Appendix 1 The method of data augmentation

Due to the small number of training samples, we amplify the data through a data augmentation method named "multi-view". The amplification process is shown in Figure S1.

In the study, the doctors manually segmented the tumor in the CT image by marking the cuboid. The center of each cuboid will be recorded as the center of rotation. All CT images were rotated through this center of rotation, and the region was cropped by the largest diameter of the tumor to generate new samples. We have selected 9 perspectives in this rotation, and these rotation methods were recorded as follows:

(I) Without rotation.
(II) Rotate 90 degrees around the X axis and 315 degrees around the Z axis.
(III) Rotate 90 degrees around the X axis and 45 degrees around the Z axis.
(IV) Rotate 45 degrees around the X axis.
(V) Rotate 90 degrees around the X axis.
(VI) Rotate 315 degrees around the X axis.
(VII) Rotate 45 degrees around the Y axis.
(VIII) Rotate 90 degrees around the Y axis.
(IX) Rotate 315 degrees around the Y axis.

Appendix 2 Survival network design and training parameters

The main ideas for the main innovation of survival network in this study come from DeepHit and Circle loss (39,44). We guess that there are two main reasons that the existing survival network cannot obtain satisfactory results. First, the number of patients is too small, which brings great obstacles to the convergence and generalization of the network. Second, the patient may have noise in the time dimension. No matter which point brings greater difficulty to the experiment.

The survival problem can be deemed as a two-step process: 1) The risk ranking of patients with endpoints. 2) The ranking of patients without endpoints based on the above ranking. In this study, we used the tertiary points of the follow-up time of patients with endpoints as the cut-off point to ensure that the number of patients in each category is almost similar. This approach makes the network easier to train. The target question of the network is whether the patient has an endpoint in this time interval. In this study, we divided patients into 3 categories, which is equivalent to two cut-off points for OS and PFS. For survival-related research using this method, the number of categories can be determined according to the target topic and sample size. The cut-off points of OS were 135 days and 282 days, which means that the patients were divided into three groups according to the time interval of the endpoint event from 0 to 135 days, 135 days to 282 days, 282 days later. The cut-off points of PFS were 98 days and 213 days.

We named the three output scores as risk vectors, which integrate patient prognosis information through the network. The method to encode patients is similar to DeepHit (39). If the patient has an end point in the time interval, it is marked as 1, and if the end point is not present, it is marked as 0. It is worth noting that patients without an endpoint are marked as 1 in the interval where an endpoint is likely to occur. At this time, patients can be divided into two types, one is the multi-label patients, in other words the patients have multiple time intervals for the occurrence of endpoints. The other is the single-label patients. We set different loss functions for these two kinds of patients.

Drawing lessons from the circle loss (44), we first display the cross-entropy loss function in the following form:

$$loss = \log\left(1 + \sum_{i \in N, j \in O} e^{s_i - s_j}\right)$$

Next, we can regard the survival problem as a multi-label classification problem. We introduced the zeroth category to stratify the score, and scale factor and relaxation factor in pairwise learning. At this time, for the single-label patients, the loss function of the sample is calculated as:

$$loss = \log\left(1 + \sum_{i \in N, j \in O} e^{\gamma(s_i - s_j + m)} + \sum_{i \in N} e^{\gamma s_i - s_0} + \sum_{j \in O} e^{\gamma(s_j + m) - s_0}\right)$$

https://dx.doi.org/10.21037/tlcr-22-244

We interpret the above formula as the score of the patient's target category needs to be greater than the score of the zeroth category. The score of the non-target category needs to be less than the score of the zeroth category. There are 3 hyperparameters in the formula, namely γ, m and s0. γ is the scale factor, m is the relaxation factor, and s0 is the score of the zeroth category. Since there is no approach to constrain the target score for patients without endpoints, we only constrain the scores of non-target categories. At this time, the loss function of the patients without endpoint is:

$$loss = \log\left(1 + \sum_{i \in N, j \in O} e^{\gamma(s_i - s_j + m)} + \sum_{i \in N} e^{\gamma s_i - s_0}\right)$$

Through the formula, we can observe that in the training process, patients with endpoints tend to have larger loss values, so the network first sorts patients with endpoints. The loss of patients with a shorter follow-up time is automatically compressed to 0. During training, we also choose AdamP as our optimizer and set 1500 epochs (45). For the learning rate, we set the learning rate to 1e-4 and also added epoch-based attenuation. The loss of the validation cohort does not change more than 0.01 in 3 consecutive epochs, which is regarded as stable by us.

Appendix 3 Parameters and training procedure of dual-task network

We train the network including three processes. First, we trained the convolution module and directly connect to the output layer, then we freezed the parameters of the convolution module and trained the fully connected network. Finally, we set a small learning rate for end-to-end training.

We set that the loss of the validation cohort does not change more than 0.01 for three consecutive epochs as the sigh of model stability. In training, we set a total of 1500 epochs, and each epoch attenuates the learning rate (the attenuation coefficient is 1e-3). We chose the AdamP that the radial component (i.e. parallel to the weight) vector is deleted in each iteration as the optimizer (45). For the loss function of a single task, we selected the cross-entropy loss function. Because it is dual-task training, and the category ratio of PR is larger than that of PD. So, we set the calculation method of loss as follows:

$$loss = 0.55 \times loss_{PR} + 0.45 \times loss_{PD}$$

Appendix 4 Packages

For Python packages, the Mann-Whitney U test and chi-square test we calculated by python package named “stats”. The AUC and macro-accuracy are calculated by the python package named “scikit-learn” (version 0.21.3; https://scikit-learn.org/stable/). For drawing ROC, line graph, bar graph and 3D view, we used python package named “matplotlib” and “seaborn”. For the development deep learning networks and coding survival loss function, we employed the deep learning tensor library named “pytorch” (version 1.4; https://pytorch.org/). For R packages, we drew Kaplan-Meier curves by the R packages named ‘survival’, ‘survminer’ and ‘ggplot2’. For the multivariate analysis of clinical characteristics and MOSRS, we used R package named ‘survival’ for calculation.

References

44. Sun Y, Cheng C, Zhang Y, et al. Circle loss: A unified perspective of pair similarity optimization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020:6398-407.
45. Heo B, Chun S, Oh SJ, et al. AdamP: Slowing down the slowdown for momentum optimizers on scale-invariant weights. In Proceedings of the International Conference on Learning Representations (ICLR); 2021:3-7.

https://dx.doi.org/10.21037/tlcr-22-244

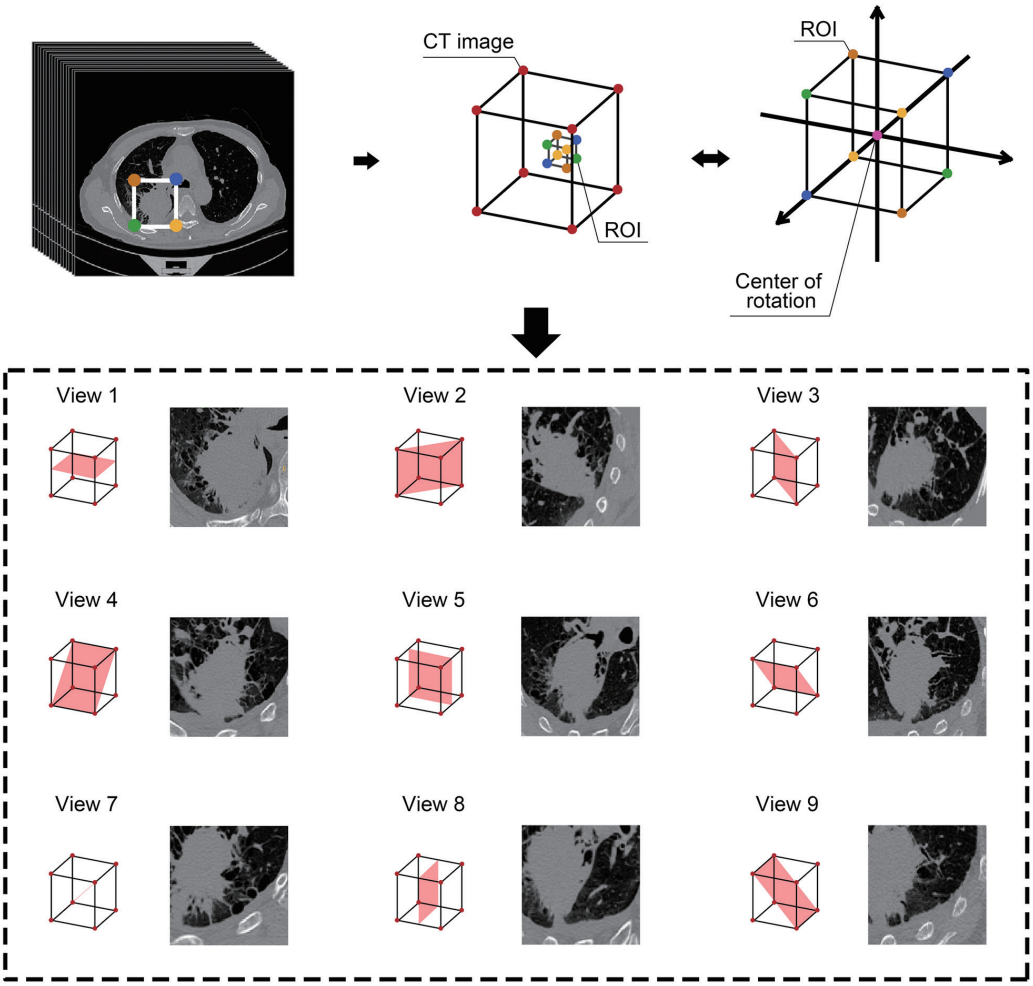

Figure S1 The flow chart of data augmentation. CT, computed tomography; ROI, region of interest.

https://dx.doi.org/10.21037/tlcr-22-244

Table S1 Clinical characteristics of the high- and low-risk groups based on OSRS

Characteristics	All (N=236)	Low-risk group (N=171)	High-risk group (N=65)	P value
Gender				0.72
Female	40 (0.17)	29 (0.17)	11 (0.17)	
Male	196 (0.83)	142 (0.83)	54 (0.83)	
Optimal immune response				1.00
Progressive disease	62 (0.26)	39 (0.23)	23 (0.35)	
Stable disease	119 (0.50)	91 (0.53)	28 (0.43)	
Partial response	55 (0.23)	41 (0.24)	14 (0.22)	
Median age (range)	64 (57, 70)	64 (56, 70)	65 (61, 70)	0.06
Smoking status				0.86
Never smoked	45 (0.19)	34 (0.20)	11 (0.17)	
Current or former smoker	83 (0.35)	57 (0.33)	26 (0.40)	
Unknown	108 (0.46)	80 (0.47)	28 (0.43)	
ECOG performance-status score				0.04
0	11 (0.05)	11 (0.06)	0 (0.00)	
1	112 (0.47)	79 (0.46)	33 (0.51)	
2	6 (0.03)	2 (0.01)	4 (0.06)	
Unknown	107 (0.45)	79 (0.46)	28 (0.43)	
Clinical stage				0.90
III	13 (0.06)	8 (0.05)	5 (0.08)	
IV	116 (0.49)	84 (0.49)	32 (0.49)	
Unknown	107 (0.45)	79 (0.46)	28 (0.43)	
Tumor histologic type				0.55
Squamous cell carcinoma	71 (0.30)	50 (0.29)	21 (0.32)	
Adenocarcinoma	119 (0.50)	94 (0.55)	25 (0.38)	
Others	46 (0.19)	27 (0.16)	19 (0.29)	
Tumor mutation				0.76
No mutation	104 (0.44)	74 (0.43)	30 (0.46)	
Mutation	25 (0.11)	18 (0.11)	7 (0.11)	
Unknown	107 (0.45)	79 (0.46)	28 (0.43)	
OSRS	-0.00 ± 0.98	-0.43 ± 0.78	1.10 ± 0.47	0.00
PFSRS	0.01 ± 0.99	-0.20 ± 0.98	0.55 ± 0.8	0.00
MOSRS	0.00 ± 1.27	-0.46 ± 1.10	1.21 ± 0.8	0.00

Categorical data are shown as numbers (%) and continuous data as mean ± standard deviation (SD) or median (range). ECOG, Eastern Cooperative Oncology Group; OSRS, overall survival risk score; PFSRS, progression-free survival risk score; MOSRS, multi-overall survival risk score.

https://dx.doi.org/10.21037/tlcr-22-244

Table S2 Clinical characteristics of the high- and low-risk groups based on PFSRS

Characteristics	All (N=236)	Low-risk group (N=173)	High-risk group (N=63)	P value
Gender				0.72
Female	40 (0.17)	21 (0.12)	19 (0.30)	
Male	196 (0.83)	152 (0.88)	44 (0.70)	
Optimal immune response				1.00
Progressive disease	62 (0.26)	31 (0.18)	31 (0.49)	
Stable disease	119 (0.50)	95 (0.55)	24 (0.38)	
Partial response	55 (0.23)	47 (0.27)	8 (0.13)	
Median age (range)	64 (57,70)	64 (57,70)	65 (58,68)	0.46
Smoking status				0.86
Never smoked	45 (0.19)	27 (0.16)	18 (0.29)	
Current or former smoker	83 (0.35)	62 (0.36)	21 (0.33)	
Unknown	108 (0.46)	84 (0.49)	24 (0.38)	
ECOG performance-status score				0.04
0	11 (0.05)	10 (0.06)	1 (0.02)	
1	112 (0.47)	75 (0.43)	37 (0.59)	
2	6 (0.03)	5 (0.03)	1 (0.02)	
Unknown	107 (0.45)	83 (0.48)	24 (0.38)	
Clinical stage				0.90
III	13 (0.06)	7 (0.04)	6 (0.10)	
IV	116 (0.49)	83 (0.48)	33 (0.52)	
Unknown	107 (0.45)	83 (0.48)	24 (0.38)	
Tumor histologic type				0.55
Squamous cell carcinoma	71 (0.30)	50 (0.29)	21 (0.33)	
Adenocarcinoma	119 (0.50)	87 (0.50)	32 (0.51)	
Others	46 (0.19)	36 (0.21)	10 (0.16)	
Tumor mutation				0.76
No mutation	104 (0.44)	71 (0.41)	33 (0.52)	
Mutation	25 (0.11)	19 (0.11)	6 (0.10)	
Unknown	107 (0.45)	83 (0.48)	24 (0.38)	
OSRS	-0.01 ± 1.08	-0.22 ± 1.06	0.56 ± 0.92	0.00
PFSRS	0.01 ± 0.99	-0.41 ± 0.78	1.15 ± 0.50	0.00
MOSRS	-0.00 ± 1.38	-0.47 ± 1.21	1.29 ± 0.92	0.00

Categorical data are shown as numbers (%) and continuous data as mean ± standard deviation (SD) or median (range). ECOG, Eastern Cooperative Oncology Group; OSRS, overall survival risk score; PFSRS, progression-free survival risk score; MOSRS, multi-overall survival risk score.

https://dx.doi.org/10.21037/tlcr-22-244

Table S3 Clinical characteristics of the high- and low-risk groups based on MOSRS

Characteristics	All (N=236)	Low-risk group (N=173)	High-risk group (N=63)	P value
Gender				0.72
Female	40 (0.17)	27 (0.16)	13 (0.21)	
Male	196 (0.83)	146 (0.84)	50 (0.79)	
Optimal immune response				1.00
Progressive disease	62 (0.26)	33 (0.19)	29 (0.46)	
Stable disease	119 (0.50)	94 (0.54)	25 (0.40)	
Partial response	55 (0.23)	46 (0.27)	9 (0.14)	
Median age (range)	64 (57,70)	64 (57,70)	66 (61,70)	0.06
Smoking status				0.86
Never smoked	45 (0.19)	30 (0.17)	15 (0.24)	
Current or former smoker	83 (0.35)	59 (0.34)	24 (0.38)	
Unknown	108 (0.46)	84 (0.49)	24 (0.38)	
ECOG performance-status score				0.04
0	11 (0.05)	11 (0.06)	0 (0.00)	
1	112 (0.47)	76 (0.44)	36 (0.57)	
2	6 (0.03)	3 (0.02)	3 (0.05)	
Unknown	107 (0.45)	83 (0.48)	24 (0.38)	
Clinical stage				0.90
III	13 (0.06)	5 (0.03)	8 (0.13)	
IV	116 (0.49)	85 (0.49)	31 (0.49)	
Unknown	107 (0.45)	83 (0.48)	24 (0.38)	
Tumor histologic type				0.55
Squamous cell carcinoma	71 (0.30)	49 (0.28)	22 (0.35)	
Adenocarcinoma	119 (0.50)	92 (0.53)	27 (0.43)	
Others	46 (0.19)	32 (0.18)	14 (0.22)	
Tumor mutation				0.76
No mutation	104 (0.44)	70 (0.40)	34 (0.54)	
Mutation	25 (0.11)	20 (0.12)	5 (0.08)	
Unknown	107 (0.45)	83 (0.48)	24 (0.38)	
OSRS	-0.0 ± 0.98	-0.35 ± 0.86	0.94 ± 0.62	0.00
PFSRS	0.01 ± 0.99	-0.34 ± 0.86	0.95 ± 0.65	0.00
MOSRS	0.00 ± 1.27	-0.51 ± 1.02	1.40 ± 0.7	0.00

Categorical data are shown as numbers (%) and continuous data as mean ± standard deviation (SD) or median (range). ECOG, Eastern Cooperative Oncology Group; OSRS, overall survival risk score; PFSRS, progression-free survival risk score; MOSRS, multi-overall survival risk score.

https://dx.doi.org/10.21037/tlcr-22-244

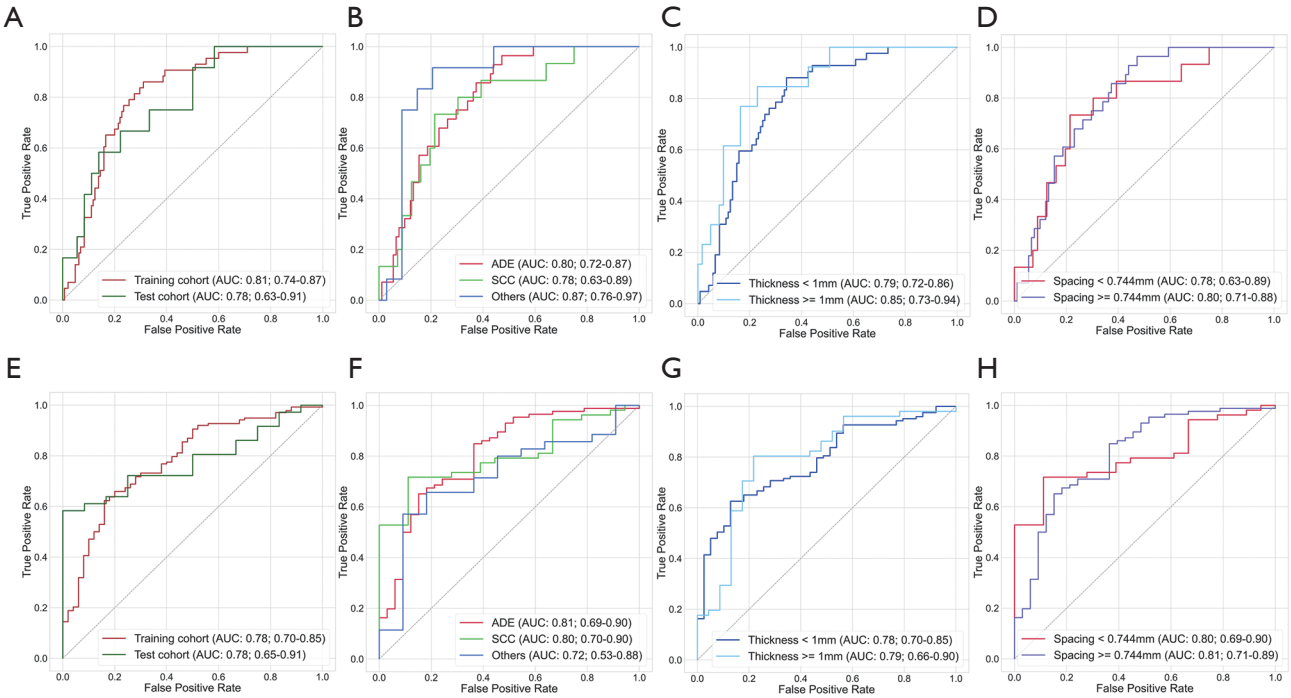

Figure S2 Ability of PRS and PDS to predict optimal immune response. The first and second rows in the figure are the evaluation results of PRS and PDS, respectively. For each column, from left to right, the ROC of PRS and PDS in different cohorts (A,E), different tumor histologic type (B,F), different thickness (C,G), and different voxel spacing (D,H). ROC, receiver operating characteristic curve; PRS, partial response score; PDS, progressive disease score; ADE, adenocarcinoma; SCC, squamous cell carcinoma; AUC, area under the curve.

https://dx.doi.org/10.21037/tlcr-22-244