.639 versus 0.607 and 0.596 on Ext-validation dataset 2. By integrating PDAC characteristics, our final model (ClsfromPDAC + ClsbyLNSeg w/ Attn) gives the best results among all ablation models: balanced accuracy **0.620** on Ext-validation dataset 1 and **0.684** on Ext-validation dataset 2.

Comparison with Baselines. Table 5 validates the generalization performance of our method compared with

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

Table 5 Performance comparison and ablation study on Patient-level LN metastasis status prediction of two external datasets. Predictions are averaged across 5 folds. Wilcoxon signed rank test is conducted on balanced accuracy. * indicates p-value < 0.05. NS indicates no significance

Dataset	Method	Balanced Accuracy [95% CI]	AUC [95% CI]	Accuracy [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]
Ext-validation dataset 1	Radiomics	0.493* [0.451−0.537]	0.511 [0.400−0.620]	0.672 [0.626−0.710]	0.051 [0.000−0.128]	**0.935** [0.880−0.978]
	ResNet3D	0.508* [0.415−0.612]	0.509 [0.409−0.617]	0.527 [0.450−0.611]	0.462 [0.308−0.615]	0.554 [0.457−0.663]
	ResNet2D	0.563* [0.470−0.656]	0.564 [0.460−0.676]	0.542 [0.466−0.626]	0.615 [0.462−0.769]	0.511 [0.413−0.609]
	DeepTEN	0.556* [0.467−0.647]	0.557 [0.454−0.666]	0.511 [0.427−0.595]	**0.667** [0.513−0.795]	0.446 [0.348−0.544]
	ClsfromPDAC	0.515* [0.423−0.609]	0.554 [0.450−0.661]	0.485 [0.402−0.568]	0.590 [0.436−0.744]	0.441 [0.344−0.548]
	ClsbyLNSeg w/o Attn	0.579* [0.498−0.662]	0.555 [0.454−0.641]	0.689 [0.621−0.750]	0.308 [0.179−0.462]	0.849 [0.774−0.914]
	ClsbyLNSeg w/ Attn	0.589* [0.511−0.672]	0.580 [0.474−0.694]	**0.705** [0.644−0.765]	0.308 [0.154−0.462]	0.871 [0.796−0.935]
	LNSegRadiomics	0.564* [0.478−0.659]	0.542 [0.442−0.654]	0.598 [0.515−0.682]	0.475 [0.325−0.625]	0.652 [0.554−0.750]
	ClsfromPDAC + LNSegRadiomics	0.581* [0.486−0.671]	0.566 [0.455−0.675]	0.583 [0.500−0.659]	0.575 [0.425−0.725]	0.587 [0.489−0.674]
	ClsfromPDAC + ClsbyLNSeg w/o Attn	0.607* [0.523−0.695]	0.599 [0.497−0.701]	0.659 [0.583−0.735]	0.475 [0.325−0.625]	0.739 [0.652−0.826]
	Ours (ClsfromPDAC + ClsbyLNSeg w/ Attn)	**0.620**^{ref} [0.538−0.713]	**0.603** [0.498−0.712]	0.674 [0.598−0.742]	0.487 [0.333−0.641]	0.753 [0.667−0.839]
Ext-validation dataset 2	Radiomics	0.508* [0.391−0.626]	0.609 [0.452−0.757]	0.441 [0.339−0.542]	0.243 [0.108−0.378]	0.773 [0.591−0.909]
	ResNet3D	0.461* [0.334−0.584]	0.442 [0.300−0.593]	0.475 [0.356−0.594]	0.514 [0.351−0.676]	0.409 [0.182−0.591]
	ResNet2D	0.681* [0.536−0.810]	0.687 [0.508−0.849]	0.650 [0.500−0.800]	0.577 [0.385−0.731]	0.786 [0.571−0.930]
	DeepTEN	0.613* [0.465−0.747]	0.647 [0.474−0.806]	0.640 [0.520−0.760]	**0.697** [0.545−0.848]	0.529 [0.294−0.765]
	ClsfromPDAC	0.620* [0.493−0.734]	0.639 [0.490−0.781]	0.593 [0.475−0.712]	0.514 [0.378−0.676]	0.727 [0.545−0.909]
	ClsbyLNSeg w/o Attn	0.607* [0.503−0.716]	0.690 [0.554−0.818]	0.542 [0.441−0.661]	0.351 [0.216−0.487]	**0.864** [0.682−1.000]
	ClsbyLNSeg w/ Attn	0.639* [0.525−0.752]	0.695 [0.552−0.833]	0.593 [0.475−0.695]	0.459 [0.297−0.622]	0.818 [0.636−0.955]
	LNSegRadiomics	0.596* [0.464−0.712]	0.625 [0.457−0.778]	0.576 [0.457−0.695]	0.526 [0.368−0.684]	0.667 [0.476−0.857]
	ClsfromPDAC + LNSegRadiomics	0.647* [0.514−0.763]	0.647 [0.494−0.782]	0.627 [0.508−0.746]	0.579 [0.421−0.737]	0.714 [0.524−0.905]
	ClsfromPDAC + ClsbyLNSeg w/o Attn	0.654* [0.530−0.755]	0.700 [0.558−0.818]	0.610 [0.491−0.729]	0.500 [0.342−0.658]	0.810 [0.619−0.952]
	Ours (ClsfromPDAC + ClsbyLNSeg w/ Attn)	**0.684**^{ref} [0.570−0.797]	**0.703** [0.554−0.846]	**0.661** [0.542−0.780]	0.595 [0.432−0.757]	0.773 [0.591−0.909]

Bold values represent the best results

radiomics and 2D/3D deep learning models. Note that we skip methods involved with CT-reported LN status since there is no CT report available in two external datasets. The radiomics model shows poor generalization ability with large drops in performance as compared with that in Table 4 where the balanced accuracy value decreases from **0.597** in discovery, to only **0.493** in Ext-validation 1 and **0.508** in Ext-validation 2 dataset, respectively. Deep learning methods are comparably more robust. Notably, our method demonstrates reduced sensitivity when applied to the external datasets as compared to the discovery dataset. This discrepancy can be attributed to variations in the distribution of tumor size, tumor stage, and tumor location between the internal and external datasets (refer to Table 1). Specifically, the external datasets contains patients with smaller median tumor sizes, a higher proportion of patients with early-stage tumors (T1/T2) and a greater prevalence of tumors located in the head and uncinate regions of the pancreas. Despite this, our method significantly surpasses all of 2D/2D deep learning models (with p-value < 0.05) on both external datasets, achieving the best balanced accuracy of **0.620** and **0.684**; and the best AUC score of **0.603** and **0.703** for Ext-validation 1 and Ext-validation 2 dataset, respectively. This demonstrates the power of our model to generalize well across patient populations of different clinical sites.

5 Discussion

Pre-operative LN metastasis status prediction is of vital significance for patients with PDAC for three main reasons. **(1)** If diagnosed with LN metastasis, patients with resectable PDAC are recommended to receive neoadjuvant therapy first before surgery, according to NCCN guidelines (Tempero et al., 2021). **(2)** Pancreatectomy could be guided by whether and where their LNs have metastasized, that is, whether or not a standard or an extended lymphadenectomy should be performed. This could make the surgical procedure being more targeted beforehand which could lead to better patient outcome and avoid over-treatment. **(3)** LN metastasis is highly associated with patients' overall survival, which can significantly assist with good prognosis prediction value (Yao et al., 2021). Note that it is very time consuming and highly dependent on (board-certified radiologist) human observer's experience to manually determine whether a patient has LN metastasis primarily from CT scans (even it is indeed a very desirable task to perform for the benefits of patient care). CT-reported LN status in this study shows limited performance with an accuracy of 0.599, thus accurate computerized LN metastasis status prediction is highly needed.

In the literature, LN metastasis status prediction is predominantly studied using tumor feature extraction, combined with CT report information, under radiomics (Ji et al., 2019; Wang et al., 2020; Li et al., 2020; Bian et al., 2019; Yang et al., 2020; Gao et al., 2020; Meng et al., 2020) or deep learning approaches (Jin et al., 2021; Dong et al., 2020). Leveraging LN radiomics requires manual delineation and only the LN with the largest size is considered (Yang et al., 2020). An automated accurate process of LN segmentation and nodal positivity identification is hence of high importance for assisting metastasis prediction. Predicting the metastasis status from segmented LNs is formulated by detecting metastatic LNs with Faster R-CNN (Lu et al., 2018), where the spatial context priors towards LNs are not exploited. Our work proposes an automated geometric attention mechanism using LN segmentation and identification to further predict the patient-level status of LN metastasis. Note that our method is fully automated for the purpose of LN metastasis status prediction during inference, relying solely on automatically generated PDAC masks and CT images without requiring any manual input.

To demonstrate the effectiveness of our method, we provide extensive quantitative experiments on LN segmentation/identification and LN metastasis status prediction. **1)** For instance-level positive/negative LN identification, our classification model (Fig. 2a) yields an average AUC of **0.854** in 5-fold nested Cross Validation (in Table 2). **The metastasized LN instances can be flagged or identified at the Sensitivity of 0.742 and Specificity of 0.800 which should be very helpful to assist radiologists. Human readers tend to have significantly lower detection sensitivity on this task if performed alone.** Our LN segmentation model statistically significantly outperforms the strong baseline nnUNet in voxel-wise and instance-wise metrics. For LN instance-wise detection, our model achieves considerable quantitative improvements (4.6%) in precision (with respect to a similar recall level) as compared to nnUNet (see Table 3). **This observation clearly validates that the proposed distance-guided attention mechanism is beneficial to remove LN false positive detections as we expect. The effectiveness of our model can be attributed to its attention map design and the informative negative selection scheme.** The former confines LN-plausible regions that deserve the network's focus and the latter helps to throw out non-informative negative training patches out of training. **2)** To verify the effect of LN detection improvements on patient-level metastasis status prediction, we perform instance-wise positivity identification and holistic patient-wise aggregation on detected LN instances to classify each patients into metastasis-positive or -negative. Our model presents better prediction performance than nnUNet w/o Attn (balanced accuracy 0.563 versus 0.545, in Table 4). We draw the following observations.

(i) **Combining both tumor and LN CT imaging characteristics into our final prediction model achieves sta-**

tistically significant performance gains compared to radiomics methods and other deep 2D/3D model variants (see Table 4),showing the effectiveness of integrating tumor morphology and lymphatic anatomy.

(ii) It is worth noting that our method obtains statistically significantly better performance (balanced accuracy 0.633 versus 0.604) than the approach where even radiologists are involved ("Radiomics + CT-reported LN status" in Table 4). Using our method, the time-consuming, and highly subjective manual process of CT-reported LN status may be alleviated and improved.

(iii) External multi-center clinical validation is conducted on patient datasets from two other hospitals. The results evidently exhibit our superior performance accuracy and generalization with the best results (balanced accuracy 0.620 and 0.684 on the two external datasets, respectively) reported among several compared models (see Table 5). With all experiments, our model reports better generalized prediction performance (0.633 in training and 0.620–0.684 in external validation) on multi-center datasets and robust improvements over radiomics (0.493–0.508 in validation) and deep learning models.

Recent progress in gastric cancer (Meng et al., 2020) enrolls over 500 patients from multiple hospitals, and yields AUC scores of 0.615–0.712 in validation using 2D/2.5D/3D radiomics features and under different patient splits, which is probably suitable to serve as reference baseline for our work. We employ 940 patients in total in this study, in which 749 patients are from a high-volume pancreatic cancer clinical center and 191 are from two external centers (AUC scores of 0.682 in cross-validation, and 0.603–0.793 in external validation). The studied patient population is arguably much closer and more realistic to the real-world patient data distributions. Additionally, we present a very promising approach that explicitly explores the role of automated LN segmentation in promoting LN metastasis status prediction to facilitate future clinical adoption as a fully-automated and generalizable clinical tool.

The limitations of our method are as follows: (1) The empirical d_{min} and d_{max} determined for attention map generation might possibly limit the adaptability of our method across different patient populations. One potential direction to address this issue would be to explore methods for automatically estimating patient-specific anatomical scales based on imaging data, and then using these scales to normalize the d_{min} and d_{max} values; (2) We extract tumor and LNs imaging information separately and then integrates them by feature concatenation. Our framework can be further improved by designing an enriched deep learning geometric network representation to encode the tumor-LN topology information and spatial anatomical interactions, by modeling the clinical pathways of nodal metastasis explicitly. Last, without loss of generality, our proposed method should be applicable on predicting the pre-operative LN metastasis status of other solid tumors or cancers, such as liver or gastric cancers.

6 Conclusion

We present an attention based LN segmentation network and utilize it on predicting the LN metastasis status in patients with PDAC. The proposed LN segmentation network involves an attention mechanism that encourages the network to focus on regions geometrically confined by certain anatomical organs/vessels. It outperforms the strong baseline nnUNet by leveraging the context information of surrounding anatomical structures. Our segmentation model, followed by a nodal positivity identification model, can serve as a single predictor for LN metastasis. Combined with tumor imaging characteristics, we further build a holistic patient-level LN metastasis status prediction model that is validated to surpass the performance of CT-reported results, radiomics based method, and other 2D/3D deep learning models. Further investigations include conceiving a more sophisticated way to determine the parameters for attention map generation and encode the tumor-LN interaction relationship, and exploring its applications to prognosis and treatment planning in cancer patient management.

Appendix A LN station Definition

The definition of 18 LN stations is shown in Table 6.

Table 6 The definition of 18 LN stations according to JPS (Kanehara, 2017)

Station No.	Name
1 #	Right cardial LNs
2 #	Left cardial LNs
3 #	LNs along the lesser curvature of the stomach
4 #	LNs along the greater curvature of the stomach
5 #	Suprapyloric LNs
6 #	Infrapyloric LNs
7 #	LNs along left gastric artery
8 #	a. LNs in the anterosuperior group along common hepatic artery
	p. LNs in the posterior group along common hepatic artery

Table 6 continued

Station No.	Name
9 #	LNs around celiac artery
10 #	LNs at the splenic hilum
11 #	p. LNs along the proximal splenic artery
	d. LNs along the distal splenic artery
12 #	a. LNs along the hepatic artery
	p. LNs along portal vein
	b. LNs along the bile duct
13 #	a. LNs on the posterior aspect of the superior portion of the head of the pancreas
	b. LNs on the posterior aspect of the inferior portion of the head of the pancreas
14 #	p. LNs along the proximal superior mesenteric artery
	d. LNs along the distal superior mesenteric artery
15 #	LNs along middle colic artery
16 #	a1. LNs around the aortic hiatus of the diaphragm
	a2 & b1 & b2. LNs around the abdominal aorta
17 #	a. LNs on the anterior surface of the superior portion of the head of the pancreas
	b. LNs on the anterior surface of the inferior portion of the head of the pancreas
18 #	LNs along the inferior margin of the pancreas

Appendix B Non-linear Mapping Parameters

To determine the non-linear mapping parameters for attention map generation, we visualize the distance distribution of GT LN voxels to the closest organs/vessels, as shown in Fig. 6 These histograms show a right-skewed distribution, indicating that most LNs are located near their associated organs or vessels. Nonetheless, the tail in each histogram suggests substantial variability in distances, making it necessary to define specific ranges for each organ/vessel (from d_{min} to d_{max}) to ensure comprehensive coverage. Parameters (i.e. d_{min} and d_{max}) of non-linear mapping function for each organ or vessel are shown in Table 7.

Table 7 Parameters of non-linear mapping function for each organ or vessel. Negative values may indicate that voxels are inside organ/vessel. SMA: superior mesenteric artery. TC&SA: truncus coeliacus and splenic artery. LGA: left gastric artery; CHA&PHA: common hepatic artery and proper hepatic artery

Organ/Vessel	d_{min} (mm)	d_{max} (mm)
Spleen	0	16
Esophagus	0	25
Stomach	-2	18
Aorta	0	28
Pancreas	-5	20
Duodenum	-5	22
SMA	-1	20
TC&SA	-2	18
LGA	0	21
CHA&PHA	0	20

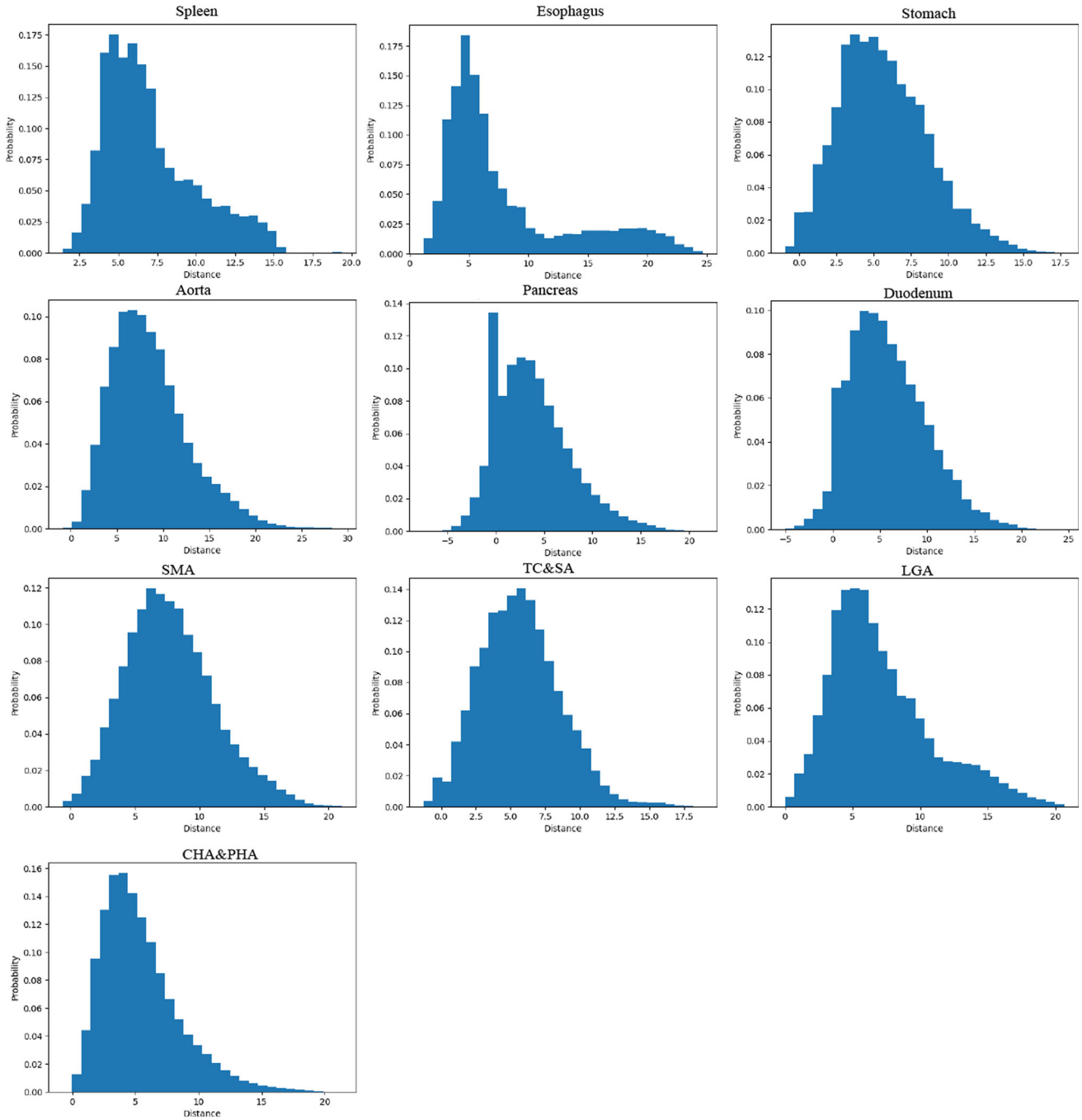


Fig. 6 Distance distribution of GT LN voxels to the closest organs/vessels

Appendix C Tumor Annotation Details

For the annotation of 3D tumor masks in the 163 patients, both radiologists (XF and MZ) were involved in labeling each patient's tumor mask. The annotation and review process comprised three key steps:

(i) **Independent Annotation:** Both radiologists independently annotated the tumor masks without consulting each other's annotations. This allowed us to capture the individual interpretations and judgments of each radiologist.

(ii) **Consensus Review:** After the initial independent annotation, a consensus review was performed under the supervision of a senior radiologist (YB) with 17 years of experience in pancreatic imaging. During this review, any discrepancies between the two radiologists' anno-

tations were resolved through discussion, with the senior radiologist providing guidance where needed.

(iii) **Final Annotation:** The final tumor masks used in our study were obtained based on this consensus review, ensuring that any inter-rater variability was minimized and that the annotations reflected the most accurate and clinically meaningful interpretations.

Appendix D Network Architecture

The LN segmentation network consists of an encoder with six convolution blocks and a decoder with five convolution blocks. Each convolution blocks is composed of two $3 \times 3 \times 3$ convolution layers followed by instance normalization and leaky ReLU activation. In the encoder, the first block uses convolution with stride of 1 and channel number of 32, and the rest five blocks down-samples the feature map by using stride of 2 in the first convolution of the block and the channel number is doubled in the subsequent blocks. In the decoder, each feature map is first up-sampled by deconvolution with stride of 2 and kernel size of 2, and then passed to the convolution block with stride of 1 and the halved channel number. Finally, the multi-resolution outputs are generated by add two $1 \times 1 \times 1$ convolution layers to the outputs of every decoder blocks.

The instance-wise LN identification network shares the same backbone as LN segmentation, and a classification head on the output of final decoder block. The head consists of a GAP layer followed by a fully-connected layer with 32 units, a leaky ReLU layer and a fully-connected layer with 2 units.

The LN metastasis status prediction status model adopts ImageNet-pre-trained ResNet18 (He et al., 2016) backbone. The side branch is composed of a 3×3 convolution layer followed by a ReLU layer, and maps the input mask to a feature map with the same shape as the output of "Conv1" layer in ResNet. It is then integrated into the backbone by element-wise multiplication with the "Conv1" feature. To be better aligned with the pre-trained backbone and eliminate the initial effect of the side branch, the weights and biases in the convolution layer are initialized to 0 and 1 respectively. Texture Encoding Layer (TEL) serves as an orderless feature pooling layer that encodes spatially invariant representation describing the feature distributions, which benefits texture recognition of the PDAC region. The dimension and number of codewords in TEL are set to 64 and 8 respectively, giving a texture feature of 512 units. It is concatenated with the original ResNet18 feature after the GAP layer to form the enhanced feature of 1024 units.

Appendix E Evaluation of Organ&Vessel Segmentation and Attention Maps

To evaluate the performance of organ&vessel segmentation, a testing set of 19 randomly selected CT volumes with ten classes of organ/vessel is manually annotated by a radiologist (XF). To reduce the annotation burden, all CT volumes are downsampled to 5mm in the slice thickness dimension. We compare our self-learning model with the pseudo annotation generator (Yao et al., 2021), which is able to segment eight of ten classes (except for LGA and CHA&PHA) on single-phase CT. Dice score, ASD (*mm*), and HD (*mm*) are adopted as the evaluation metrics and the results are provided in Table 8. Our model that is trained on two phases outperforms the results of Yao et al. (2021) on seven of eight organs/vessels. Note that SMA and TC&SA masks segmented by Yao et al. (2021) contain shorter parts compared with those segmented by our model, which resulting in significantly lower performance than ours (e.g., 0.331 lower Dice score in SMA, and 0.171 lower in TC&SA).

Table 8 Quantitative Performance of Organ&Vessel Segmentation. A: arterial. V: venous. SMA: superior mesenteric artery. TC&SA: truncus coeliacus and splenic artery. LGA: left gastric artery; CHA&PHA: common hepatic artery and proper hepatic artery

Organ/vessel	Methods	CT Phases	Dice	ASD (*mm*)	HD (*mm*)
Spleen	(Yao et al., 2021)	A	0.938	0.643	14.252
	(Yao et al., 2021)	V	0.954	0.422	11.107
	Ours	A+V	**0.959**	**0.384**	**8.129**
Esophagus	(Yao et al., 2021)	A	0.557	0.936	13.897
	(Yao et al., 2021)	V	0.598	0.854	11.003
	Ours	A+V	**0.745**	**0.641**	**8.125**
Stomach	(Yao et al., 2021)	A	0.846	2.223	35.338
	(Yao et al., 2021)	V	0.813	3.765	43.114
	Ours	A+V	**0.905**	**1.519**	**19.183**
Aorta	(Yao et al., 2021)	A	0.893	0.519	8.130
	(Yao et al., 2021)	V	**0.924**	0.417	6.158
	Ours	A+V	0.920	**0.359**	**5.863**
Pancreas	(Yao et al., 2021)	A	0.712	2.905	25.880
	(Yao et al., 2021)	V	0.756	1.897	19.258
	Ours	A+V	**0.847**	**0.975**	**12.859**
Duodenum	(Yao et al., 2021)	A	0.613	2.976	34.187
	(Yao et al., 2021)	V	0.665	3.366	32.174
	Ours	A+V	**0.764**	**1.892**	**29.131**
SMA	(Yao et al., 2021)	A	0.387	**0.663**	68.869
	(Yao et al., 2021)	V	0.415	0.710	68.098
	Ours	A+V	**0.746**	0.860	**28.840**
TC&SA	(Yao et al., 2021)	A	0.563	0.780	43.974
	(Yao et al., 2021)	V	0.407	1.245	51.432
	Ours	A+V	**0.734**	**0.305**	**22.224**
LGA	(Yao et al., 2021)	A	–	–	–
	(Yao et al., 2021)	V	–	–	–
	Ours	A+V	0.651	0.371	10.420
CHA&PHA	(Yao et al., 2021)	A	–	–	–
	(Yao et al., 2021)	V	–	–	–
	Ours	A+V	0.715	1.424	24.239

Bold values represent the best results

Appendix F Visualization of the Side Branch in LN Metastasis Status Prediction Network

In the LN metastasis status prediction network, the convolution in the side branch learns to focus on specific regions of the PDAC mask when integrated into the backbone network by element-wise multiplication. To illustrate the convolution learning process, we provide visual results in Fig. 7 to show the feature map of each channel learned by the convolution layer in the side branch. Each feature map is first normalized to a range of 0 to 1 and then overlaid on the original image for better visualization. These visualizations confirm that the convolution layer effectively learns and highlights specific regions inside the mask, around the mask border, or outside the mask in each channel.

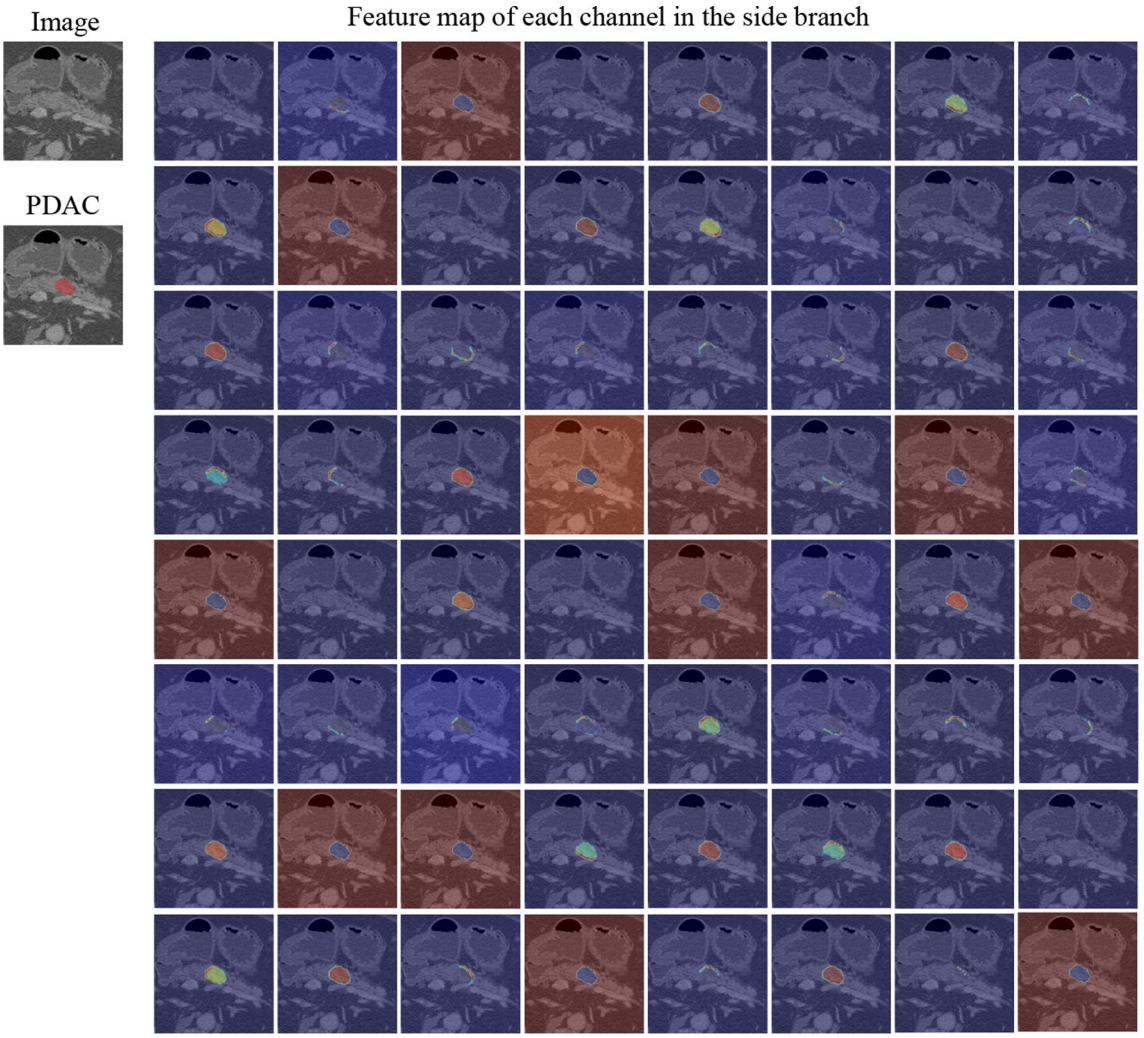


Fig. 7 Visualization of the feature map of each channel learned by the convolution layer in the side branch of LN metastasis status prediction network

Acknowledgements This work was supported in part by the National Science Foundation for Scientists of China (81871352, 82171915, and 82171930), Clinical Research Plan of SHDC (SHDC2020CR4073), 234 Platform Discipline Consolidation Foundation Project (2019YPT001, 2020YPT001), and The Natural Science Foundation of Shanghai Science and Technology Innovation Action Plan (21ZR1478500, 21Y11910300).

Author Contributions Conceptualization: Le Lu, Ling Zhang, Chengwei Shao, and Yun Bian; Methodology: Zhilin Zheng, Jiawen Yao, and Ling Zhang; Data curation: Zhilin Zheng, and Xu Fang; Formal analysis and investigation: Zhilin Zheng, Xu Fang, Jiawen Yao, Mengmeng Zhu, Le Lu, Yu Shi, Hong Lu, Ling Zhang, and Yun Bian; Writing - original draft preparation: Zhilin Zheng, and Jiawen Yao; Writing - review and editing: Zhilin Zheng, Jiawen Yao, Le Lu, and Ling Zhang; Funding acquisition: Chengwei Shao, and Yun Bian; Supervision: Le Lu, Jianping Lu, Chengwei Shao, and Yun Bian.

Data availability The data are not publicly available due to patient privacy restrictions. Please refer any questions or requests regarding data used in this manuscript to the corresponding author(Y.B.).

Declarations

Conflict of interest The authors declare no Conflict of interest.

Ethics approval This study was reviewed and approved by the Biomedical Research Ethics Committee of our institution (No. CHEC2021164), and was performed in accordance with the ethical standards of the 1964 Declaration of Helsinki.

Consent to participate The requirement for patient informed consent was waived by the Institutional Review Board due to the retrospective nature of the study and because all procedures performed were part of routine care.

References

Asagi, A., Ohta, K., Nasu, J., Tanada, M., Nadano, S., Nishimura, R., Teramoto, N., Yamamoto, K., Inoue, T., & Iguchi, H. (2013). Utility of contrast-enhanced fdg-pet/ct in the clinical management of pancreatic cancer: impact on diagnosis, staging, evaluation of treatment response, and detection of recurrence. *Pancreas, 42*(1), 11–19.

Bian, Y., Guo, S., Jiang, H., Gao, S., Shao, C., Cao, K., Fang, X., Li, J., Wang, L., Hua, W., et al. (2019). Relationship between radiomics and risk of lymph node metastasis in pancreatic ductal adenocarcinoma. *Pancreas, 48*(9), 1195.

Bouget, D., Pedersen, A., Vanel, J., Leira, H.O., & Langø, T. (2021). Mediastinal lymph nodes segmentation using 3d convolutional neural network ensembles and anatomical priors guiding. arXiv:2102.06515

Bouget, D., Jørgensen, A., Kiss, G., Leira, H. O., & Langø, T. (2019). Semantic segmentation and detection of mediastinal lymph nodes and anatomical structures in CT data for lung cancer staging. *International Journal of computer assisted radiology and surgery, 14*(6), 977–986.

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: an international retrospective study. *Clinical Cancer Research, 27*(14), 3948–3959. https://doi.org/10.1158/1078-0432.CCR-20-4935

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., & Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In: *MICCAI*, pp. 424–432.

Dahmarde, H., Parooie, F., & Salarzaei, M. (2020). Is 18f-fdg pet/ct an accurate way to detect lymph node metastasis in colorectal cancer: A systematic review and meta-analysis. *Contrast Media & Molecular Imaging,2020*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In: *IEEE CVPR*, pp. 248–255. Ieee

Dong, D., Fang, M.-J., Tang, L., Shan, X.-H., Gao, J.-B., Giganti, F., Wang, R.-P., Chen, X., Wang, X.-X., Palumbo, D., et al. (2020). Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Annals of Oncology, 31*(7), 912–920.

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., et al. (2009). New response evaluation criteria in solid tumours: Revised recist guideline (version 1.1). *European Journal of cancer, 45*(2), 228–247.

Eppel, S. (2018). Classifying a specific image region using convolutional nets with an roi mask as input. arXiv:1812.00291

Feuerstein, M., Glocker, B., Kitasaka, T., Nakamura, Y., Iwano, S., & Mori, K. (2012). Mediastinal atlas creation from 3-d chest computed tomography images: application to automated detection and station mapping of lymph nodes. *Medical Image Analysis, 16*(1), 63–74.

Gao, J., Han, F., Jin, Y., Wang, X., & Zhang, J. (2020). A radiomics nomogram for the preoperative prediction of lymph node metastasis in pancreatic ductal adenocarcinoma. *Frontiers in Oncology, 10*, 1654.

Gerard, S.E., & Reinhardt, J.M. (2019). Pulmonary lobe segmentation using a sequence of convolutional neural networks for marginal learning. In: *IEEE ISBI,* pp. 1207–1211.

Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: Images are more than pictures, they are data. *Radiology, 278*(2), 563–577.

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: A Cancer Journal for Clinicians, 70*(5), 375–403.

Guo, D., Ye, X., Ge, J., Di, X., Lu, L., Huang, L., Xie, G., Xiao, J., Lu, Z., Peng, L., et al. (2021). Deepstationing: thoracic lymph node station parsing in ct scans using anatomical context encoding and key organ auto-search. *In: MICCAI,* pp. 3–12.

Gupta, S., Hu, X., Kaan, J., Jin, M., Mpoy, M., Chung, K., Singh, G., Saltz, M., Kurc, T., & Saltz, J., et al. (2022). Learning topological interactions for multi-class medical image segmentation. In: European Conference on Computer Vision, pp. 701–718. Springer.

Hara, K., Kataoka, H., & Satoh, Y. (2018). Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: *IEEE CVPR*, pp. 6546–6555.

Harmon, S. A., Sanford, T. H., Brown, G. T., Yang, C., Mehralivand, S., Jacob, J. M., Valera, V. A., Shih, J. H., Agarwal, P. K., Choyke, P. L., et al. (2020). Multiresolution application of artificial intelligence in digital pathology for prediction of positive lymph nodes from primary tumors in bladder cancer. *JCO Clinical Cancer Informatics, 4*, 367–382.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In: *IEEE ICCV,* pp. 2961–2969.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In: *IEEE CVPR,* pp. 770–778.

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: *MICCAI*, pp. 187–194.

Huang, C., Han, H., Yao, Q., Zhu, S., & Zhou, S.K. (2019). 3d u^2-net: a 3d universal u-net for multi-domain medical image segmentation. In: *MICCAI,* pp. 291–299.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In: *IEEE CVPR,* pp. 4700–4708.

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. *Nature Methods, 18*(2), 203–211.

Jin, C., Jiang, Y., Yu, H., Wang, W., Li, B., Chen, C., Yuan, Q., Hu, Y., Xu, Y., Zhou, Z., et al. (2021). Deep learning analysis of the primary tumour and the prediction of lymph node metastases in gastric cancer. *British Journal of Surgery, 108*(5), 542–549.

Ji, G.-W., Zhang, Y.-D., Zhang, H., Zhu, F.-P., Wang, K., Xia, Y.-X., Zhang, Y.-D., Jiang, W.-J., Li, X.-C., & Wang, X.-H. (2019). Biliary tract cancer at CT: A radiomics-based model to predict lymph node metastasis and survival outcomes. *Radiology, 290*(1), 90–98.

Jung, W., Park, K. R., Lee, K.-J., Kim, K., Lee, J., Jeong, S., Kim, Y.-J., Kim, J., Yoon, H.-J., Kang, B.-C., et al. (2017). Value of imaging study in predicting pelvic lymph node metastases of uterine cervical cancer. *Radiation Oncology Journal, 35*(4), 340.

Kanda, M., Fujii, T., Nagai, S., Kodera, Y., Kanzaki, A., Sahin, T. T., Hayashi, M., Yamada, S., Sugimoto, H., Nomoto, S., et al. (2011). Pattern of lymph node metastasis spread in pancreatic cancer. *Pancreas, 40*(6), 951–955.

Kanehara & Co., L. (2017). Classification of Pancreas Carcinoma (Fourth English Edition). *Japan Pancreas Society*.

Kay, W., Carreira, J., &Simonyan, K., al. (2017). The kinetics human action video dataset. arXiv:1705.06950.

Kazemifar, S., Balagopal, A., Nguyen, D., McGuire, S., Hannan, R., Jiang, S., & Owrangi, A. (2018). Segmentation of the prostate and organs at risk in male pelvic CT images using deep learning. *Biomedical Physics & Engineering Express, 4*(5), 055003.

Kim, S. H., Song, B.-I., Kim, B. W., Kim, H. W., Won, K. S., Bae, S. U., Jeong, W. K., & Baek, S. K. (2019). Predictive value of [18f]

fdg pet/ct for lymph node metastasis in rectal cancer. *Scientific Reports, 9*(1), 1–7.

Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., Forster, K., Aerts, H. J., Dekker, A., Fenstermacher, D., et al. (2012). Radiomics: The process and the challenges. *Magnetic Resonance Imaging, 30*(9), 1234–1248.

Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. G., Granton, P., Zegers, C. M., Gillies, R., Boellard, R., Dekker, A., et al. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer, 48*(4), 441–446.

Liu, J., Zhao, J., Hoffman, J., Yao, J., Zhang, W., Turkbey, E.B., Wang, S., Kim, C., & Summers, R.M. (2014). Mediastinal lymph node detection on thoracic ct scans using spatial prior from multi-atlas label fusion. In: *Medical Imaging 2014: Computer-Aided Diagnosis*, vol. 9035, p. 90350.

Liu, J., Hoffman, J., Zhao, J., Yao, J., Lu, L., Kim, L., Turkbey, E. B., & Summers, R. M. (2016). Mediastinal lymph node detection and station mapping on chest ct using spatial priors and random forest. *Medical Physics, 43*(7), 4362–4374.

Li, K., Yao, Q., Xiao, J., Li, M., Yang, J., Hou, W., Du, M., Chen, K., Qu, Y., Li, L., et al. (2020). Contrast-enhanced CT radiomics for predicting lymph node metastasis in pancreatic ductal adenocarcinoma: a pilot study. *Cancer Imaging, 20*(1), 1–10.

Lu, Y., Yu, Q., Gao, Y., Zhou, Y., Liu, G., Dong, Q., Ma, J., Ding, L., Yao, H., Zhang, Z., et al. (2018). Identification of metastatic lymph nodes in mr imaging with faster region-based convolutional neural networks. *Cancer Research, 78*(17), 5135–5143.

Meng, L., Dong, D., Chen, X., Fang, M., Wang, R., Li, J., Liu, Z., & Tian, J. (2020). 2d and 3d CT radiomic features performance comparison in characterization of gastric cancer: a multi-center study. *IEEE Journal of Biomedical and Health Informatics, 25*(3), 755–763.

Min, K., Lee, G.-H., & Lee, S.-W. (2022). Attentional feature pyramid network for small object detection. *Neural Networks, 155*, 439–450.

Monfort, M., Andonian, A., Zhou, B., Ramakrishnan, K., Bargal, S. A., Yan, T., Brown, L., Fan, Q., Gutfreund, D., Vondrick, C., et al. (2019). Moments in time dataset: One million videos for event understanding. *IEEE Trans Pat Analy Mach Intel, 42*(2), 502–508.

Oda, H., Bhatia, K.K., Oda, M., Kitasaka, T., Iwano, S., Homma, H., Takabatake, H., Mori, M., Natori, H., & Schnabel, J.A., et al. (2017). Hessian-assisted supervoxel: structure-oriented voxel clustering and application to mediastinal lymph node detection from ct volumes. In: *Medical Imaging 2017: Computer-Aided Diagnosis*.

Oda, H., Roth, H.R., Bhatia, K.K., Oda, M., Kitasaka, T., Iwano, S., Homma, H., Takabatake, H., Mori, M., & Natori, H., et al. (2018). Dense volumetric detection and segmentation of mediastinal lymph nodes in chest ct images. In: Medical Imaging 2018: Computer-Aided Diagnosis.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., & Heinrich, M., al. (2018). Attention u-net: Learning where to look for the pancreas. arXiv:1804.03999.

Roland, C. L., Yang, A. D., Katz, M. H., Chatterjee, D., Wang, H., Lin, H., Vauthey, J. N., Pisters, P. W., Varadhachary, G. R., Wolff, R. A., et al. (2015). Neoadjuvant therapy is associated with a reduced lymph node ratio in patients with potentially resectable pancreatic cancer. *Annals of Surgical Oncology, 22*(4), 1168–1175.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In:*MICCAI,* pp. 234–241.

Rosenfeld, A., & Pfaltz, J. L. (1968). Distance functions on digital pictures. *Pattern Recognition, 1*(1), 33–61.

Salehi, S.S.M., Erdogmus, D., & Gholipour, A. (2017). Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: *MLMI,* pp. 379–387.

Sang, S., Zhou, Y., Islam, M. T., & Xing, L. (2022). Small-object sensitive segmentation using across feature map attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(5), 6289–6306.

Seo, H., Huang, C., Bassenne, M., Xiao, R., & Xing, L. (2019). Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images. *IEEE Trans Medical Imaging, 39*(5), 1316–1325.

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer statistics 2021. *CA: A Cancer Journal for Clinicians, 71*, 7–33.

Sironi, A., Lepetit, V., & Fua, P. (2014). Multiscale centerline detection by learning a scale-space distance transform. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 2697–2704.

Tempero, M. A., Malafa, M. P., Al-Hawary, M., Behrman, S. W., Benson, A. B., Cardin, D. B., Chiorean, E. G., Chung, V., Czito, B., Del Chiaro, M., et al. (2021). Pancreatic adenocarcinoma, version 2.2021, nccn clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network,19*(4), 439–457.

Tseng, D. S., Pranger, B. K., Leeuwen, M. S., Pennings, J. P., Brosens, L. A., Mohammad, N. H., Meijer, V. E., Santvoort, H. C., Erdmann, J. I., & Molenaar, I. Q. (2021). The role of ct in assessment of extraregional lymph node involvement in pancreatic and periampullary cancer: A diagnostic accuracy study. *Radiology: Imaging Cancer***3**(2).

Tseng, D. S., Santvoort, H. C., Fegrachi, S., Besselink, M. G., Zuithoff, N. P., Rinkes, I. H. B., Leeuwen, M. S., & Molenaar, I. Q. (2014). Diagnostic accuracy of CT in assessing extra-regional lymphadenopathy in pancreatic and peri-ampullary cancer: A systematic review and meta-analysis. *Surgical Oncology, 23*(4), 229–235.

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 Research, 77*(21), 104–107.

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.

Wang, Y., Liu, W., Yu, Y., Liu, J.-J., Xue, H.-D., Qi, Y.-F., Lei, J., Yu, J.-C., & Jin, Z.-y. (2020). Ct radiomics nomogram for the preoperative prediction of lymph node metastasis in gastric cancer. *European Radiology, 30*(2), 976–986.

Wei, Q., Chen, Z., Tang, Y., Chen, W., Zhong, L., Mao, L., Hu, S., Wu, Y., Deng, K., Yang, W., et al. (2023). External validation and comparison of mr-based radiomics models for predicting pathological complete response in locally advanced rectal cancer: a two-centre, multi-vendor study. *European Radiology, 33*(3), 1906–1917.

Xia, Y., Yao, J., Lu, L., Huang, L., Xie, G., Xiao, J., Yuille, A., Cao, K., & Zhang, L. (2021). Effective pancreatic cancer screening on non-contrast ct scans via anatomy-aware transformers. In: *MICCAI,* pp. 259–269.

Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, I., Mak, R. H., & Aerts, H. J. W. L. (2019). Deep learning predicts lung cancer treatment response from serial medical imaging. *Clinical Cancer Research,25*(11), 3266–3275. https://doi.org/10.1158/1078-0432.CCR-18-2495https://clincancerres.aacrjournals.org/content/25/11/3266.full.pdf

Yang, J., Wu, Q., Xu, L., Wang, Z., Su, K., Liu, R., Yen, E. A., Liu, S., Qin, J., Rong, Y., et al. (2020). Integrating tumor and nodal radiomics to predict lymph node metastasis in gastric cancer. *Radiotherapy and Oncology, 150*, 89–96.

Yao, J., Shi, Y., Lu, L., Xiao, J., & Zhang, L. (2020). Deepprognosis: Preoperative prediction of pancreatic cancer survival and surgical

margin via contrast-enhanced ct imaging. In: *MICCAI,* pp. 272–282.

Yao, J., Shi, Y., Cao, K., Lu, L., Lu, J., Song, Q., Jin, G., Xiao, J., Hou, Y., & Zhang, L. (2021). Deepprognosis: Preoperative prediction of pancreatic cancer survival and surgical margin via comprehensive understanding of dynamic contrast-enhanced ct imaging and tumor-vascular contact parsing. *Medical Image Analysis, 73*, 102150.

Zhang, H., Xue, J., & Dana, K. (2017). Deep ten: Texture encoding network. In: *IEEE CVPR,* pp. 708–717

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: *IEEE CVPR,* pp. 13743–13752.

Zheng, X., Yao, Z., Huang, Y., Yu, Y., Liu, Y., Mao, R., Li, F., Xiao, Y., Wang, Y., Wang, Y., et al. (2020). Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nature Communications, 11*(1), 1–9.