Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy

Hongming Li[*], Maya Galperin-Aizenberg[*], Daniel Pryma[*], Charles B. Simone II[#], Yong Fan[*]
[*]Department of Radiology, Perelman School of Medicine, the University of Pennsylvania, Philadelphia, PA 19104
[#]Maryland Proton Treatment Center, University of Maryland School of Medicine, Baltimore, MD 21201

Supplementary data

Radiomic features adopted in this study

All the radiomic features used in this study are listed in table S1, and more details could be found at http://pyradiomics.readthedocs.io/en/1.1.1/features.html.

Table S1. Radiomic features adopted in this study.

Morphometry	Volume (V) Elongation $\frac{\lambda_{longest}}{\lambda_{intermediate}}$ Spherical Disproportion $\frac{A}{\sqrt[3]{36\pi V^2}}$ Sphericity $\frac{\sqrt[3]{36\pi V^2}}{A}$ Surface Volume Ratio $\frac{A}{V}$ Flatness $\frac{\lambda_{intermediate}}{\lambda_{shortest}}$ Surface Area (A) Maximum 3D Diameter Maximum 2D Diameter axial Maximum 2D Diameter coronal Maximum 2D Diameter sagittal where $\lambda_{longest}$, $\lambda_{intermediate}$, and $\lambda_{shortest}$ are the lengths of the largest, second largest and smallest principal component axes.
First-order statistics	Minimum $min(X)$ Maximum $max(X)$ Mean $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X(i)$ Standard Deviation $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^2}$ Variance $\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^2$ Skewness $\frac{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^3}{(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^2})^3}$ Kurtosis $\frac{\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^4}{(\frac{1}{N}\sum_{i=1}^{N}(X(i)-\bar{X})^2)^2}$ Median Range $max(X) - min(X)$ 10-th Percentile P_{10} 90-th Percentile P_{90} Interquartile Range $P_{75} - P_{25}$ Mean Absolute Deviation $\frac{1}{N}\sum_{i=1}^{N}\lvert X(i)-\bar{X}\rvert$

	Robust Mean Absolute Deviation $\frac{1}{N_{10-90}}\sum_{i=1}^{N_{10-90}}\lvert X_{10-90}(i)-\bar{X}_{10-90}\rvert$ Root Mean Squared $\sqrt{\frac{1}{N}\sum_{i=1}^{N}X(i)^2}$ Uniformity $\sum_{i=1}^{N_l}p(i)^2$ Energy $\sum_{i=1}^{N}X(i)^2$ Total Energy $V_{voxel}\sum_{i=1}^{N}X(i)^2$ Entropy $-\sum_{i=1}^{N_l}p(i)\log_2(p(i)+\epsilon)$ where X denotes the 3D image matrix with N voxels, $p(i)$ is the normalized first order histogram with N_l intensity levels, V_{voxel} is the volume of the voxel in mm^3, $P_d$is the d-th percentile of the image matrix, X_{10-90} denotes the subset of image matrix with intensity levels in between the 10-th and 90-th percentile, and ϵ is an arbitrary small positive number.
Gray level co-occurrence matrix (GLCM)	Energy $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}p(i,j)^2$ Contrast $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-j)^2p(i,j)$ Entropy $-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}p(i,j)\log_2(p(i,j)+\epsilon)$ Homogeneity 1 $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+\lvert i-j\rvert}$ Correlation $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}p(i,j)ij-u_xu_y}{\sigma_x\sigma_y}$ Sum Average (SA) $\sum_{k=2}^{2N_g}kp_{x+y}(k)$ Autocorrelation $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}p(i,j)ij$ Sum Variance $\sum_{k=2}^{2N_g}(k-SE)^2p_{x+y}(k)$ Homogeneity 2 $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+\lvert i-j\rvert^2}$ Cluster Shade $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i+j-u_x-u_y)^3p(i,j)$ Maximum Probability $max(p(i,j))$ Inverse Difference Moment Normalized $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+\frac{\lvert i-j\rvert^2}{N_g^2}}$ Difference Entropy $\sum_{k=0}^{N_g-1}p_{x-y}(k)\log_2(p_{x-y}(k)+\epsilon)$ Inverse Variance $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{\lvert i-j\rvert^2}, i\neq j$ Dissimilarity $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}p(i,j)\lvert i-j\rvert$ Difference Variance $\sum_{k=0}^{N_g-1}(1-DA)^2p_{x-y}(k)$ Inverse Difference Normalized $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+\frac{\lvert i-j\rvert}{N_g}}$ Inverse Difference moment $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+\lvert i-j\rvert^2}$ Sum Entropy (SE) $\sum_{k=2}^{2N_g}p_{x+y}(k)\log_2(p_{x+y}(k)+\epsilon)$ Sum of Squares $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-u_x)^2p(i,j)$ Cluster Prominence $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i+j-u_x-u_y)^4p(i,j)$ Informal Measure of Correlation 1 $\frac{HXY-HXY1}{\max\{HX,HY\}}$ Informal Measure of Correlation 2 $\sqrt{1-e^{-2(HXY2-HXY)}}$ Difference Average (DA) $\sum_{k=0}^{N_g-1}kp_{x-y}(k)$ Inverse Difference $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+\lvert i-j\rvert}$ Cluster Tendency $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i+j-u_x-u_y)^2p(i,j)$

	Sum Variance 2 $\sum_{k=2}^{2N_g}(k-SA)^2 p_{x+y}(k)$ Average Intensity $\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)i$ where $p(i,j)$ is the normalized co-occurrence matrix, $N_g$is the number of intensity levels, $p_x(i)=\sum_{j=1}^{N_g} p(i,j)$, $p_y(j)=\sum_{i=1}^{N_g} p(i,j)$, $u_x=\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)i$, $u_y=\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)j$, $\sigma_x$and $\sigma_y$are the standard deviation of p_x and p_y, $p_{x+y}(k)=\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)$ where $i+j=k$, $p_{x-y}(k)=\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)$ where $\lvert i-j\rvert=k$, $HX=-\sum_{i=1}^{N_g} p_x(i)\log_2(p_x(i)+\epsilon)$, $HY=-\sum_{j=1}^{N_g} p_y(j)\log_2(p_y(j)+\epsilon)$, $HXY=-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log_2(p(i,j)+\epsilon)$, $HXY1=-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log_2(p_x(i)p_y(j)+\epsilon)$, $HXY2=-\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_x(i)p_y(j)\log_2(p_x(i)p_y(j)+\epsilon)$, and ϵ is an arbitrary small positive number.
Gray level run length matrix (GLRLM)	Short Run Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\lvert\theta)}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Long Run Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)j^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Gray Level Non-uniformity $\frac{\sum_{i=1}^{N_g}(\sum_{j=1}^{N_r}P(i,j\lvert\theta))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Run Length Non-uniformity $\frac{\sum_{j=1}^{N_r}(\sum_{i=1}^{N_g}P(i,j\lvert\theta))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Low Gray Level Run Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\lvert\theta)}{i^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ High Gray Level Run Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)i^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Short Run Low Gray Level Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\lvert\theta)}{i^2j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Short Run High Gray Level Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\lvert\theta)i^2}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Long Run Low Gray Level Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\lvert\theta)j^2}{i^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Long Run High Gray Level Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)i^2j^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)}$ Gray Level Variance $\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)(i-u)^2$ where $u=\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)\,i$ Run Variance $\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)(j-u)^2$ where $u=\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)\,j$ Run Entropy $-\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)\log_2(P(i,j\lvert\theta)+\epsilon)$ Run Percentage $\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\lvert\theta)}{N_p}$ Gray Level Non-uniformity Normalized $\frac{\sum_{i=1}^{N_g}(\sum_{j=1}^{N_r}P(i,j\lvert\theta))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j\lvert\theta)^2}$

	Run Length Non-uniformity Normalized $\frac{\sum_{j=1}^{N_r}(\sum_{i=1}^{N_g} P(i,j\|\theta))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} P(i,j\|\theta)^2}$ where $P(i,j\|\theta)$ is the run length matrix for direction θ, N_g is the number of intensity levels, N_r is the number of run lengths, N_p is the number of voxels, and ϵ is an arbitrary small positive number.
Gray level size zone matrix (GLSZM)	Small Area Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{P(i,j)}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Large Area Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)j^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Gray Level Non-uniformity $\frac{\sum_{i=1}^{N_g}(\sum_{j=1}^{N_s} P(i,j))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Size Zone Non-uniformity $\frac{\sum_{j=1}^{N_s}(\sum_{i=1}^{N_g} P(i,j))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Zone Percentage $\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{P(i,j)}{N_p}$ Low Gray Level Zone Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{P(i,j)}{i^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ High Gray Level Zone Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)i^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Small Area Low Gray Level Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{P(i,j)}{i^2 j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Small Area High Gray Level Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{P(i,j)i^2}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Large Area Low Gray Level Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{P(i,j)j^2}{i^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Large Area High Gray Level Emphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)i^2 j^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)}$ Gray Level Variance $\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)(i-u)^2$ where $u = \sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)\, i$ Zone Variance $\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)(j-u)^2$ where $u = \sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)\, j$ Gray Level Non-uniformity Normalized $\frac{\sum_{i=1}^{N_g}(\sum_{j=1}^{N_s} P(i,j))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)^2}$ Size Zone Non-uniformity Normalized $\frac{\sum_{j=1}^{N_s}(\sum_{i=1}^{N_g} P(i,j))^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)^2}$ Zone Entropy $-\sum_{i=1}^{N_g}\sum_{j=1}^{N_s} P(i,j)\log_2(P(i,j)+\epsilon)$ where $P(i,j)$ is the size zone matrix, N_g is the number of intensity levels, N_s is the number of zone sizes, N_p is the number of voxels, and ϵ is an arbitrary small positive number.

Unsupervised two-way clustering

The clustering result with 3 clusters regarding patients and 10 clusters regarding features obtained by the unsupervised two-way clustering results is illustrated in Fig. S1. As shown in the Fig. S1 (bottom left and right), highly correlated features and subjects with similar radiomic features were grouped into the same clusters simultaneously, and blocks along the diagonal demonstrated high correlation coefficients within the same cluster. The meta-features obtained by the two-way clustering were weighted combination representations of the original features within the same feature clusters, which reduced the feature dimensionality without losing discriminative information of the original features. The clustering result regarding subjects was obtained based on the meta-features, which was more robust to feature noise and could lead to better subject stratification than simple clustering methods, such as k-means. As both the clustering results of features and subjects were optimized jointly, they could benefit from each other and lead to better subject stratification at a group level and generate more discriminative meta-features for survival prediction analyses at an individual subject level.

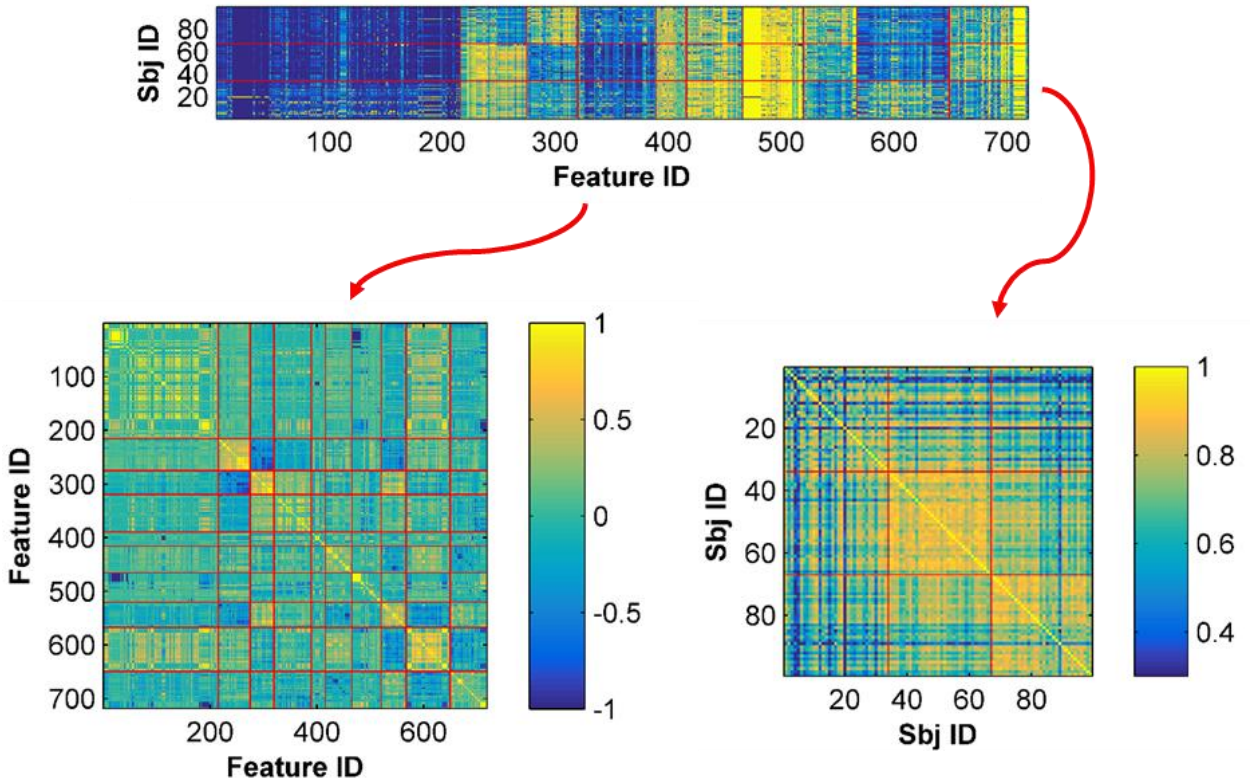


Fig. S1. Clustering results obtained by the unsupervised two-way clustering method, with 3 clusters regarding subjects and 10 clusters regarding features. Clustering results overlaid on the original feature matrix of all the subjects (top), Pearson correlation matrix between all the features (bottom left), and Pearson correlation matrix between all the subjects (bottom right). Different clusters are separated by red lines.

Patient stratification based on T stages

Patient stratification has also been carried out based on their T stages. Three subgroups of T1a, T1b, and T2 were obtained, and no significant difference was observed between different subgroups. Kaplan–Meier plots of survival and nodal failure are shown in Fig. S2.

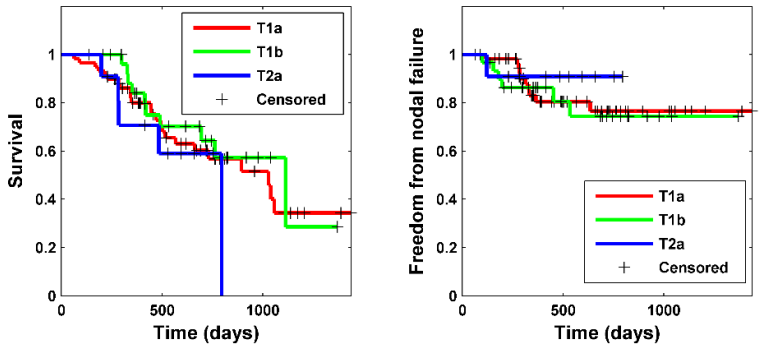


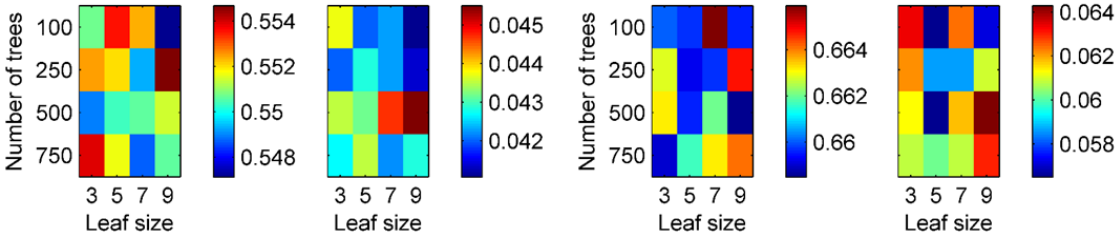
Fig. S2. Kaplan–Meier plots of subgroups with different T stages regarding survival and freedom from nodal failure.

Identification of the number of meta-features used for survival analysis at group level

The number of meta-features was set to 10 according to its cross-validation performance for survival prediction. Specifically, a grid searching was adopted to evaluate survival models with the number of patient clusters set to 2 or 3, and the number of meta-features ranging from 5 to 11 with an increment of 1. Cox regression, Cox_lasso model, and RSF model were trained and evaluated regarding survival and nodal failure under the same 3-fold cross-validation setting. The cross-validation procedure was repeated 100 times, and the number of meta-features (i.e., 10) most frequently selected with the best prediction performance was adopted to evaluate the performance of patient stratification.

Parameters in the random survival forests based model

We have tested different settings for the RSF model, with different number of trees including 100, 250, 500, and 750, and leaf sizes varying from 3 to 9 with an increment of 2. The prediction performance with different parameter settings regarding survival and nodal failure is shown in Fig. S3, indicating that the prediction was stable in terms of both survival and nodal failure.



(a) Survival (b) Nodal failure

Fig. S3. Prediction performance using random survival forests with different parameter settings regarding survival (a) and nodal failure (b). The left matrix shows the mean c-index measures of 100 repetitions of cross-validation, and the right matrix shows their standard deviations.

Prediction models built upon the tumor volume and SUVmax measures

We have built Cox regression models to explore whether the tumor volume and SUVmax measures are correlated with survival and nodal failure. The resulting Cox models indicated that the tumor volume was significantly correlated with survival ($p < 0.05$) and nodal failure ($p < 0.0005$), but the SUVmax measure was not significantly correlated with survival or nodal failure ($p > 0.5$).

We have further investigated the performance of prediction models built upon the tumor volume and SUVmax measures using the same 3 fold cross-validation. For the survival prediction, c-index values obtained by the Cox model, Cox_lasso, and the RSF model were 0.551±0.047, 0.560±0.081, and 0.562±0.033 respectively. For the prediction of nodal failure, c-

index values of these models were 0.663±0.044, 0.604±0.059, and 0.637±0.044 respectively. These prediction models had worse performance than those built upon more complex radiomic features (0.640±0.029 and 0.664±0.063 for survival and nodal failure, respectively).

Prediction models built upon radiomic features obtained with alternative gray-level discretization scheme

Max-Lloyd algorithm [1] was adopted as an alternative discretization method for extracting radiomic features, in order to investigate its prediction performance regarding survival and nodal failure. Particularly, different numbers of gray levels were adopted (32 and 64) to extract radiomic features, and prediction models for survival and local nodal failure were then built upon meta-features extracted using different methods and evaluated using the same cross-validation setting as described in the main document.

With the number of gray levels set to 32, the Cox, Cox_lasso, and RSF models built upon the meta-features extracted using the two-way clustering method for predicting survival obtained c-index values of 0.566 ± 0.043, 0.531 ± 0.059, and 0.493 ± 0.046 respectively. The Cox, Cox_lasso, and RSF models built upon the two-way clustering based meta-features for predicting nodal failure obtained c-index values of 0.688 ± 0.043, 0.675 ± 0.047, and 0.640 ± 0.047 respectively. The Cox, Cox_lasso, and RSF models built upon the PCA based meta-features for predicting survival obtained c-index values of 0.558 ± 0.047, 0.518 ± 0.056, and 0.495 ± 0.042 respectively. The Cox, Cox_lasso, and RSF models built upon the PCA based meta-features for predicting nodal failure obtained c-index values of 0.681 ± 0.052, 0.661 ± 0.046, and 0.577 ± 0.054 respectively.

With the number of gray level set to 64, the Cox, Cox_lasso, and RSF models built upon the meta-features extracted using the two-way clustering method for predicting survival obtained c-index values of 0.572 ± 0.048, 0.545 ± 0.057, and 0.496 ± 0.042 respectively. The Cox, Cox_lasso, and RSF models built upon the two-way clustering based meta-features for predicting nodal failure obtained c-index values of 0.688 ± 0.047, 0.677 ± 0.047, and 0.638 ± 0.044 respectively. The Cox, Cox_lasso, and RSF models built upon the PCA based meta-features for predicting survival obtained c-index values of 0.565 ± 0.042, 0.526 ± 0.047, and 0.484 ± 0.039 respectively. The Cox, Cox_lasso, and RSF models built upon the PCA based meta-features for predicting nodal failure obtained c-index values of 0.684 ± 0.041, 0.659 ± 0.051, and 0.556 ± 0.049 respectively.

Overall, the prediction models built upon the meta-features extracted using the two-way clustering method obtained better prediction performance than models built upon the PCA based meta-features when Max-Lloyd algorithm was adopted as an alternative discretization method for extracting radiomic features.

References

[1] Vallieres M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Physics in medicine and biology. 2015;60:5471-96.