**Supplementary Material**

1. **Performance of training and testing cohort for choosing deep learning features extractor**

For selecting the best model of extracting deep learning features, we used three types of deep learning network architectures, including CNN, MedicalNet, and ViT. We elaborated the details in the main text, and show the predictive performance of the classification task on preoperative MRI training cohort.

Table SI showed the predictive performance of postoperative MRI training cohort, and Table SII and Table SIII demonstrated the predictive performance of preoperative and postoperative MRI testing cohort, respectively.

**Table. SI. Predictive performance of different types of models in identification of long/short OS on patients in the postoperative MRI training cohort with 10-fold cross-validation.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Framework** | **Model** | **ACC(%)** | **SEN(%)** | **SPE(%)** | **PPR(%)** | **NPR(%)** |
| CNN | Vgg16 | 66.31±3.20 | 70.84±3.51 | 62.42±3.23 | 61.90±3.10 | 71.42±3.63 |
| Vgg19 | 70.29±3.53 | 73.69±3.26 | 65.75±3.60 | 64.34±3.18 | 74.72±3.45 |
| ResNet50 | 73.77±2.42 | 78.65±2.44 | 68.43±2.08 | 65.10±2.17 | 80.29±2.67 |
| DenseNet201 | 78.10±3.51 | 82.34±3.66 | 72.15±3.41 | **70.45±3.14** | 83.84±3.32 |
| InceptionV3 | 77.30±3.84 | **82.50±3.89** | 73.04±3.60 | 69.96±3.81 | 82.98±4.00 |
| EfficientNetB7 | **78.82±1.94** | 81.85±2.16 | **74.03±2.09** | 71.95±1.82 | **83.33±2.22** |
| InceptionResNetV2 | 73.66±1.55 | 75.30±2.05 | 70.64±1.88 | 68.65±1.70 | 76.64±1.64 |
| MedicalNet | 3D-ResNet10 | 65.93±3.89 | 68.01±4.15 | 60.34±3.72 | 56.64±3.91 | 69.07±4.28 |
| 3D-ResNet18 | 67.91±3.63 | 73.82±4.07 | 61.87±3.55 | 62.13±3.32 | 75.34±3.33 |
| 3D-ResNet34 | 69.41±2.51 | 74.61±2.24 | 63.45±1.98 | 61.74±2.03 | 75.38±2.40 |
| 3D-ResNet50 | **73.82±2.57** | **77.75±2.82** | 68.04±2.69 | 65.24±2.38 | **78.66±2.80** |
| 3D-ResNet101\* | 72.35±4.15 | 76.19±4.19 | **69.93±3.70** | **67.72±3.71** | 77.66±4.00 |
| ViT | ViT-B/16 | 74.17±1.32 | 77.44±0.80 | 70.04±1.10 | 67.06±0.94 | 78.90±1.12 |
| ViT-B/32 | 73.65±1.65 | 75.32±1.45 | 68.39±1.57 | 67.05±1.68 | 76.74±1.48 |
| ViT-L/16 | 78.44±3.62 | 83.05±3.91 | 72.97±3.44 | 69.83±3.16 | 84.81±4.11 |
| ViT-L/32 | 77.20±3.17 | 82.94±2.94 | 73.39±3.11 | 71.63±2.98 | 83.65±2.88 |
| ViT-B/16-imagenet1k | 79.86±1.83 | 83.18±2.16 | 75.74±1.69 | 73.57±1.76 | 84.54±2.22 |
| ViT-B/32-imagenet1k | 76.08±3.16 | 79.40±2.88 | 70.23±2.59 | 68.32±3.03 | 81.77±3.97 |
| ViT-L/16-imagenet1k | **81.53±1.72** | **84.57±1.78** | 75.40±1.45 | 72.75±1.46 | **85.74±1.80** |
| ViT-L/32-imagenet1k | 80.07±3.04 | 84.39±3.32 | **77.29±2.80** | **74.43±3.31** | 84.05±3.06 |

Abbreviations: ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; PPR, Positive predictive ratio; NPR, Negative predictive ratio;

\*3D-ResNet101 model is pre-trained from 8 medical datasets, while other MedicalNet models is pre-trained from 23 medical datasets.

**Table. SII. Predictive performance of different types of models in identification of long/short OS on patients in the preoperative MRI testing cohort with 10-fold cross-validation.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Framework** | **Model** | **ACC(%)** | **SEN(%)** | **SPE(%)** | **PPR(%)** | **NPR(%)** |
| CNN | Vgg16 | 64.40±3.07 | 68.84±3.15 | 61.02±3.23 | 59.53±3.41 | 69.54±3.20 |
| Vgg19 | 69.27±3.88 | 73.59±4.07 | 63.45±3.73 | 62.88±3.94 | 73.62±4.02 |
| ResNet50 | 71.75±2.87 | 76.30±3.05 | 68.84±2.77 | 65.93±2.82 | 78.04±2.85 |
| DenseNet201 | 75.34±3.50 | 77.54±3.62 | 71.03±3.49 | 70.40±3.41 | 78.83±3.66 |
| InceptionV3 | 74.64±2.55 | 77.48±2.42 | 70.58±2.58 | 67.32±2.23 | 77.94±2.75 |
| EfficientNetB7 | **76.09±2.04** | **79.47±2.21** | **72.54±1.97** | **71.51±1.89** | **81.50±2.10** |
| InceptionResNetV2 | 70.33±1.92 | 75.58±2.29 | 65.32±1.94 | 63.64±2.00 | 78.91±2.83 |
| MedicalNet | 3D-ResNet10 | 62.14±3.85 | 65.35±3.56 | 58.34±4.10 | 56.35±3.93 | 66.39±3.76 |
| 3D-ResNet18 | 67.35±3.37 | 70.93±3.59 | 64.00±3.28 | 62.07±3.50 | 71.73±3.32 |
| 3D-ResNet34 | 70.91±3.21 | 76.74±3.20 | 65.55±3.29 | 64.92±2.89 | 78.01±2.99 |
| 3D-ResNet50 | **74.52±2.82** | **78.33±2.71** | **70.98±3.16** | **68.93±2.80** | **79.05±2.66** |
| 3D-ResNet101\* | 67.28±3.08 | 71.50±3.24 | 61.85±3.20 | 60.57±2.95 | 73.23±2.91 |
| ViT | ViT-B/16 | 72.83±1.51 | 75.43±1.60 | 69.94±1.77 | 67.14±1.38 | 77.82±1.62 |
| ViT-B/32 | 72.22±2.20 | 75.81±1.92 | 68.74±2.52 | 65.71±2.37 | 76.05±1.98 |
| ViT-L/16 | 77.58±2.69 | 80.49±2.70 | 73.77±3.01 | 71.53±2.46 | 80.42±2.79 |
| ViT-L/32 | 76.14±2.94 | 78.89±3.11 | 71.35±2.90 | 70.52±3.13 | 78.61±2.81 |
| ViT-B/16-imagenet1k | 76.12±1.82 | 79.11±1.54 | 73.53±1.67 | 72.08±1.99 | 80.29±1.68 |
| ViT-B/32-imagenet1k | 77.08±2.37 | 79.10±2.25 | 72.22±2.44 | 69.16±2.31 | 80.19±2.50 |
| ViT-L/16-imagenet1k | **79.20±2.35** | **82.44±2.24** | 74.16±2.29 | 72.30±2.17 | **82.66±2.49** |
| ViT-L/32-imagenet1k | 78.17±3.52 | 81.85±3.87 | **74.83±4.00** | **73.96±3.66** | 82.41±3.61 |

**Table. SIII. Predictive performance of different types of models in identification of long/short OS on patients in the postoperative MRI testing cohort with 10-fold cross-validation.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Framework** | **Model** | **ACC(%)** | **SEN(%)** | **SPE(%)** | **PPR(%)** | **NPR(%)** |
| CNN | Vgg16 | 59.82±3.28 | 66.94±3.45 | 54.00±3.31 | 53.18±3.16 | 67.21±3.25 |
| Vgg19 | 61.71±3.64 | 67.82±3.66 | 58.13±3.39 | 57.03±3.57 | 68.64±3.70 |
| ResNet50 | 69.25±2.50 | 73.11±2.73 | 66.94±2.61 | 65.93±2.79 | 74.73±2.48 |
| DenseNet201 | 73.64±3.33 | **78.74±3.27** | 68.75±3.60 | 66.91±3.71 | **80.82±3.25** |
| InceptionV3 | 72.30±3.41 | 75.20±3.39 | 69.22±3.18 | 66.49±3.44 | 75.04±3.20 |
| EfficientNetB7 | **74.37±2.05** | 77.15±1.95 | **71.94±2.14** | **70.73±1.85** | 79.61±2.23 |
| InceptionResNetV2 | 66.55±1.77 | 70.65±1.82 | 62.02±2.03 | 61.64±1.88 | 72.57±1.75 |
| MedicalNet | 3D-ResNet10 | 58.31±4.04 | 61.04±4.26 | 53.55±3.91 | 51.92±4.09 | 62.21±4.17 |
| 3D-ResNet18 | 60.90±3.87 | 64.84±3.73 | 56.38±4.10 | 53.28±3.83 | 64.90±3.99 |
| 3D-ResNet34 | 63.76±3.09 | 67.01±3.35 | 59.25±3.42 | 58.93±3.20 | 68.74±2.96 |
| 3D-ResNet50 | **69.84±2.61** | **71.93±2.64** | **64.41±2.77** | 63.78±2.84 | **72.33±2.57** |
| 3D-ResNet101\* | 67.05±3.92 | 70.74±4.12 | 64.92±3.98 | **63.08±3.79** | 72.46±4.07 |
| ViT | ViT-B/16 | 71.94±1.50 | 74.85±1.41 | 67.58±1.39 | 66.55±1.55 | 76.01±1.64 |
| ViT-B/32 | 69.37±1.62 | 73.66±1.80 | 64.28±1.53 | 62.57±1.66 | 73.87±1.75 |
| ViT-L/16 | 73.99±2.86 | 76.86±3.02 | 69.49±3.19 | 67.26±3.14 | 77.81±3.06 |
| ViT-L/32 | 73.50±2.90 | 77.29±2.97 | 68.21±2.82 | 65.18±3.10 | 78.53±2.91 |
| ViT-B/16-imagenet1k | 73.54±2.21 | 77.17±2.04 | 70.49±2.33 | 68.05±2.29 | 79.27±2.45 |
| ViT-B/32-imagenet1k | 71.75±2.43 | 74.52±2.62 | 65.05±2.31 | 64.92±2.40 | 75.62±2.38 |
| ViT-L/16-imagenet1k | **75.28±2.39** | **78.22±2.30** | **72.27±2.24** | **69.61±2.15** | **80.16±2.48** |
| ViT-L/32-imagenet1k | 73.69±3.17 | 77.65±2.96 | 70.90±3.01 | 67.05±3.11 | 78.70±3.21 |

2**. The effectiveness comparison of global max pooling (GMP) and global average pooling (GAP)**

The effectiveness comparison of GMP and GAP was shown in Figure S1. Among them, EfficientNetB7 was used as the baseline for extracting deep learning features. In comparison, GMP is better than GAP in the localization of our task. That is, GMP is more accurate in localizing the tumor area and locates the boundaries more clearly. Therefore, after comparing a large number of examples, we believed that GMP can precisely represent the deep learning features of tumor regions.

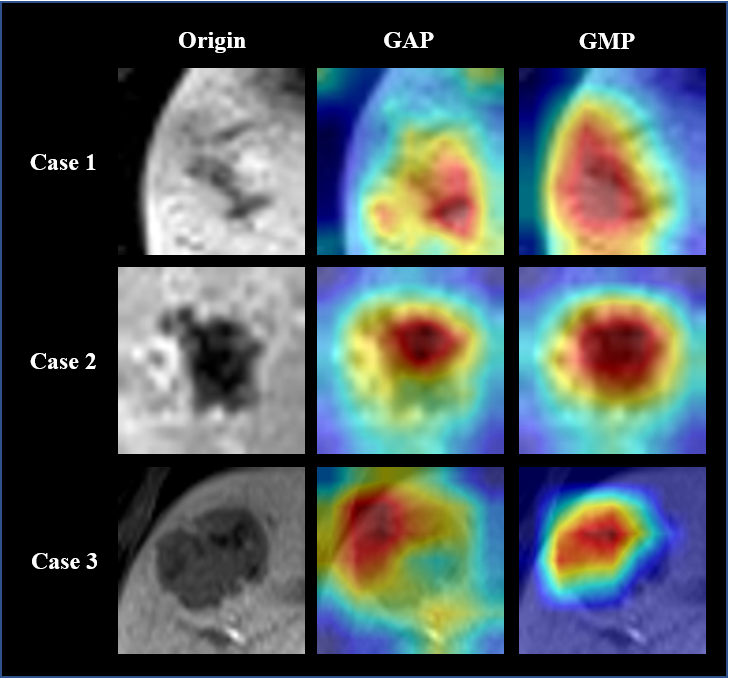


Fig S1. The effectiveness comparison of GMP and GAP

**3. The partial feature correlation matrices of different types of features**

We calculated the correlation matrix for radiomics, deep learning, and immune features, respectively. Of note, radiomics feature matrix exhibit strong correlation.

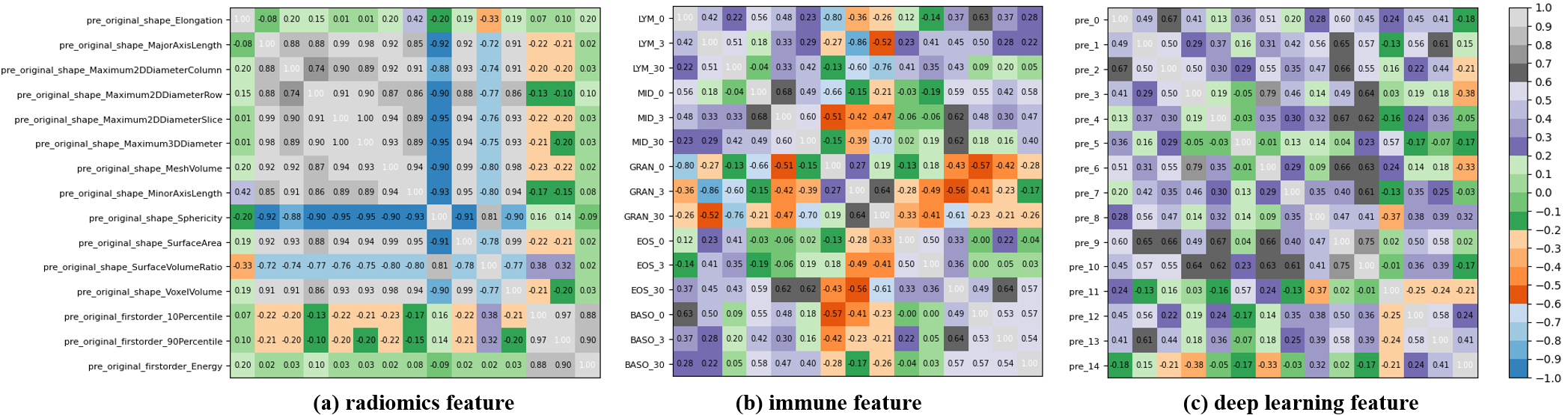


Fig S2. Partial correlation metrices of (a) radiomics, (b) immune, and (c) deep learning features.

**4. Individual survival probability for each multimode thermal therapy patient**

We showed individual survival probability plots for all 17 patients separately. For each patient, we selected the highest risk score from their tumor slices. We used the median risk score to divide the patients into high-risk and low-risk groups. 11 patients were assigned to high risk group and 6 patients were assigned to low risk group. The high-risk and low-risk patients shown in Fig S7 were the longest and shortest follow-ups which must be included in the training set. Fig S2-S6 demonstrated the individual survival probability of randomly selected patients from high and low risk of tumor recurrence. Individual survival probability plots for the remaining high-risk group patients were shown in Figure S8-S12.

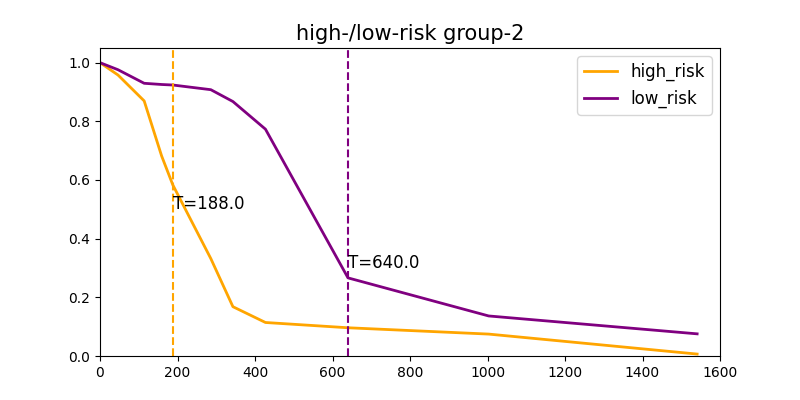
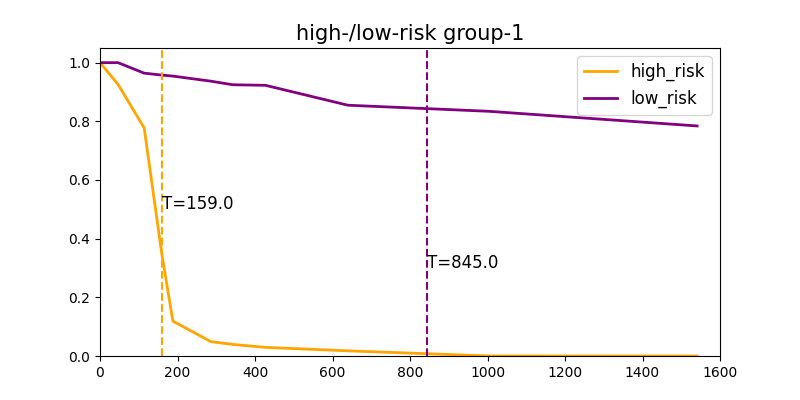


Fig S3. Individual survival probability of high-risk patient-6 and low-risk patient-3 Fig S4. Individual survival probability of high-risk patient-7 and low-risk patient-2

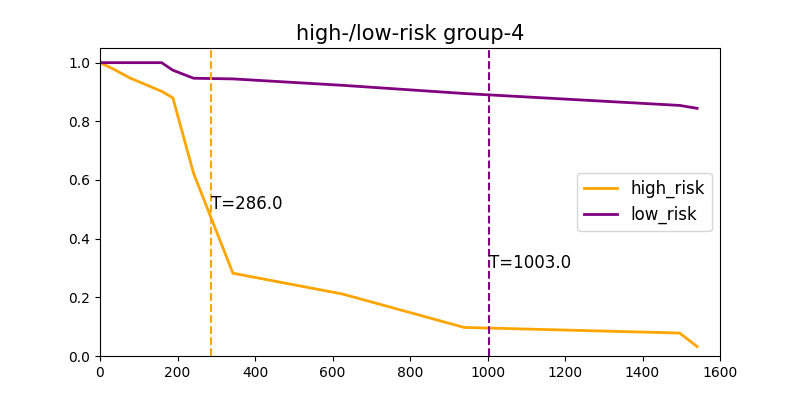
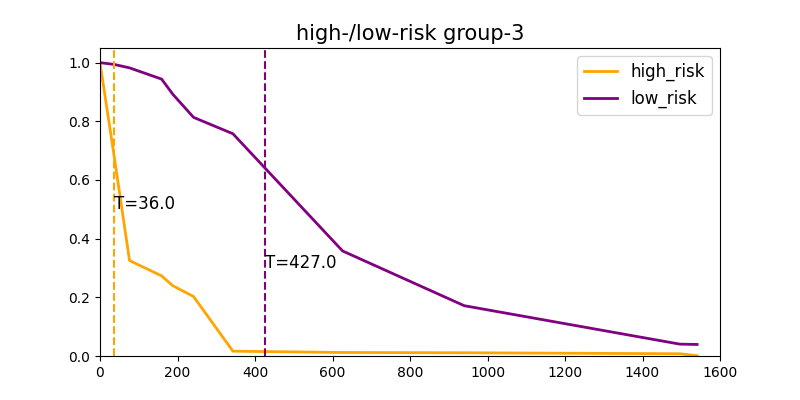


Fig S5. Individual survival probability of high-risk patient-3 and low-risk patient-1 Fig S6. Individual survival probability of high-risk patient-9 and low-risk patient-4

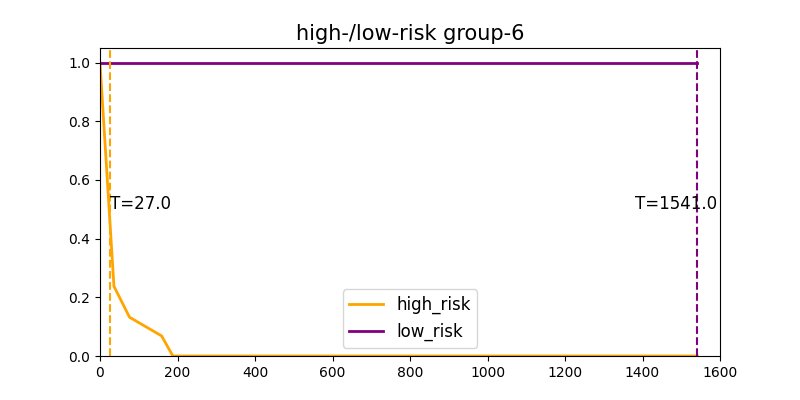
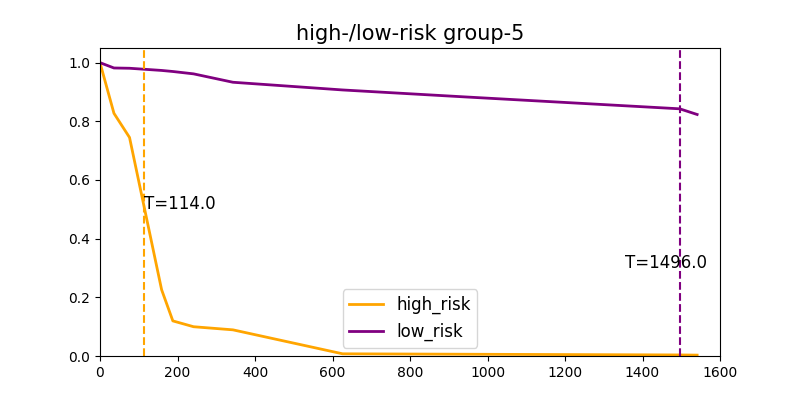


Fig S7. Individual survival probability of high-risk patient-5 and low-risk patient-5 Fig S8. Individual survival probability of high-risk patient-1 and low-risk patient-6

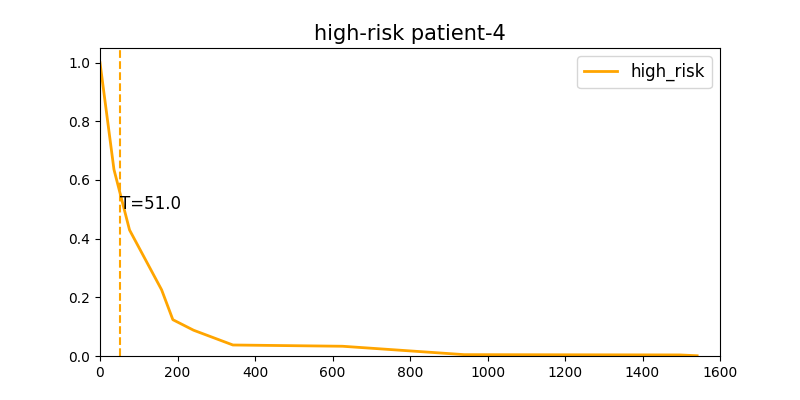
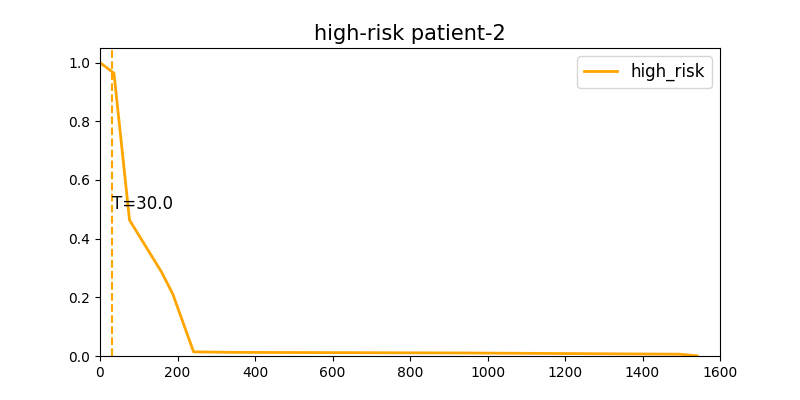


Fig S9. Individual survival probability of high-risk patient-2 Fig S10. Individual survival probability of high-risk patient-4

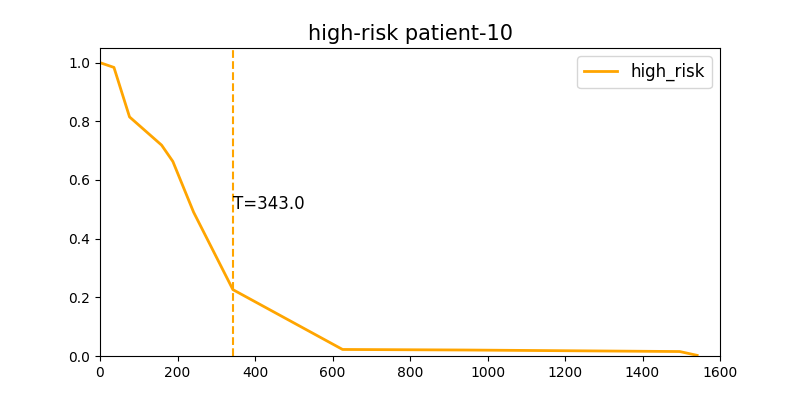
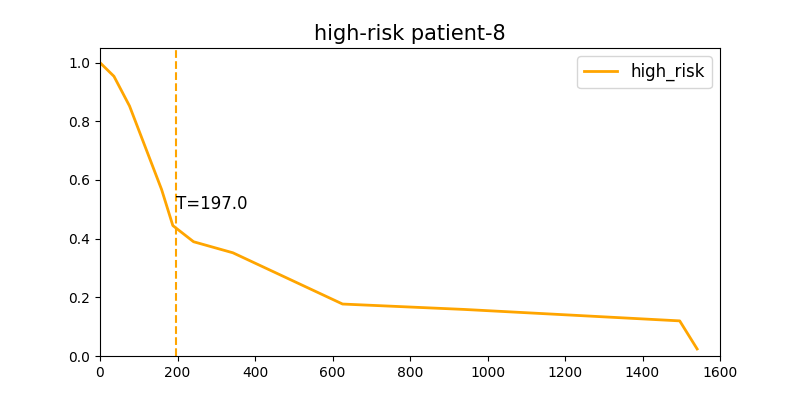


Fig S11. Individual survival probability of high-risk patient-8 Fig S12. Individual survival probability of high-risk patient-10

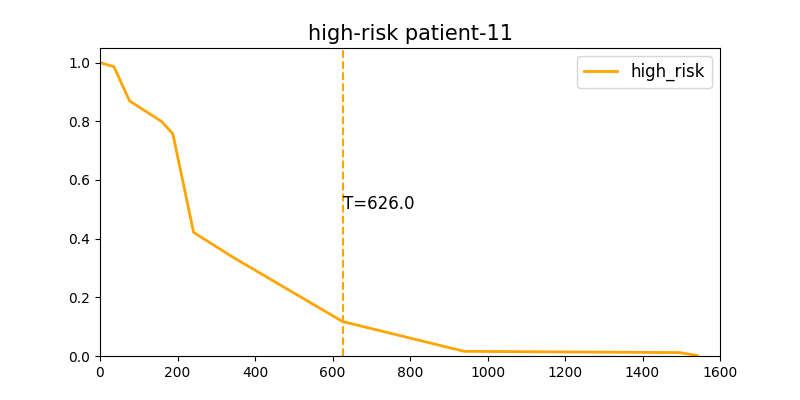


Fig S13. Individual survival probability of high-risk patient-11

**5. The comparison of different CAM methods**

Figure S14 was the comparison of the different CAM methods using our dataset. Among them, EfficientNetB7 was used as the baseline for extracting deep learning features, and global maximum pooling replaced the last fully connected layer as the final output. Compared with other methods, Score-CAM can more accurately locate the tumor area and have clearer edges. Particularly, the Smoothed Score-CAM (SS-CAM) develops an enhanced visual interpretation that produces centralized localization of object features in images through smoothing operations. However, for our experiments, the localization effect of SS-CAM method was too centralized and smooth, resulting in some tumor areas not being noticed. After weighing a large number of examples, we finally chose Score-CAM with the best effect as the method for deep learning feature visualization.

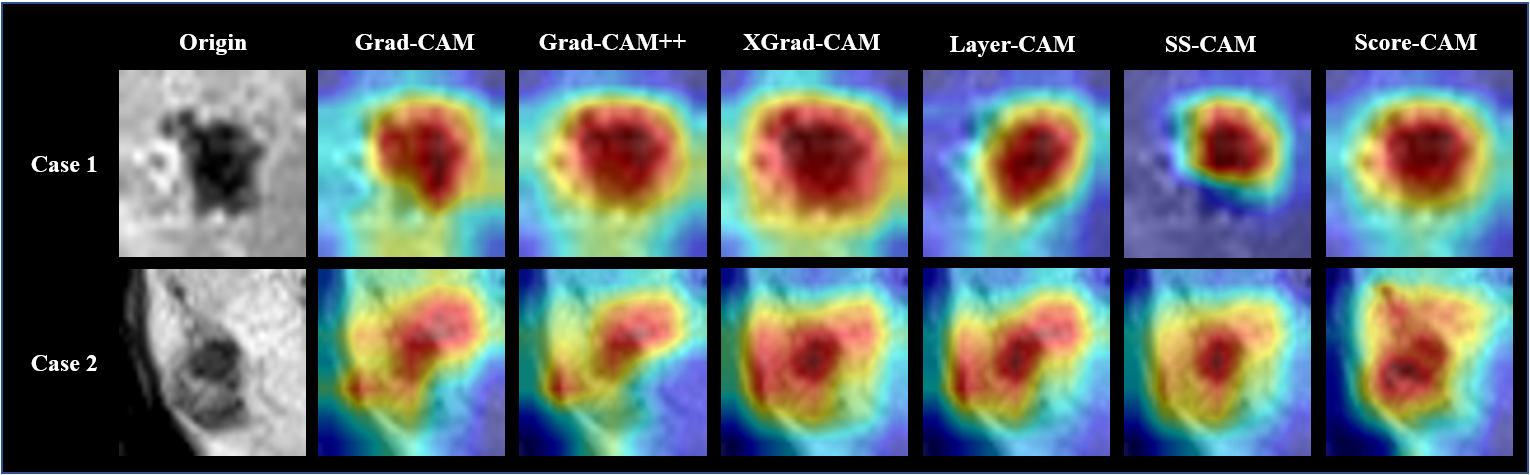


Fig S14. The comparison of different CAM methods