Improved Prognostic Prediction of Pancreatic Cancer Using Multi-phase CT by Integrating Neural Distance and Texture-Aware Transformer

Hexin Dong[1,2,3], Jiawen Yao[1,3(✉)], Yuxing Tang[1], Mingze Yuan[1,2,3], Yingda Xia[1], Jian Zhou[4,5], Hong Lu[6], Jingren Zhou[1,3], Bin Dong[2,7], Le Lu[1], Zaiyi Liu[8], Li Zhang[2(✉)], Yu Shi[9(✉)], and Ling Zhang[1]

[1] DAMO Academy, Alibaba Group, Hangzhou, China
yaojiawen.yjw@alibaba-inc.com
[2] Peking University, Beijing, China
zhangli_pku@pku.edu.cn
[3] Hupan Lab, Hangzhou 310023, China
[4] Sun Yat-sen University Cancer Center, Guangzhou, China
[5] South China Hospital, Shenzhen University, Shenzhen, China
[6] Tianjin Medical University Cancer Institute and Hospital, Tianjin, China
[7] Peking University Changsha Institute for Computing and Digital Economy, Changsha, China
[8] Guangdong Provincial People's Hospital, Guangzhou, China
[9] Shengjing Hospital, Shenyang, China
18940259980@163.com

Abstract. Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal cancer in which the tumor-vascular involvement greatly affects the resectability and, thus, overall survival of patients. However, current prognostic prediction methods fail to explicitly and accurately investigate relationships between the tumor and nearby important vessels. This paper proposes a novel learnable neural distance that describes the precise relationship between the tumor and vessels in CT images of different patients, adopting it as a major feature for prognosis prediction. Besides, different from existing models that used CNNs or LSTMs to exploit tumor enhancement patterns on dynamic contrast-enhanced CT imaging, we improved the extraction of dynamic tumor-related texture features in multi-phase contrast-enhanced CT by fusing local and global features using CNN and transformer modules, further enhancing the features extracted across multi-phase CT images. We extensively evaluated and compared the proposed method with existing methods in the multi-center (n = 4) dataset with 1,070 patients with PDAC, and statistical

---

---

analysis confirmed its clinical effectiveness in the external test set consisting of three centers. The developed risk marker was the strongest predictor of overall survival among preoperative factors and it has the potential to be combined with established clinical factors to select patients at higher risk who might benefit from neoadjuvant therapy.

Keywords: Pancreatic ductal adenocarcinoma (PDAC) · Survival prediction · Texture-aware Transformer · Cross-attention · Nerual distance

1   Introduction

Pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest forms of human cancer, with a 5-year survival rate of only 9% [16]. Neoadjuvant chemotherapy can increase the likelihood of achieving a margin-negative resection and avoid unnecessary surgery in patients with aggressive tumor types [23]. Providing accurate and objective preoperative biomarkers is crucial for triaging patients who are most likely to benefit from neoadjuvant chemotherapy. However, current clinical markers such as larger tumor size and high carbohydrate antigen (CA) 19-9 level may not be sufficient to accurately tailor neoadjuvant treatment for patients [19]. Therefore, multi-phase contrast-enhanced CT has a great potential to enable personalized prognostic prediction for PDAC, leveraging its ability to provide a wealth of texture information that can aid in the development of accurate and effective prognostic models [2,10].

Previous studies have utilized image texture analysis with hand-crafted features to predict the survival of patients with PDACs [1], but the representational
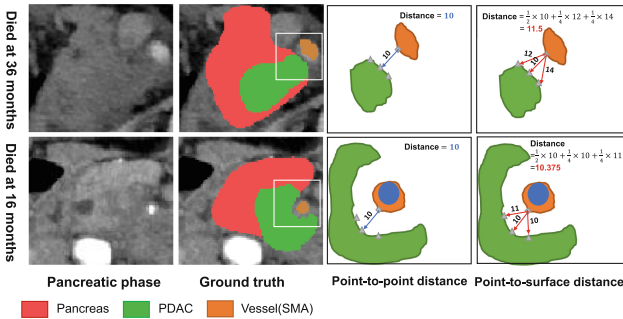

Fig. 1. Two examples of spatial information between vessel (orange region) and tumor (green region). The minimum distance, which refers to the closest distance between the Superior Mesenteric Artery (SMA) and the PDAC tumor region, is almost identical in these two cases. We define the surface-to-surface distance based on point-to-surface distance (weighted-average of red lines from ◇ to △) instead of point-to-point distance (blue lines) to better capture the relationship between the tumor and the perivascular tissue. Here ◇ and △ are points sampled from subset $\hat{\mathcal{V}}_c$ and $\hat{\mathcal{P}}_c$ defined in Eq. 4. The distances and weights shown in the figure is for illustration purposes only. (Color figure online)

power of these features may be limited. In recent years, deep learning-based methods have shown promising results in prognosis models [3,6,12]. However, PDACs differ significantly from the tumors in these studies. A clinical investigation based on contrast-enhanced CT has revealed a dynamic correlation between the internal stromal fractions of PDACs and their surrounding vasculature [14]. Therefore, focusing solely on the texture information of the tumor itself may not be effective for the prognostic prediction of PDAC. It is necessary to incorporate tumor-vascular involvement into the feature extraction process of the prognostic model. Although some studies have investigated tumor-vascular relationships [21,22], these methods may not be sufficiently capable of capturing the complex dynamics between the tumor and its environment.

We propose a novel approach for measuring the relative position relationship between the tumor and the vessel by explicitly using the distance between them. Typically, Chamfer distance [7], Hausdorff distance [8], or other surface-awareness metrics are used. However, as shown in Fig. 1, these point-to-point distances cannot differentiate the degree of tumor-vascular invasion [18]. To address this limitation, we propose a learnable neural distance that considers all relevant points on different surfaces and uses an attention mechanism to compute a combined distance that is more suitable for determining the degree of invasion. Furthermore, to capture the tumor enhancement patterns across multi-phase CT images, we are the first to combine convolutional neural networks (CNN) and transformer [4] modules for extracting the dynamic texture patterns of PDAC and its surroundings. This approach takes advantage of the visual transformer's adeptness in capturing long-distance information compared to the CNN-only-based framework in the original approach. By incorporating texture information between PDAC, pancreas, and peripancreatic vessels, as well as the local tumor information captured by CNN, we aim to improve the accuracy of our prognostic prediction model.

In this study, we make the following contributions: (1) We propose a novel approach for aiding survival prediction in PDAC by introducing a learnable neural distance that explicitly evaluates the degree of vascular invasion between the tumor and its surrounding vessels. (2) We introduce a texture-aware transformer block to enhance the feature extraction approach, combining local and global information for comprehensive texture information. We validate that the cross-attention is utilized to capture cross-modality information and integrate it with in-modality information, resulting in a more accurate and robust prognostic prediction model for PDAC. (3) Through extensive evaluation and statistical analysis, we demonstrate the effectiveness of our proposed method. The signature built from our model remains statistically significant in multivariable analysis after adjusting for established clinical predictors. Our proposed model has the potential to be used in combination with clinical factors for risk stratification and treatment decisions for patients with PDAC.

2   Methods

As shown in Fig. 2, the proposed method consists of two main components. The first component combines the CNN and transformer to enhance the extraction of tumor dynamic texture features. The second component proposes a neural distance metric between PDAC and important vessels to assess their involvements.

2.1   Texture-Aware Vision Transformer: Combination of CNN and Transformer

Recently, self-attention models, specifically vision transformers (ViTs [4]), have emerged as an alternative to CNNs in survival prediction [15,25]. Our proposed texture-aware transformer, inspired by MobileViT [13], aims to combine both local information (such as PDAC texture) and global information (such as the relationship between PDAC and the pancreas). This approach is different from previous methods that rely solely on either CNN-based or transformer-based backbones, focusing only on local or global information, respectively.

The texture-aware transformer (Fig. 2) comprises three blocks, each consisting of a texture-aware CNN block and a texture-aware self-attention block. These blocks encode the input feature of an image $\mathbf{F}_i \in \mathbb{R}^{H \times W \times D \times C}$ to the hidden feature $\mathbf{F}_c \in \mathbb{R}^{H \times W \times D \times C_l}$ using a $3 \times 3 \times 3$ convolutional layer, followed by a
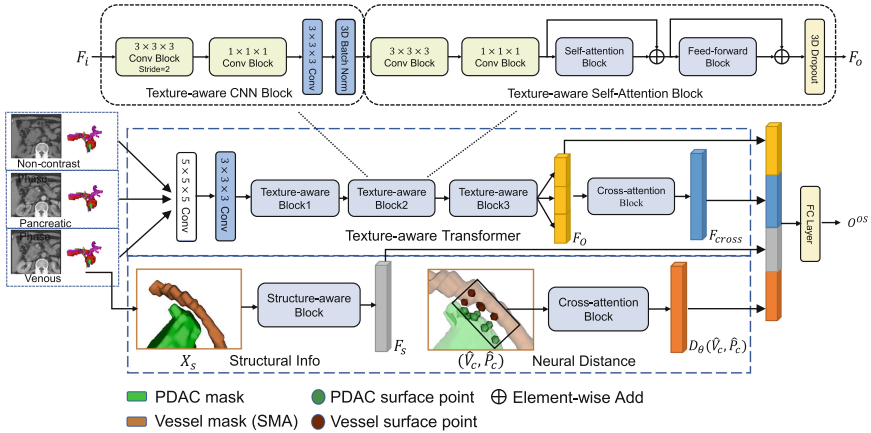

Fig. 2. An overview of the proposed method. The texture-aware transformer captures texture information among PDAC, Pancreas and vessels around Pancreas with our proposed texture-aware transformer block and a cross-attention block to fusion cross-modality features. The structure-aware block extracts the structure relationship between PDAC and four related vessels. The neural distance calculates the distances between the PDAC surface and the vessel surface with our proposed neural distance. We first select related points set from the closest sub-surface on PDAC and vessels respectively. Then we use a cross-attention block to obtain the neural distance. Finally, we concatenate features from three branches to obtain the survival outcome O_{OS}.

$1 \times 1 \times 1$ convolutional layer. The $3 \times 3 \times 3$ convolution captures local spatial information, while the $1 \times 1 \times 1$ convolution maps the input tensor to a higher-dimensional space (*i.e.*, $C_l > C$). The texture-aware CNN block downsamples the input, and the texture-aware self-attention block captures long-range non-local dependencies through a patch-wise self-attention mechanism.

In the texture-aware self-attention block, the input feature $\mathbf{F}_c$ is divided into N non-overlapping 3D patches $\mathbf{F}_u \in \mathbb{R}^{V \times N \times C_u}$, where $V = hwd$ and $N = HWD/V$ is the number of patches, and h, w, d are the height, width, and depth of a patch, respectively. For each voxel position within a patch, we apply a multi-head self-attention block and a feed-forward block following [20] to obtain the output feature $\mathbf{F}_o$. In this study, preoperative multi-phase CE-CT pancreatic imaging includes the non-contrast phase, the pancreatic phase and venous phase. Therefore, we obtain three outputs from the transformer block with the input of these phases, denoted as $\mathbf{F}_o^1, \mathbf{F}_o^2, \mathbf{F}_o^3 \in \mathbb{R}^{D \times C}$, resulting in the concatenated output $\mathbf{F}_o \in \mathbb{R}^{D \times 3C}$.

Instead of directly fusing the outputs as in previous work, we employ a 3-way cross-attention block to extract cross-modality information from these phases. The cross-attention is performed on the concatenated self-attention matrix with an extra mask $\mathbf{M} \in \{0, -\infty\}^{3C \times 3C}$, defined as:

$$\mathbf{F}_{cross} = \text{Softmax}(\mathbf{Q}\mathbf{K}^{\mathrm{T}} + \mathbf{M})\mathbf{V},$$

$$\mathbf{M}(i,j) = \begin{cases} -\infty & kC < i, j \leq (k+1)C, \quad k = 0, 1, 2, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

Here, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are the query, key, and value matrices, respectively, obtained by linearly projecting the input $\mathbf{F}_o^{\mathrm{T}} \in \mathbb{R}^{3C \times D}$. The cross-modality output $\mathbf{F}_{cross}$ and in-modality output $\mathbf{F}_o^{\mathrm{T}}$ are then concatenated and passed through an average pooling layer to obtain the final output feature of the texture branch, denoted as $\mathbf{F}_t \in \mathbb{R}^{C_t}$.

2.2   Neural Distance: Positional and Structural Information Between PDAC and Vessels

The vascular involvement in patients with PDAC affects the resectability and treatment planning [5]. In this study, we investigate four important vessels: portal vein and splenic vein (PVSV), superior mesenteric artery (SMA), superior mesenteric vein (SMV), and truncus coeliacus (TC). We used a semi-supervised nnUnet model to segment PDAC and the surrounding vessels, following recent work [11,21]. We define a general distance between the surface boundaries of PDAC ($\mathcal{P}$) and the aforementioned four types of vessels ($\mathcal{V}$) as $D(\mathcal{V}, \mathcal{P})$, which can be derived as follows:

$$D(\mathcal{V}, \mathcal{P}) = d_{ss}(\mathcal{V}, \mathcal{P}) + d_{ss}(\mathcal{P}, \mathcal{V}) = \frac{1}{\|\mathcal{V}\|} \int_{\mathcal{V}} d_{ps}(v, \mathcal{P}) \mathrm{d}v + \frac{1}{\|\mathcal{P}\|} \int_{\mathcal{P}} d_{ps}(p, \mathcal{V}) \mathrm{d}p, \tag{2}$$

where $v \in \mathcal{V}$ and $p \in \mathcal{P}$ are points on the surfaces of blood vessels and PDAC, respectively. The point-to-surface distance $d_{ps}(v, \mathcal{P})$ is the distance from a point v on $\mathcal{V}$ to $\mathcal{P}$, defined as $d_{ps}(v, \mathcal{P}) = \min_{p \in \mathcal{P}} \|v - p\|_2^2$, and vice versa.

To numerically calculate the integrals in the previous equation, we uniformly sample from the surfaces $\mathcal{V}$ and $\mathcal{P}$ to obtain the sets $\hat{\mathcal{V}}$ and $\hat{\mathcal{P}}$ consisting of N_v points and N_p points, respectively. The distance is then calculated between the two sets using the following equation:

$$D(\hat{\mathcal{V}}, \hat{\mathcal{P}}) = \frac{1}{N_v} \sum_{v \in \hat{\mathcal{V}}} d_{ps}(v, \hat{\mathcal{P}}) + \frac{1}{N_p} \sum_{p \in \hat{\mathcal{P}}} d_{ps}(p, \hat{\mathcal{V}}). \tag{3}$$

However, the above distance treats all points equally and may not be flexible enough to adapt to individualized prognostic predictions. Therefore, we improve the above equation in two ways. Firstly, we focus on the sub-sets $\hat{\mathcal{V}}_c$ and $\hat{\mathcal{P}}_c$ of $\hat{\mathcal{V}}$ and $\hat{\mathcal{P}}$, respectively, which only contain the K closest points to the opposite surfaces $\hat{\mathcal{P}}$ and $\hat{\mathcal{V}}$, respectively. The sub-sets are defined as:

$$\hat{\mathcal{V}}_c = \text{argmin}_{\{v_1, v_2, \cdots, v_K\} \subset \hat{\mathcal{V}}} \sum_{i=1}^{K} d_{ps}(v_i, \hat{\mathcal{P}}),$$
$$\hat{\mathcal{P}}_c = \text{argmin}_{\{p_1, p_2, \cdots, p_K\} \subset \hat{\mathcal{P}}} \sum_{i=1}^{K} d_{ps}(p_i, \hat{\mathcal{V}}). \tag{4}$$

Secondly, we regard the entire sets $\hat{\mathcal{V}}_c$ and $\hat{\mathcal{P}}_c$ as sequences and calculate the distance using a 2-way cross-attention block (similar to Eq. 1) to build a neural distance based on the 3D spatial coordinates of each point:

$$D_\theta(\hat{\mathcal{V}}, \hat{\mathcal{P}}) = \text{CrossAttention}(\hat{\mathcal{V}}_c, \hat{\mathcal{P}}_c), \quad \hat{\mathcal{V}}_c, \hat{\mathcal{P}}_c \in \mathbb{R}^{K \times 3}. \tag{5}$$

Neural distance allows for the flexible assignment of weights to different points and is able to find positional information that is more suitable for PDAC prognosis prediction. In addition to neural distance, we use the 3D-CNN model introduced in [22] to extract the structural relationship between PDAC and the vessels. Specifically, we concatenate each PDAC-vessel pair $\mathbf{X}_s^v \in \mathbb{R}^{2 \times H \times W \times D}$, where $v \in \{$PVSV, SMV, SMA, TC$\}$ and obtain the structure feature $\mathbf{F}_s \in \mathbb{R}^{C_s}$.

Finally, we concatenate the features extracted from the two components and apply a fully-connected layer to predict the survival outcome, denoted as O^{OS}, which is a value between 0 and 1. To optimize the proposed model, we use the negative log partial likelihood as the survival loss [9].

3   Experiments

Dataset. In this study, we used data from Shengjing Hospital to train our method with 892 patients, and data from three other centers, including Guangdong Provincial People's Hospital, Tianjin Medical University and Sun Yatsen University Cancer Center for independent testing with 178 patients. The

contrast-enhanced CT protocol included non-contrast, pancreatic, and portal venous phases. PDAC masks for 340 patients were manually labeled by a radiologist from Shengjing Hospital with 18 years of experience in pancreatic cancer, while the rest were predicted using self-learning models [11,24] and checked by the same annotator. Other vessel masks were generated using the same semi-supervised segmentation models. **C-index** was used as our primary evaluation metric for survival prediction. We also reported the survival **AUC**, which estimates the cumulative area under the ROC curve for the first 36 months.

Implementation Details: We used nested 5-fold cross-validation and augmented the training data by rotating volumetric tumors in the axial direction and randomly selecting cropped regions with random shifts. We also set the output feature dimensions to $C_t = 64$ for the texture-aware transformer, $C_s = 64$ for the structure extraction and $K = 32$ for the neural distance. The batch size was 16 and the maximum iteration was set to 1000 epochs, and we selected the model with the best performance on the validation set during training for testing. We implemented our experiments using PyTorch 1.11 and trained the models on a single NVIDIA 32G-V100 GPU.

Ablation Study. We first evaluated the performance of our proposed texture-aware transformer (TAT) by comparing it with the ResNet18 CNN backbone and ViT transformer backbone, as shown in Table 1. Our model leverages the strengths of both local and global information in the pancreas and achieved the best result. Next, we compared different methods for multi-phase stages, including LSTM, early fusion (Fusion), and cross-attention (Cross) in our method. Cross-attention is more effective and lightweight than LSTM. Moreover, we separated texture features into in-phase features and cross-phase features, which is more reasonable than early fusion.

Secondly, we evaluated each component in our proposed method, as shown in Fig. 2, and presented the results in Table 1. Combining the texture-aware transformer and regular structure information improved the results from 0.630 to 0.648, as tumor invasion strongly affects the survival of PDAC patients. We also employed a simple 4-variable regression model that used only the Chamfer distance of the tumor and the four vessels for prognostic prediction. The resulting C-index of 0.611 confirmed the correlation of the distance with the survival, which is consistent with clinical findings [18]. Explicitly adding the distance measure further improved the results. Our proposed neural distance metric outperformed traditional surface distance metrics like Chamfer distance, indicating its suitability for distinguishing the severity of PDAC.

Comparisons. To further evaluate the performance of our proposed model, we compared it with recent deep prediction methods [17,21] and report the results in Table 2. We modified baseline deep learning models [12,17] and used their network architectures to take a single pancreatic phase or all three phases as inputs. DeepCT-PDAC [21] is the most recent method that considers both tumor-related and tumor-vascular relationships using 3D CNNs. Our proposed method, which uses the transformer and structure-aware blocks to capture tumor enhancement

Table 1. Ablation tests with different network backbones including ResNet18 (Res), ViT and texture-aware transformer (TAT) and methods for multi-phases including LSTM, early fusion (Fusion) and cross-attention (Cross).

Network Backbone	Structural Info	Distance	Model Size (M)	C-index
Res-LSTM	–	–	65.06	0.618 ± 0.017
Res-Cross	–	–	43.54	0.625 ± 0.016
ViT-Cross	–	–	23.18	0.628 ± 0.018
TAT-Fusion	–	–	**3.64**	0.626 ± 0.022
TAT-Cross	–	–	15.13	0.630 ± 0.019
TAT-Cross	✓	–	15.90	0.648 ± 0.021
–	–	Chamfer distance	–	0.611 ± 0.029
TAT-Cross	✓	Chamfer distance	15.93	0.652 ± 0.019
TAT-Cross	✓	Nerual distance	16.28	$\mathbf{0.656 \pm 0.017}$

Table 2. Results of different methods on nested 5-fold CV and independent set.

|	Nested 5-fold CV (n = 892)		Independent test (n = 178)	
	C-index	AUC	C-index	AUC
3DCNN-P [12]	0.630 ± 0.009	0.668 ± 0.019	0.674	0.740
Early Fusion [17]	0.635 ± 0.011	0.670 ± 0.024	0.696	0.779
DeepCT-PDAC [21]	0.640 ± 0.018	0.680 ± 0.036	0.697	0.773
Ours	$\mathbf{0.656 \pm 0.017}$	$\mathbf{0.695 \pm 0.023}$	**0.710**	**0.792**

patterns and tumor-vascular involvement, demonstrated its effectiveness with better performance in both nested 5-fold cross-validation and the multi-center independent test set.

In Table 3, we used univariate and multivariate Cox proportional-hazards models to evaluate our signature and other clinicopathologic factors in the independent test set. The proposed risk stratification was a significant prognostic factor, along with other factors like pathological TNM stages. After selecting significant variables ($p < 0.05$) in univariate analysis, our proposed staging remained strong in multivariable analysis after adjusting for important prognostic markers like pT and resection margins. Notably, our proposed marker remained the strongest among all pre-operative markers, such as tumor size and CA 19-9.

Neoadjuvant Therapy Selection. To demonstrate the added value of our signature as a tool to select patients for neoadjuvant treatment before surgery, we plotted Kaplan-Meier survival curves in Fig. 3. We further stratify patients by our signature after grouping them by tumor size and CA19-9, two clinically used preoperative criteria for selection, and also age. Our signature could significantly stratify patients in all cases and those in the high-risk group had worse outcomes and might be considered as potential neoadjuvant treatment candidates (e.g. 33 high-risk patients with larger tumor size and high CA19-9).

Table 3. Univariate and Multivariate Cox regression analysis. HR: hazard ratio.

Independent test set (n = 178)	Univariate Analysis		Multivariate Analysis	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Proposed (High vs low risk)	2.42(1.64-3.58)	<0.0001	1.85(1.08-3.17)	0.027
Age (> 60 vs = 60)	1.49(1.01-2.20)	0.043	1.01(0.65-1.58)	0.888
Sex (Male vs Female)	1.28(0.86-1.90)	0.221	-	-
pT (pT3-pT4 vs pT1-pT2)	3.17(2.10-4.77)	<0.0001	2.44(1.54-3.86)	0.00015
pN (Positive ve Negative)	1.47(0.98-2.20)	0.008	1.34(0.85-2.12)	0.210
Resection margin (R1 vs R0)	2.84(1.64-4.93)	<0.0001	1.68(0.92-3.07)	0.091
CA19-9 (> 210 vs ≤ 210 U/mL)	0.94(0.64-1.39)	0.759	-	-
Tumor Size (> 25 vs ≤ 25 mm)	2.36(1.59-3.52)	<0.0001	0.99(0.52-1.85)	0.963
Tumor Location (Head vs Tail)	1.06(0.63-1.79)	0.819	-	-
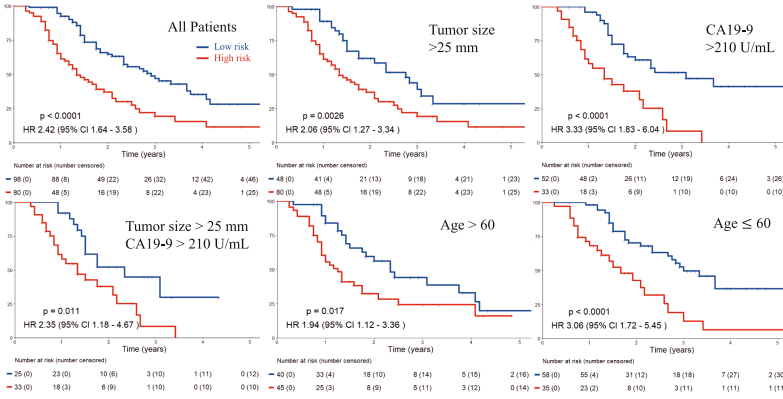
Fig. 3. Kaplan-Meier analyses of overall survival according to the proposed signature in all patients in the independent test set (n = 178) and subgroups defined by preoperative factors. High risk group indicated by the proposed method is the potential patient group that could benefit from neoadjuvant treatment before surgery.

4   Conclusion

In our paper, we propose a multi-branch transformer-based framework for predicting cancer survival. Our framework includes a texture-aware transformer that captures both local and global information about the PDAC and pancreas. We also introduce a neural distance to calculate a more reasonable distance between PDAC and vessels, which is highly correlated with PDAC survival. We have extensively evaluated and statistically analyzed our proposed method, demonstrating its effectiveness. Furthermore, our model can be combined with established high-risk features to aid in the patient selections who might benefit from neoadjuvant therapy before surgery.

References

1. Attiyeh, M.A., et al.: Survival prediction in pancreatic ductal adenocarcinoma by quantitative computed tomography image analysis. Ann. Surg. Oncol. **25**(4), 1034–1042 (2018)
2. Bian, Y., et al.: Artificial intelligence to predict lymph node metastasis at CT in pancreatic ductal adenocarcinoma. Radiology **306**(1), 160–169 (2023)
3. Cheng, N.M., et al.: Deep learning for fully automated prediction of overall survival in patients with oropharyngeal cancer using FDG-PET imaging. Clin. Cancer Res. **27**(14), 3948–3959 (2021)
4. Dosovitskiy, A., et al.: An image is worth 16x16 words: transformers for image recognition at scale. In: ICLR (2021)
5. Ducreux, M., et al.: Cancer of the pancreas: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann. Oncol. **26**, v56–v68 (2015)
6. Feng, Y., Wang, J., An, D., Gu, X., Xu, X., Zhang, M.: End-to-end evidential-efficient net for radiomics analysis of brain MRI to predict oncogene expression and overall survival. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13433, pp. 282–291. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_27
7. Haoqiang Fan, H.S., Guibas, L.: A point set generation network for 3D object reconstruction from a single image. In: CVPR (2017)
8. Huttenlocher, D.P., Rucklidge, W.J., Klanderman, G.A.: Comparing images using the hausdorff distance under translation. In: Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2002)
9. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC Med. Res. Methodol. **18**(1), 24 (2018)
10. Koay, E.J., et al.: Computed tomography-based biomarker outcomes in a prospective trial of preoperative folfirinox and chemoradiation for borderline resectable pancreatic cancer. JCO Precis. Oncol. **3**, 1–15 (2019)
11. Koehler, G., Isensee, F., Maier-Hein, K.: A noisy nnU-Net student for semi-supervised abdominal organ segmentation. In: Ma, J., Wang, B. (eds.) MICCAI 2022. LNCS, vol. 13816, pp. 128–138. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-23911-3_12
12. Lou, B., et al.: An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. Lancet Digit. Health **1**(3), e136–e147 (2019)
13. Mehta, S., Rastegari, M.: Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. In: ICLR (2022)
14. Prokesch, R.W., Chow, L.C., Beaulieu, C.F., Bammer, R., Jeffrey, R.B., Jr.: Isoattenuating pancreatic adenocarcinoma at multi-detector row CT: secondary signs. Radiology **224**(3), 764–768 (2002)

15. Saeed, N., Sobirov, I., Al Majzoub, R., Yaqub, M.: TMSS: an end-to-end transformer-based multimodal network for segmentation and survival prediction. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13437, pp. 319–329. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_31
16. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2019. CA Cancer J. Clin. **69**(1), 7–34 (2019)
17. Tang, Z., et al.: Deep learning of imaging phenotype and genotype for predicting overall survival time of glioblastoma patients. IEEE Trans. Med. Imaging **39**(6), 2100–2109 (2020)
18. Tempero, M.A., et al.: Pancreatic adenocarcinoma, version 2.2021, NCCN clinical practice guidelines in oncology. J. Natl. Compr. Cancer Netw. **19**(4), 439–457 (2021)
19. Tsai, S., et al.: Importance of normalization of ca19-9 levels following neoadjuvant therapy in patients with localized pancreatic cancer. Ann. Surg. **271**(4), 740–747 (2020)
20. Vaswani, A., et al.: Attention is all you need. In: Guyon, I., et al. (eds.) NeurIPS, vol. 30. Curran Associates, Inc. (2017)
21. Yao, J., et al.: Deep learning for fully automated prediction of overall survival in patients undergoing resection for pancreatic cancer: a retrospective multicenter study. Ann. Surg. **278**(1), e68–e79 (2023)
22. Yao, J., et al.: Deepprognosis: Preoperative prediction of pancreatic cancer survival and surgical margin via comprehensive understanding of dynamic contrast-enhanced CT imaging and tumor-vascular contact parsing. Med. Image Anal. **73**, 102150 (2021)
23. Yuan, M., et al.: Devil is in the queries: advancing mask transformers for real-world medical image segmentation and out-of-distribution localization. In: CVPR, pp. 23879–23889 (2023)
24. Zhang, L., et al.: Robust pancreatic ductal adenocarcinoma segmentation with multi-institutional multi-phase partially-annotated CT scans. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 491–500. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_48
25. Zheng, H., et al.: Multi-transSP: multimodal transformer for survival prediction of nasopharyngeal carcinoma patients. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022, pp. 234–243. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_23