

Supplementary Materials

Censor-aware semi-supervised survival time prediction in lung cancer using clinical and radiomics features

Arman Groji¹, Ali Fathi Jouzdani¹, Nima Sanati¹, Ren Yuan^{2,3}, Arman Rahmim^{3,4}, Mohammad R. Salmanpour^{3,4,5}

¹Department of Neuroscience, Hamadan University of Medical Sciences, Hamadan 65176-19657, Iran.

²BC Cancer, Vancouver Center, Vancouver V5Z 4E6, Canada.

³Department of Radiology, University of British Columbia, Vancouver V5Z 1M9, Canada.

⁴Department of Basic and Translational Research, BC Cancer Research Institute, Vancouver V5Z 1L3, Canada.

⁵Technological Virtual Collaboration (TECVICO CORP.), Vancouver V5E 3H7, Canada.

Correspondence to: Mohammad R. Salmanpour, Department of Basic and Translational Research, BC Cancer Research Institute, Vancouver V5Z 1L3, Canada.
E-mail: msalman@bccrc.ca

1. Methods and Material

Supplemental Table S1. Clinical features description

Features	Meaning
Surgery category	The types of surgery that patients have such as 'pneumonectomy', 'lobectomy', 'segmentectomy'
Ethnicity	Tells the patient is Asian or not.
Postoperative residual tumor	Explains post-operation residual tumor with 'Microscopic', 'Macroscopic', and 'Not Applicable' values
Primary tumour resected	Describes Primary Tumour Resected in surgery with 'Yes', 'No', and 'Not Applicable' values
Radiation therapy	Shows if a person has radiation therapy or not

Dose category	The total dose of radiotherapy
Histology	Describe the type of tumor Squamous cell carcinoma, Adenocarcinoma etc.
Stage at diagnose	Describes the cancer stage at diagnosis time
Smoking category	Describes three categories of smoking including current smokers, former smokers, and non-smokers
Smoking Pack Years	The pack-year is a unit for measuring the amount a person has smoked over a long period of time. It is calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the person has smoked
weight loss category	Describe the amount of weight loss
Chemotherapy	Some patients have chemotherapy, and some others don't.
Sex	Male or Female
Recurrence status	means that the cancer has come back in the same place it first started.regional recurrence means that the cancer has come back in the lymph nodes near the place it first started. distant recurrence means the cancer has come back in another part of the body; some distance from where it started
Metastasis	Describes whether a patient has metastasis or not
Correct baseline ECOG	Describes 0—Fully active, 1—Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, 2—Ambulatory and capable of all self-care but unable to carry out any work activities, 3—Capable of only limited self-care, 4—Completely disabled, 5—Dead
duration	Between the diagnosis date and Recurrence date
Age at diagnosis	Describes the age at diagnosis

1.1. Image preprocessing and mask preparation

1.1.1. Attenuation correction (AC)

Attenuation correction (AC) is one of the most important corrections that need to be performed in PET imaging. AC methods aim to account for the photon attenuation

along each line of response (LOR). For this purpose, maps of the linear attenuation coefficients (LACs) for all the tissues and materials located in the PET field of view are generated or integrals of these values along all the LORs are directly measured. The procedure for including this information in the reconstruction (performing the actual correction) is similar [1].

1.1.2. Standardized Uptake Value (SUV) correction

SUV is a quantitative measure used in PET imaging to supplement visual interpretation of PET images. It is a ratio of tissue radioactivity concentration to injected dose per unit body weight, normalized to the whole-body radioactivity concentration. SUV is a semiquantitative measure, simpler to calculate than fully quantitative measures like the influx constant K. It can be corrected for factors like body weight, lean body mass, and body surface area [2].

1.1.3. Min/max normalization technique

Min-max normalization is a technique that transforms the values of a numerical variable to a standard range, usually between 0 and 1. It is useful for comparing data that have different scales or units of measurement. The formula for min-max normalization is:

$$x' = (x - \min) / (\max - \min)$$

where x is the original value, x' is the normalized value, $\min$ is the minimum value of the variable, and $\max$ is the maximum value of the variable. This technique is also known as feature scaling or rescaling [3, 4].

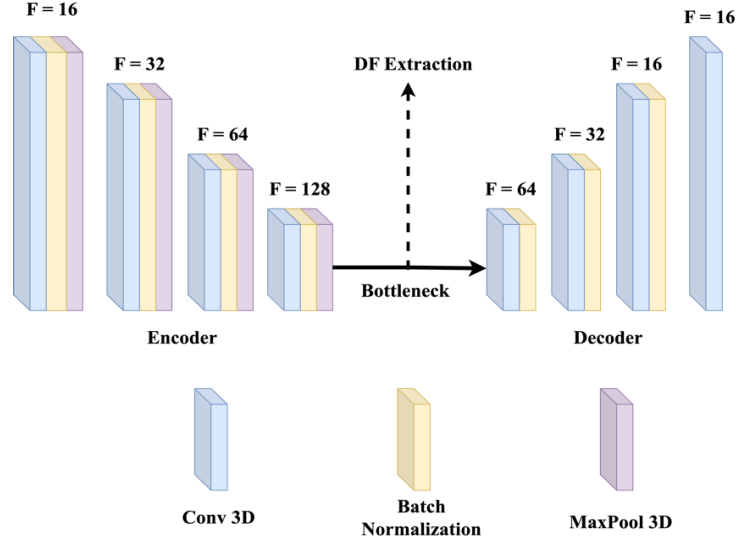
1.2. Feature extraction

1.2.1. HRF Extraction

In our investigation, we utilized two distinct strategies for image feature extraction: HRF and DRF. Within the HRF strategy, a total of 215 quantitative HRFs were derived from each delineated tumor region using the PySERA standardization process available in the ViSERA software [5]. The radiomic feature generation process in ViSERA aligns with the protocols established by the Image Biomarker Standardization Initiative (ISBI). The 215 HRFs encompassed a variety of feature types: 29 shape descriptors, 20 first-order statistics (FO), 30 intensity histogram parameters (IH), and 136 textural features. These textural features included 50 co-

occurrence matrix attributes (CMs), 32 run-length matrix characteristics (RLMs), 16 size zone metrics (SZMs), 16 distance zone parameters (DZMs), 5 neighborhood gray-tone difference matrix elements (NGTs), and 17 neighboring gray level dependence matrix indices (NGLs). HRFs were extracted from the segmented PET and CT images.

1.2.2. DRF Extraction



Supplemental Figure S1. Structure of our autoencoder model. The Encoder includes four convolutional layers, each followed by a batch normalization and a max-pooling operation. The decoder path includes four convolutional layers that three of them are followed by a batch normalization. F: Filters.

1.3. Pearson's correlation coefficient regression (R_Regression)

Compute Pearson's r for each feature and the target. Pearson's r is also known as the Pearson correlation coefficient. Linear model for testing the individual effect of each of many regressors. This is a scoring function to be used in a feature selection procedure, not a free-standing feature selection procedure. The cross-correlation between each regressor and the target is computed as [6]:

$$E[(X[:, i] - \text{mean}(X[:, i])) * (y - \text{mean}(y))] / (\text{std}(X[:, i]) * \text{std}(y))$$

1.4.F-test for regression (F_Regression)\

Univariate linear regression tests returning F-statistic and p-values. Quick linear model for testing the effect of a single regressor, sequentially for many regressors. This is done in 2 steps: The cross correlation between each regressor and the target is computed using `r_regression` as:

$$E[(X[:, i] - \text{mean}(X[:, i])) * (y - \text{mean}(y))] / (\text{std}(X[:, i]) * \text{std}(y))$$

It is converted to an F score and then to a p-value. `F_Regression` is derived from `R_Regression` and will rank features in the same order if all the features are positively correlated with the target [7].

1.5.Principal Component Analysis (PCA)

PCA is a statistical unsupervised technique used in data analysis and Machine Learning (ML) to simplify the complexity of high-dimensional data while retaining most of the information. It works by identifying the directions, called principal components, along which the variability of the data is maximized. These components are orthogonal to each other and are derived from the eigenvectors of the data's covariance matrix, ranked according to their eigenvalues. In this study, PCA reduces the dimensionality by projecting the original data onto a smaller set of 10 significant components to prevent overfitting in ML prediction algorithm [8].

1.6. Weibull Accelerated failure time (AFT)

Accelerated failure time (AFT) models are used widely in medical research, though to a much lesser extent than proportional hazards models. In an AFT model, the effect of covariates act to accelerate or decelerate the time to event of interest, that is, shorten or extend the time to event [9]. The Weibull model is a parametric survival model in which the outcome is assumed to follow a Weibull distribution. A unique property of the Weibull model is that when the accelerated failure time (AFT) assumption holds, the proportional hazards (PH) assumption also holds, making it compatible with both frameworks. In the PH formulation, the key assumption is that hazard ratios remain constant over time, whereas in the AFT formulation, survival times are assumed to accelerate or decelerate by a constant factor across covariate levels. This dual property is specific to the Weibull model and makes it especially useful in survival analysis [10].

1.7. Regression Algorithms (RA)

1.7.1. K-Nearest Neighbour

In supervised machine learning, the K-Nearest Neighbor (KNN) algorithm is a straightforward yet powerful technique for regression and classification. It works on the tenet that data points with comparable traits have a tendency to group together. KNN finds the 'K' nearest points (neighbors) in the training data when a new data point is added, and then assigns a value based on the average value for regression tasks or the majority class for classification tasks. The algorithm's non-parametric nature, which means it doesn't assume any particular distribution for the data points, is one of its advantages. This makes it adaptable to a variety of datasets. The performance of KNN is largely dependent on the value of 'K', the number of neighbors considered. An ideal 'K' balances the model's bias and variance to prevent overfitting or underfitting. Higher 'K' values produce more stable predictions because they analyze more data points and are less sensitive to noise. However, this can result in smoother decision boundaries, perhaps missing local nuances in the data. The approach relies on an appropriate distance measure, such as Euclidean or Manhattan distance, to locate the nearest neighbors [11].

1.7.2. AdaBoost Regressor (ABR)

Adaboost learning is a potent machine learning technique that enables the construction of highly accurate prediction models through the combination of multiple weak models. Adaboost, fundamentally, operates as an ensemble method wherein decision trees are iteratively trained to rectify the errors introduced by the preceding tree. By training each subsequent decision tree using the residuals of the previous tree, the model is able to evolve and develop gradually. The ultimate prognosis is computed by combining the predictions generated by each individual tree, assigning a weight to its accuracy-based contribution. This procedure enables Adaboost to effortlessly manage high-dimensional data and capture intricate interactions among features. In addition, built-in regularisation in Adaboost ensures that the model generalises well to new data and prevents overfitting. In general, Adaboost learning provides a resilient and adaptable approach for constructing accurate prediction models, rendering it a favoured option across an extensive variety of applications encompassing computer vision, natural language processing, and other domains [12].

1.7.3. Random Forest Regressor (RFR)

A Random Forest Regressor is a form of machine learning model that predicts continuous outcomes using a collection of decision trees. It is widely used in regression problems that aim to predict a numerical value due to its capability of managing complicated relationships among variables and delivering precise estimates of uncertainty. Similar to other ensemble methods, a Random Forest Regressor generates a final output by combining the predictions of multiple trees. A random subset of the training data and features is used to train each tree, and the final prediction is calculated by summing the outputs of all trees. The utilisation of randomness in the feature and data selection processes serves to mitigate overfitting and enhances the model's capacity to extrapolate to novel data. In addition, the model is robust against outliers and can accommodate absent values. Applications such as finance, marketing, and healthcare, where predicting continuous outcomes is crucial and the model's ability to provide accurate estimates of uncertainty is valuable, frequently employ Random Forest Regressors [13].

1.7.4. Multi-Layer Perceptron (MLPR)

For regression tasks, Multi-Layer Perceptron (MLP) regressors are a type of neural network. An MLP regressor, as compared to a linear regression model that operates under the assumption of a linear association between the input features and output variables, employs multiple layers of artificial neurons to discover non-linear relationships. An input layer, one or more concealed layers, and an output layer create the model. Every hidden layer is composed of a specific group of neurons. In turn, each neuron executes a computation on the inputs it receives from the previous layer and subsequently transmits the result to the neurons in the following layer. The final output is generated by applying an activation function to the output of the final hidden layer. One notable benefit of MLP regressors is their capacity to identify intricate patterns within datasets, which enables them to model non-linear associations between inputs and outputs with remarkable accuracy. They are frequently implemented in medical diagnosis, stock price prediction, and weather forecasting, among other applications. Nevertheless, MLP regressors are susceptible to overfitting, particularly when challenged with enormous datasets, and require precise hyperparameter optimisation in order to attain peak performance [14].

1.7.5. Decision Tree Regressor (DTR)

A Decision Tree Regressor is a form of machine learning model that predicts continuous outcomes using a tree-based structure. In contrast to the predefined categories employed by a conventional decision tree, a decision tree regressor generates predictions for continuous values, such as probabilities or prices. Based on the values of the input features, the model recursively divides the data into smaller subsets before fitting a statistical model, such as linear regression, to the data points within each subset. In order to arrive at the ultimate prediction, the tree is traversed from its root to each leaf node, where the predicted value is calculated using the attributes linked to that specific leaf node. When the relationships between the input features and the output variable are complex, decision tree regressors are beneficial; they can process both numeric and categorical features. In addition to being interpretable, which enables users to comprehend the methodology behind the model's predictions, these models are also applicable to classification and regression tasks. Nevertheless, decision tree regressors are susceptible to overfitting, which occurs

when the dataset contains an excessive number of irrelevant features; therefore, care must be taken to select the proper features and adjust the model's parameters [15].

1.7.6. Linear Regression (LR)

In machine learning, linear regression is a widely implemented algorithm employed for tasks involving the prediction of a continuous outcome variable from one or more input features. Linear regression is a statistical method that models the relationship between an output variable and input features through the use of a linear equation. The coefficients of the features are acquired through training with the data. The model operates under the assumption of a linear relationship between the output variable and the features, whereby any change in a feature results in a constant effect on the output variable, irrespective of the feature's level. As an algorithm for supervised learning, Linear Regression requires labelled training data in order to discover the connection between the input variable and the features. As a result of minimising the sum of squared differences between the predicted and actual values, the algorithm determines the linear equation that best describes the relationship between the output variable and the features. Linear regression is an exceptionally potent instrument used across many fields, including finance, economics, social sciences, and engineering, to predict continuous outcomes. Additionally, it provides an excellent basis for comprehending advanced algorithms in machine learning, including logistic regression and support vector machines [16].

1.7.7. Support Vector Machine (SVR)

For regression tasks, Support Vector Machine (SVM) regression is a type of machine learning algorithm. In contrast to conventional linear regression, which models the relationship between independent and dependent variables using a linear function, SVM regression maps inputs to outputs using a non-linear function. SVM regression operates on the principle of locating the hyperplane that divides the data points into distinct classes at the highest level possible. The objective of regression tasks is to identify the hyperplane where the mean squared error between the predicted and actual values is minimised. Support vector regression accomplishes this objective by optimising the margin separating the hyperplane from the closest data points. As the distance between the hyperplane and the support vectors, the margin signifies the model's robustness. A greater margin indicates an enhanced capacity of the model to

extrapolate to novel data. Support Vector Machine (SVM) regression proves to be highly advantageous when confronted with chaotic or unstructured data. Its usage extends across diverse domains, including finance, biology, and computer vision [17].

1.8. Hazard Ratio Survival Analysis (HRSA)

1.8.1. Fast Survival Support Vector Machines (FSSVM)

An expansion of the regular Support Vector Machine to right-censored time-to-event data is called a Survival Support Vector Machine. Its main benefit is that, with the use of the so-called kernel trick, it can take into account intricate, non-linear correlations between survival and attributes. The input features are implicitly mapped by a kernel function into high-dimensional feature spaces where a hyperplane can describe survival. Because of this, Survival Support Vector Machines may be applied to a large variety of data and are highly versatile [18].

1.8.2. Component-wise Gradient Boosting Survival Analysis (CWGBSA)

Gradient Boosting is a flexible framework for optimizing a variety of loss functions rather than a specific model. It builds a strong overall model by aggregating the predictions of several base learners, according to the strength in numbers concept. The base learners are sometimes known as weak learners since they are frequently relatively basic models that perform slightly better than random guessing. The predictions are combined additively, meaning that the addition of a base model improves the model as a whole [19].

1.8.3. Random Survival Forest (RSF)

An ensemble of tree-based learners makes up a Random Survival Forest, just like its well-known counterparts for regression and classification. By 1) generating each tree using a unique bootstrap sample of the original training data and 2) only evaluating the split criterion for a randomly chosen subset of features and thresholds at each node, a Random Survival Forest ensures that individual trees are de-correlated. The ensemble's predictions are created by adding up the predictions of each individual tree [20].

1.9.Pseudo-labeling

Pseudo-labeling [21] is a technique for semi-supervised learning, which is a type of ML that uses both labeled and unlabeled data. Pseudo-labeling works by using a model trained on labeled data to predict the labels for unlabeled data, and then using those “pseudo labels” to train the model in a supervised way on the unlabeled data. This can help improve the accuracy and generalization of the model, especially when there is not enough labeled data available. Of course, ultimately the performance of such a framework is tested on a fully labeled dataset.

2. Results

Supplemental Table S2. Results provided by regression algorithms (RA)

<i>Dataset</i>	<i>C- Index</i>	<i>FSA</i>	<i>Algorit hm</i>	<i>SL RMSE</i>	<i>SL MAE</i>	<i>SL R2</i>	<i>SSL RMSE</i>	<i>SSL MAE</i>	<i>SSL R2</i>	<i>T-test p- value</i>
<i>CF</i>	0.68	FR	KNN	1.41	1.07	0.00	1.52	1.19	-0.17	0.00
<i>CF</i>	0.68	RR	KNN	1.41	1.07	0.00	1.52	1.19	-0.17	0.00
<i>CF</i>	0.56	RR	ADB	1.42	1.09	-0.02	1.39	1.04	0.03	0.05
<i>CF</i>	0.68	FR	LRR	1.41	1.08	0.00	1.58	1.24	-0.25	0.13
<i>CF</i>	0.68	RR	LRR	1.41	1.08	0.00	1.58	1.24	-0.25	0.13
<i>CF</i>	0.68	FR	DTC	2.33	1.66	-1.73	1.79	1.37	-0.61	0.13
<i>CF</i>	0.68	RR	DTC	2.33	1.66	-1.73	1.79	1.37	-0.61	0.13
<i>CF</i>	0.68	FR	SVR	1.49	1.11	-0.11	1.53	1.21	-0.17	0.22
<i>CF</i>	0.68	RR	SVR	1.49	1.11	-0.11	1.53	1.21	-0.17	0.22
<i>CF</i>	0.68	FR	RFR	1.51	1.16	-0.14	1.54	1.21	-0.18	0.31
<i>CF</i>	0.68	RR	RFR	1.51	1.16	-0.14	1.54	1.21	-0.18	0.31
<i>CF</i>	0.68	FR	MLP	1.43	1.14	-0.02	1.53	1.21	-0.17	0.37
<i>CF</i>	0.68	RR	MLP	1.43	1.15	-0.02	1.46	1.16	-0.07	0.84
<i>CF</i>	0.64	FR	ADB	1.54	1.14	-0.19	1.47	1.13	-0.08	0.95
<i>CT_DRF</i>	0.70	FR	RFR	1.58	1.20	-0.25	9.02	3.62	-39.80	0.04

<i>CT_DRF</i>	0.70	FR	DTC	2.04	1.45	-1.08	8.79	3.71	-37.82	0.05
<i>CT_DRF</i>	0.65	FR	ADB	1.60	1.13	-0.28	1.49	1.03	-0.12	0.07
<i>CT_DRF</i>	0.50	RR	MLP	1.41	1.09	0.00	1.55	1.18	-0.21	0.07
<i>CT_DRF</i>	0.70	FR	SVR	1.54	1.15	-0.19	15.32	4.81	-116.84	0.09
<i>CT_DRF</i>	0.70	FR	LRR	2.73	1.65	-2.73	17.98	5.23	-161.17	0.11
<i>CT_DRF</i>	0.70	FR	MLP	1.79	1.38	-0.60	14.82	4.62	-109.19	0.11
<i>CT_DRF</i>	0.70	FR	KNN	1.52	1.16	-0.16	7.90	2.62	-30.31	0.18
<i>CT_DRF</i>	0.50	RR	ADB	1.41	1.06	0.00	1.41	1.09	0.00	0.26
<i>CT_DRF</i>	0.50	RR	LRR	1.41	1.08	0.00	1.49	1.10	-0.11	0.73
<i>CT_DRF</i>	0.50	RR	DTC	1.41	1.08	0.00	1.43	1.07	-0.02	0.73
<i>CT_DRF</i>	0.50	RR	RFR	1.41	1.09	0.00	1.47	1.07	-0.08	0.77
<i>CT_DRF</i>	0.50	RR	SVR	1.42	1.05	-0.02	1.44	1.06	-0.03	0.78
<i>CT_DRF</i>	0.50	RR	KNN	1.41	1.08	0.00	1.50	1.08	-0.12	0.95
<i>CT_DRF+CF</i>	0.65	FR	MLP	1.74	1.30	-0.52	1.60	1.13	-0.28	0.00
<i>CT_DRF+CF</i>	0.65	FR	RFR	1.58	1.12	-0.25	1.52	1.06	-0.16	0.12
<i>CT_DRF+CF</i>	0.50	RR	MLP	1.41	1.08	0.00	1.52	1.15	-0.17	0.19
<i>CT_DRF+CF</i>	0.50	RR	ADB	1.41	1.09	0.00	1.41	1.07	0.00	0.26
<i>CT_DRF+CF</i>	0.65	FR	LRR	1.63	1.18	-0.33	1.62	1.12	-0.32	0.45
<i>CT_DRF+CF</i>	0.65	FR	DTC	1.83	1.22	-0.67	1.69	1.13	-0.44	0.45

<i>CT_DRF+CF</i>	0.65	FR	ADB	1.49	1.04	-0.11	1.54	1.07	-0.18	0.65
<i>CT_DRF+CF</i>	0.50	RR	LRR	1.41	1.08	0.00	1.49	1.10	-0.11	0.73
<i>CT_DRF+CF</i>	0.50	RR	DTC	1.41	1.08	0.00	1.43	1.07	-0.02	0.73
<i>CT_DRF+CF</i>	0.50	RR	RFR	1.41	1.09	0.00	1.47	1.07	-0.08	0.77
<i>CT_DRF+CF</i>	0.50	RR	SVR	1.42	1.05	-0.02	1.44	1.06	-0.03	0.78
<i>CT_DRF+CF</i>	0.65	FR	SVR	1.46	1.08	-0.08	1.53	1.09	-0.17	0.86
<i>CT_DRF+CF</i>	0.65	FR	KNN	1.50	1.10	-0.12	1.54	1.09	-0.19	0.91
<i>CT_DRF+CF</i>	0.50	RR	KNN	1.41	1.08	0.00	1.50	1.08	-0.12	0.95
<i>CT_DRF+CF+CT_HRF</i>	0.65	FR	MLP	1.80	1.31	-0.62	1.56	1.12	-0.22	0.02
<i>CT_DRF+CF+CT_HRF</i>	0.65	FR	RFR	1.58	1.12	-0.25	1.52	1.06	-0.16	0.12
<i>CT_DRF+CF+CT_HRF</i>	0.50	RR	ADB	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>CT_DRF+CF+CT_HRF</i>	0.50	RR	MLP	1.41	1.09	0.00	1.49	1.13	-0.12	0.35
<i>CT_DRF+CF+CT_HRF</i>	0.65	FR	LRR	1.63	1.18	-0.33	1.62	1.12	-0.32	0.45
<i>CT_DRF+CF+CT_HRF</i>	0.65	FR	DTC	1.83	1.22	-0.67	1.69	1.13	-0.44	0.45
<i>CT_DRF+CF+CT_HRF</i>	0.65	FR	ADB	1.46	1.04	-0.08	1.45	1.02	-0.05	0.65
<i>CT_DRF+CF+CT_HRF</i>	0.50	RR	LRR	1.41	1.08	0.00	1.49	1.10	-0.11	0.73
<i>CT_DRF+CF+CT_HRF</i>	0.50	RR	DTC	1.41	1.08	0.00	1.43	1.07	-0.02	0.73
<i>CT_DRF+CF+CT_HRF</i>	0.50	RR	RFR	1.41	1.09	0.00	1.47	1.07	-0.08	0.77
<i>CT_DRF+CF+CT_HRF</i>	0.50	RR	SVR	1.42	1.05	-0.02	1.44	1.06	-0.03	0.78

<i>CT_DRF+CF+CT_HRF</i>	0.65	FR	SVR	1.46	1.08	-0.08	1.53	1.09	-0.17	0.86
<i>CT_DRF+CF+CT_HRF</i>	0.65	FR	KNN	1.50	1.10	-0.12	1.54	1.09	-0.19	0.91
<i>CT_DRF+CF+CT_HRF</i>	0.50	RR	KNN	1.41	1.08	0.00	1.50	1.08	-0.12	0.95
<i>CT_DRF+CT_HRF</i>	0.65	FR	MLP	1.76	1.30	-0.56	1.56	1.12	-0.22	0.03
<i>CT_DRF+CT_HRF</i>	0.65	FR	RFR	1.58	1.12	-0.25	1.52	1.06	-0.16	0.12
<i>CT_DRF+CT_HRF</i>	0.50	RR	MLP	1.41	1.08	0.00	1.53	1.15	-0.18	0.13
<i>CT_DRF+CT_HRF</i>	0.50	RR	ADB	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>CT_DRF+CT_HRF</i>	0.65	FR	LRR	1.63	1.18	-0.33	1.62	1.12	-0.32	0.45
<i>CT_DRF+CT_HRF</i>	0.65	FR	DTC	1.83	1.22	-0.67	1.69	1.13	-0.44	0.45
<i>CT_DRF+CT_HRF</i>	0.50	RR	LRR	1.41	1.08	0.00	1.49	1.10	-0.11	0.73
<i>CT_DRF+CT_HRF</i>	0.50	RR	DTC	1.41	1.08	0.00	1.43	1.07	-0.02	0.73
<i>CT_DRF+CT_HRF</i>	0.50	RR	RFR	1.41	1.09	0.00	1.47	1.07	-0.08	0.77
<i>CT_DRF+CT_HRF</i>	0.50	RR	SVR	1.42	1.05	-0.02	1.44	1.06	-0.03	0.78
<i>CT_DRF+CT_HRF</i>	0.65	FR	ADB	1.60	1.12	-0.28	1.57	1.11	-0.24	0.85
<i>CT_DRF+CT_HRF</i>	0.65	FR	SVR	1.46	1.08	-0.08	1.53	1.09	-0.17	0.86
<i>CT_DRF+CT_HRF</i>	0.65	FR	KNN	1.50	1.10	-0.12	1.54	1.09	-0.19	0.91
<i>CT_DRF+CT_HRF</i>	0.50	RR	KNN	1.41	1.08	0.00	1.50	1.08	-0.12	0.95
<i>CT_HRF</i>	0.74	RR	KNN	1.41	1.09	0.00	1.49	1.21	-0.11	0.12
<i>CT_HRF</i>	0.61	FR	LRR	1.50	1.10	-0.13	1.47	1.05	-0.08	0.14

<i>CT_HRF</i>	0.74	RR	SVR	1.77	1.41	-0.58	1.62	1.24	-0.31	0.24
<i>CT_HRF</i>	0.74	RR	LRR	1.45	1.10	-0.06	1.60	1.20	-0.29	0.27
<i>CT_HRF</i>	0.59	RR	ADB	1.50	1.13	-0.12	1.43	1.08	-0.03	0.28
<i>CT_HRF</i>	0.61	FR	DTC	1.71	1.28	-0.47	1.62	1.16	-0.31	0.34
<i>CT_HRF</i>	0.61	FR	ADB	1.41	1.02	0.01	1.47	1.08	-0.08	0.34
<i>CT_HRF</i>	0.74	RR	DTC	1.47	1.10	-0.08	1.55	1.20	-0.20	0.37
<i>CT_HRF</i>	0.61	FR	MLP	1.50	1.11	-0.12	1.42	1.06	-0.01	0.37
<i>CT_HRF</i>	0.61	FR	RFR	1.48	1.08	-0.11	1.54	1.13	-0.19	0.43
<i>CT_HRF</i>	0.61	FR	KNN	1.50	1.18	-0.12	1.50	1.14	-0.13	0.60
<i>CT_HRF</i>	0.74	RR	MLP	1.51	1.12	-0.14	1.46	1.15	-0.07	0.67
<i>CT_HRF</i>	0.74	RR	RFR	1.54	1.21	-0.20	1.58	1.19	-0.25	0.77
<i>CT_HRF</i>	0.61	FR	SVR	1.45	1.07	-0.06	1.46	1.07	-0.07	0.89
<i>CT_HRF+CF</i>	0.74	RR	LRR	1.46	1.15	-0.06	1.86	1.50	-0.74	0.00
<i>CT_HRF+CF</i>	0.64	FR	MLP	1.48	1.17	-0.11	2.27	1.76	-1.59	0.01
<i>CT_HRF+CF</i>	0.74	RR	SVR	1.52	1.11	-0.16	1.77	1.41	-0.57	0.02
<i>CT_HRF+CF</i>	0.74	RR	KNN	1.41	1.09	0.00	1.61	1.25	-0.30	0.09
<i>CT_HRF+CF</i>	0.64	FR	KNN	1.44	1.11	-0.05	1.39	1.02	0.04	0.09
<i>CT_HRF+CF</i>	0.74	RR	RFR	1.56	1.22	-0.22	1.81	1.40	-0.64	0.19
<i>CT_HRF+CF</i>	0.64	FR	ADB	1.55	1.17	-0.21	1.49	1.08	-0.11	0.20

<i>CT_HRF+CF</i>	0.64	FR	SVR	1.43	1.06	-0.03	1.42	1.05	-0.01	0.24
<i>CT_HRF+CF</i>	0.74	RR	MLP	1.45	1.12	-0.05	1.49	1.20	-0.11	0.29
<i>CT_HRF+CF</i>	0.64	FR	DTC	1.57	1.15	-0.23	1.46	1.08	-0.07	0.45
<i>CT_HRF+CF</i>	0.64	FR	LRR	1.50	1.10	-0.13	1.48	1.08	-0.09	0.68
<i>CT_HRF+CF</i>	0.64	FR	RFR	1.46	1.08	-0.06	1.49	1.07	-0.12	0.92
<i>CT_HRF+CF</i>	0.74	RR	DTC	1.69	1.34	-0.44	2.00	1.36	-1.01	0.92
<i>CT_HRF+CF</i>	0.59	RR	ADB	1.47	1.10	-0.08	1.46	1.10	-0.08	0.96
<i>PET_DRF</i>	0.66	FR	RFR	1.32	1.02	0.12	1.73	1.23	-0.50	0.04
<i>PET_DRF</i>	0.66	FR	SVR	1.43	1.07	-0.02	1.49	1.13	-0.11	0.14
<i>PET_DRF</i>	0.50	RR	RFR	1.41	1.09	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF</i>	0.50	RR	LRR	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF</i>	0.50	RR	ADB	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF</i>	0.66	FR	KNN	1.46	1.12	-0.07	1.62	1.20	-0.32	0.30
<i>PET_DRF</i>	0.66	FR	ADB	1.33	1.01	0.11	1.71	1.14	-0.47	0.31
<i>PET_DRF</i>	0.50	RR	DTC	1.41	1.08	0.00	1.42	1.06	-0.01	0.33
<i>PET_DRF</i>	0.66	FR	DTC	1.59	1.13	-0.27	1.96	1.31	-0.93	0.34
<i>PET_DRF</i>	0.50	RR	KNN	1.41	1.08	0.00	1.42	1.05	-0.02	0.38
<i>PET_DRF</i>	0.66	FR	MLP	1.45	1.14	-0.06	1.41	1.08	0.01	0.54
<i>PET_DRF</i>	0.66	FR	LRR	1.49	1.09	-0.11	1.59	1.14	-0.27	0.54

<i>PET_DRF</i>	0.50	RR	MLP	1.41	1.09	0.00	1.48	1.10	-0.09	0.74
<i>PET_DRF</i>	0.50	RR	SVR	1.42	1.05	-0.02	1.42	1.05	-0.02	
<i>PET_DRF+CF</i>	0.67	FR	RFR	1.32	1.02	0.13	1.59	1.17	-0.27	0.08
<i>PET_DRF+CF</i>	0.67	FR	SVR	1.43	1.07	-0.02	1.49	1.12	-0.11	0.22
<i>PET_DRF+CF</i>	0.50	RR	ADB	1.42	1.10	-0.01	1.41	1.06	0.00	0.26
<i>PET_DRF+CF</i>	0.50	RR	RFR	1.41	1.09	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF+CF</i>	0.50	RR	LRR	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF+CF</i>	0.67	FR	DTC	1.57	1.11	-0.24	1.83	1.28	-0.68	0.31
<i>PET_DRF+CF</i>	0.50	RR	DTC	1.41	1.08	0.00	1.42	1.06	-0.01	0.33
<i>PET_DRF+CF</i>	0.67	FR	MLP	1.48	1.20	-0.11	1.40	1.10	0.02	0.36
<i>PET_DRF+CF</i>	0.50	RR	KNN	1.41	1.08	0.00	1.42	1.05	-0.02	0.38
<i>PET_DRF+CF</i>	0.67	FR	ADB	1.36	1.03	0.07	1.56	1.12	-0.21	0.49
<i>PET_DRF+CF</i>	0.67	FR	LRR	1.48	1.09	-0.10	1.57	1.12	-0.24	0.50
<i>PET_DRF+CF</i>	0.50	RR	MLP	1.41	1.09	0.00	1.49	1.10	-0.12	0.63
<i>PET_DRF+CF</i>	0.67	FR	KNN	1.46	1.12	-0.06	1.50	1.13	-0.13	0.81
<i>PET_DRF+CF</i>	0.50	RR	SVR	1.42	1.05	-0.02	1.42	1.05	-0.02	
<i>PET_DRF+CF+PET_HRF</i>	0.66	FR	MLP	1.49	1.21	-0.12	2.98	2.55	-3.45	0.00
<i>PET_DRF+CF+PET_HRF</i>	0.66	FR	ADB	1.51	1.12	-0.14	1.90	1.36	-0.82	0.07
<i>PET_DRF+CF+PET_HRF</i>	0.66	FR	RFR	1.43	1.12	-0.03	1.70	1.26	-0.45	0.17

<i>PET_DRF+CF+PET_HRF</i>	0.50	RR	RFR	1.41	1.09	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF+CF+PET_HRF</i>	0.50	RR	LRR	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF+CF+PET_HRF</i>	0.50	RR	ADB	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF+CF+PET_HRF</i>	0.66	FR	DTC	1.83	1.35	-0.67	2.26	1.60	-1.57	0.29
<i>PET_DRF+CF+PET_HRF</i>	0.50	RR	DTC	1.41	1.08	0.00	1.42	1.06	-0.01	0.33
<i>PET_DRF+CF+PET_HRF</i>	0.50	RR	KNN	1.41	1.08	0.00	1.42	1.05	-0.02	0.38
<i>PET_DRF+CF+PET_HRF</i>	0.66	FR	LRR	1.50	1.16	-0.13	1.65	1.20	-0.36	0.60
<i>PET_DRF+CF+PET_HRF</i>	0.50	RR	MLP	1.41	1.09	0.00	1.48	1.10	-0.11	0.69
<i>PET_DRF+CF+PET_HRF</i>	0.66	FR	KNN	1.52	1.18	-0.17	1.53	1.21	-0.17	0.75
<i>PET_DRF+CF+PET_HRF</i>	0.66	FR	SVR	1.42	1.07	-0.01	1.40	1.07	0.01	0.87
<i>PET_DRF+CF+PET_HRF</i>	0.50	RR	SVR	1.42	1.05	-0.02	1.42	1.05	-0.02	
<i>PET_DRF+PET_HRF</i>	0.50	RR	RFR	1.41	1.09	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF+PET_HRF</i>	0.50	RR	LRR	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF+PET_HRF</i>	0.50	RR	ADB	1.41	1.08	0.00	1.41	1.07	0.00	0.26
<i>PET_DRF+PET_HRF</i>	0.50	RR	DTC	1.41	1.08	0.00	1.42	1.06	-0.01	0.33
<i>PET_DRF+PET_HRF</i>	0.50	RR	KNN	1.41	1.08	0.00	1.42	1.05	-0.02	0.38
<i>PET_DRF+PET_HRF</i>	0.66	FR	MLP	1.42	1.14	-0.02	1.44	1.18	-0.03	0.58
<i>PET_DRF+PET_HRF</i>	0.50	RR	MLP	1.41	1.09	0.00	1.48	1.10	-0.10	0.67
<i>PET_DRF+PET_HRF</i>	0.66	FR	LRR	1.51	1.15	-0.14	1.52	1.17	-0.17	0.71

<i>PET_DRF+PET_HRF</i>	0.66	FR	ADB	1.46	1.13	-0.07	1.46	1.15	-0.07	0.75
<i>PET_DRF+PET_HRF</i>	0.66	FR	SVR	1.42	1.09	-0.01	1.41	1.08	0.00	0.79
<i>PET_DRF+PET_HRF</i>	0.66	FR	RFR	1.45	1.17	-0.06	1.53	1.18	-0.17	0.91
<i>PET_DRF+PET_HRF</i>	0.66	FR	KNN	1.47	1.14	-0.08	1.49	1.14	-0.11	0.95
<i>PET_DRF+PET_HRF</i>	0.66	FR	DTC	1.74	1.36	-0.53	2.69	1.37	-2.62	0.98
<i>PET_DRF+PET_HRF</i>	0.50	RR	SVR	1.42	1.05	-0.02	1.42	1.05	-0.02	
<i>PET_HRF</i>	0.63	FR	MLP	2.17	1.79	-1.36	1.42	1.10	-0.02	0.00
<i>PET_HRF</i>	0.63	FR	ADB	1.57	1.20	-0.23	1.38	1.05	0.04	0.00
<i>PET_HRF</i>	0.68	RR	SVR	137.64	58.33	-9507.08	1.37	1.04	0.06	0.00
<i>PET_HRF</i>	0.63	FR	RFR	1.59	1.25	-0.27	1.47	1.13	-0.09	0.00
<i>PET_HRF</i>	0.63	FR	DTC	1.80	1.39	-0.63	1.51	1.11	-0.14	0.03
<i>PET_HRF</i>	0.68	RR	KNN	1.53	1.20	-0.18	1.43	1.07	-0.03	0.11
<i>PET_HRF</i>	0.63	FR	SVR	1.42	1.07	-0.01	1.39	1.05	0.03	0.22
<i>PET_HRF</i>	0.63	FR	LRR	1.48	1.13	-0.09	1.50	1.17	-0.12	0.43
<i>PET_HRF</i>	0.63	FR	KNN	1.52	1.16	-0.15	1.44	1.12	-0.04	0.52
<i>PET_HRF</i>	0.68	RR	RFR	1.58	1.23	-0.25	1.54	1.20	-0.18	0.66
<i>PET_HRF</i>	0.68	RR	DTC	1.70	1.29	-0.45	1.77	1.34	-0.57	0.70
<i>PET_HRF</i>	0.63	RR	ADB	1.52	1.14	-0.16	1.46	1.12	-0.08	0.71
<i>PET_HRF</i>	0.68	RR	LRR	1.50	1.14	-0.13	1.49	1.15	-0.12	0.75

<i>PET_HRF</i>	0.68	RR	MLP	1.45	1.06	-0.06	1.37	1.08	0.06	0.85
<i>PET_HRF+CF</i>	0.67	RR	SVR	100.18	42.79	-5035.65	1.48	1.16	-0.10	0.00
<i>PET_HRF+CF</i>	0.63	FR	DTC	1.80	1.39	-0.63	1.45	1.08	-0.06	0.01
<i>PET_HRF+CF</i>	0.63	FR	RFR	1.58	1.23	-0.25	1.49	1.14	-0.11	0.04
<i>PET_HRF+CF</i>	0.67	RR	DTC	1.67	1.25	-0.40	2.07	1.57	-1.15	0.08
<i>PET_HRF+CF</i>	0.67	RR	MLP	1.42	1.11	-0.02	1.47	1.18	-0.09	0.24
<i>PET_HRF+CF</i>	0.63	RR	ADB	1.51	1.14	-0.15	1.45	1.11	-0.06	0.39
<i>PET_HRF+CF</i>	0.63	FR	KNN	1.51	1.14	-0.14	1.44	1.09	-0.04	0.40
<i>PET_HRF+CF</i>	0.63	FR	LRR	1.50	1.14	-0.13	1.43	1.10	-0.02	0.56
<i>PET_HRF+CF</i>	0.63	FR	SVR	1.42	1.06	-0.01	1.41	1.05	0.00	0.63
<i>PET_HRF+CF</i>	0.63	FR	MLP	1.41	1.05	0.00	1.43	1.07	-0.02	0.65
<i>PET_HRF+CF</i>	0.67	RR	KNN	1.53	1.20	-0.18	1.50	1.17	-0.12	0.71
<i>PET_HRF+CF</i>	0.63	FR	ADB	1.48	1.09	-0.10	1.46	1.10	-0.07	0.89
<i>PET_HRF+CF</i>	0.67	RR	RFR	1.56	1.21	-0.22	1.56	1.21	-0.23	0.96
<i>PET_HRF+CF</i>	0.67	RR	LRR	1.49	1.15	-0.12	1.52	1.15	-0.16	0.96

HRF: handcrafted Radiomics features, DRF: deep Radiomics features, DRF_CT: DRFs extracted from segmented CT, DRF_PET: DRFs extracted from segmented PET, HRF_CT: HFRs extracted from segmented CT, HRF_PET: HFRs extracted from segmented PET, FFCV: Five-Fold Cross-Validation, SSL: Semi-supervised Learning, SL: Supervised Learning, MAE: Mean Absolute Error, STD: Standard Deviation, ABR: AdaBoost Regressor, RFR: Random Forest Regressor, MLPR: Multi-Layer Perceptron Regressor, DTR: Decision Tree Regressor, LR: Linear Regression, SVR: Support Vector Machine Regressor, KNNR: K-Nearest Neighbor Regressor.

Supplemental Table S3. Top 10 features and parameters of top results

<i>Dataset</i>	<i>FSA</i>	<i>Algorit</i> <i>hm</i>	<i>SL_Top10_Features</i>	<i>SL_BestP</i> <i>arams</i>	<i>SSL_Top10_Features</i>	<i>SSL_Best</i> <i>Params</i>
<i>PET_HR</i> <i>F</i>	FR	ADB	cm_joint_max_3D_comb (0.0890), szm_zsnu_norm_3D (0.0767), szm_szhge_3D (0.0078), cm_joint_max_3D_avg (0.0007), cm_energy_3D_comb (0.0000), szm_size_3D (0.0000), cm_energy_3D_avg (0.0000), cm_joint_entr_3D_avg (0.0000), morph_geary_c (0.0000), morph_moran_i (0.0000)	{'learning _rate': 0.1, 'loss': 'linear', 'n_estima tors': 10}	szm_szhge_3D (0.2486), cm_joint_max_3D_avg (0.0885), cm_energy_3D_comb (0.0670), cm_joint_max_3D_comb (0.0623), szm_zsnu_norm_3D (0.0515), szm_size_3D (0.0502), morph_moran_i (0.0268), morph_geary_c (0.0157), cm_energy_3D_avg (0.0039), cm_joint_entr_3D_avg (0.0000)	{'learning _rate': 0.01, 'loss': 'square', 'n_estimat ors': 10}
<i>CF</i>	RR	ADB	ageatdiagnosis (0.0761), FBase (0.0236), Wt_loss_c (0.0140), Correct_baseline_ECOG (0.0114), Fractionation_cat (0.0077), Met_no0_yes1 (0.0000), Sex (0.0000), Recurrence_status (0.0000),	{'learning _rate': 0.1, 'loss': 'square', 'n_estima tors': 10}	Wt_loss_c (0.1122), ageatdiagnosis (0.1086), Correct_baseline_ECOG (0.0849), Sex (0.0697), Recurrence_status (0.0402), Met_no0_yes1 (0.0381), FBase (0.0198), Fractionation_cat (0.0156),	{'learning _rate': 0.1, 'loss': 'square', 'n_estimat ors': 10}

			chemotherapy (0.0000), PostOpResidualTumourSx1 (0.0000)		chemotherapy (0.0125), PostOpResidualTumourSx1 (0.0034)	
<i>CT_DRF</i>	FR	ADB	459 (0.6831), 505 (0.2896), 298 (0.0926), 114 (0.0919), 456 (0.0292), 291 (0.0287), 480 (0.0282), 427 (0.0098), 420 (0.0038), 461 (0.0000)	{'learning _rate': 0.01, 'loss': 'linear', 'n_estima tors': 10}	298 (0.2403), 114 (0.1455), 480 (0.1076), 427 (0.0481), 461 (0.0361), 459 (0.0192), 291 (0.0107), 420 (0.0080), 505 (0.0000), 456 (0.0000)	{'learning _rate': 0.01, 'loss': 'linear', 'n_estimat ors': 10}
<i>PET_DR F+CF</i>	FR	RFR	DF327 (0.6999), DF330 (0.6497), DF2 (0.1416), DF816 (0.1125), DF586 (0.0967), DF98 (0.0655), DF58 (0.0483), DF625 (0.0370), DF212 (0.0204), DF863 (0.0106)	{'max_de pth': 5, 'n_estima tors': 100}	DF2 (0.2547), DF863 (0.2515), DF816 (0.1715), DF58 (0.1071), DF98 (0.0538), DF330 (0.0140), DF327 (0.0086), DF212 (0.0078), DF625 (0.0058), DF586 (0.0037)	{'max_de pth': 5, 'n_estimat ors': 50}
<i>CT+CF</i>	FR	KNN	Smoking Pack Years (0.3584), Smoking_cat (0.0179), dzm_sdlge_3D (0.0010), szm_lgze_3D (0.0004), morph_av (0.0000), ngl_lde_3D (0.0000), szm_szlge_3D (0.0000), dzm_lgze_3D	{'algorith m': 'auto', 'n_neighb ors': 7,	Smoking Pack Years (0.2774), Smoking_cat (0.1100), ngl_lde_3D (0.0238), morph_av (0.0237), dzm_sdlge_3D (0.0054), szm_lgze_3D (0.0049), dzm_lgze_3D (0.0049),	{'algorith m': 'auto', 'n_neighb ors': 9,

			(0.0000), ngl_dcnu_norm_3D (0.0000), ngl_ldlge_3D (0.0000)	'weights': 'uniform'}	ngl_dcnu_norm_3D (0.0046), szm_szlge_3D (0.0029), ngl_ldlge_3D (0.0005)	'weights': 'uniform'}
<i>PET_DR</i> <i>F</i>	FR	ADB	DF241 (0.1918), DF277 (0.1293), DF98 (0.1013), DF58 (0.0814), DF863 (0.0705), DF2 (0.0319), DF625 (0.0000), DF330 (0.0000), DF586 (0.0000), DF212 (0.0000)	{'learning _rate': 0.1, 'loss': 'linear', 'n_estima tors': 10}	DF2 (0.1790), DF863 (0.1503), DF277 (0.1022), DF58 (0.0865), DF241 (0.0693), DF98 (0.0170), DF212 (0.0069), DF625 (0.0000), DF586 (0.0000), DF330 (0.0000)	{'learning _rate': 0.01, 'loss': 'linear', 'n_estimat ors': 10}
<i>CT</i>	FR	ADB	dzm_sdlge_3D (0.3624), morph_av (0.1769), cm_info_corr1_3D_avg (0.0956), stat_qcod (0.0897), dzm_lgze_3D (0.0175), ngl_hdhge_3D (0.0000), ngl_ldlge_3D (0.0000), szm_szlge_3D (0.0000), szm_lgze_3D (0.0000), cm_info_corr2_3D_avg (0.0000)	{'learning _rate': 0.01, 'loss': 'square', 'n_estima tors': 10}	szm_szlge_3D (0.1629), dzm_sdlge_3D (0.0894), ngl_ldlge_3D (0.0566), cm_info_corr1_3D_avg (0.0533), stat_qcod (0.0487), szm_lgze_3D (0.0201), morph_av (0.0199), cm_info_corr2_3D_avg (0.0051), ngl_hdhge_3D (0.0019), dzm_lgze_3D (0.0006)	{'learning _rate': 0.01, 'loss': 'square', 'n_estimat ors': 10}

<i>PET_DR</i> <i>F+CF+P</i> <i>ET_HRF</i>	RR	KNN	DF1020 (0.0000), DF1012 (0.0000), DF1006 (0.0000), DF1001 (0.0000), DF991 (0.0000), DF988 (0.0000), DF977 (0.0000), DF965 (0.0000), DF948 (0.0000), DF928 (0.0000)	{'algorithm': 'ball_tree', , 'n_neighbors': 3, 'weights': 'uniform'}	DF1020 (0.0000), DF1012 (0.0000), DF1006 (0.0000), DF1001 (0.0000), DF991 (0.0000), DF988 (0.0000), DF977 (0.0000), DF965 (0.0000), DF948 (0.0000), DF928 (0.0000)	{'algorithm': 'auto', 'n_neighbors': 3, 'weights': 'uniform'}
<i>PET_DR</i> <i>F+PET_</i> <i>HRF</i>	RR	KNN	DF1020 (0.0000), DF1012 (0.0000), DF1006 (0.0000), DF1001 (0.0000), DF991 (0.0000), DF988 (0.0000), DF977 (0.0000), DF965 (0.0000), DF948 (0.0000), DF928 (0.0000)	{'algorithm': 'ball_tree', , 'n_neighbors': 3, 'weights': 'uniform'}	DF1020 (0.0000), DF1012 (0.0000), DF1006 (0.0000), DF1001 (0.0000), DF991 (0.0000), DF988 (0.0000), DF977 (0.0000), DF965 (0.0000), DF948 (0.0000), DF928 (0.0000)	{'algorithm': 'auto', 'n_neighbors': 3, 'weights': 'uniform'}
<i>PET_HR</i> <i>F+CF</i>	FR	SVR	szm_szhge_3D (0.1071), dzm_sdhge_3D (0.0399), Smoking_cat (0.0064), ngl_glnu_norm_3D (0.0000),	{'C': 1, 'degree': 3,	szm_szhge_3D (0.1343), dzm_sdhge_3D (0.1128), Smoking_cat (0.0064), morph_geary_c (0.0000), morph_moran_i	{'C': 1, 'degree':

			cm_joint_max_3D_comb (0.0000), morph_geary_c (0.0000), cm_joint_max_3D_avg (0.0000), cm_energy_3D_comb (0.0000), cm_energy_3D_avg (0.0000), morph_moran_i (0.0000)	'kernel': 'rbf'	(0.0000), cm_joint_max_3D_comb (0.0000), cm_joint_max_3D_avg (0.0000), ngl_glnu_norm_3D (0.0000), cm_energy_3D_comb (0.0000), cm_energy_3D_avg (0.0000)	3, 'kernel': 'rbf'
<i>CT_DRF</i> <i>+CF</i>	FR	ADB	DF459 (0.2402), DF505 (0.1674), DF114 (0.1316), DF456 (0.1215), DF480 (0.0518), DF291 (0.0206), DF427 (0.0192), DF420 (0.0190), DF461 (0.0010), DF298 (0.0009)	{'learning _rate': 0.1, 'loss': 'linear', 'n_estima tors': 10}	DF298 (0.2844), DF427 (0.1225), DF114 (0.0941), DF480 (0.0779), DF505 (0.0255), DF459 (0.0247), DF456 (0.0002), DF461 (0.0000), DF420 (0.0000), DF291 (0.0000)	{'learning _rate': 0.1, 'loss': 'linear', 'n_estimat ors': 10}
<i>CT_DRF</i> <i>+CF+CT</i> <i>_HRF</i>	FR	ADB	DF459 (0.1361), DF456 (0.1087), DF480 (0.0974), DF291 (0.0683), DF461 (0.0362), DF298 (0.0131), DF114 (0.0097), DF427 (0.0083), DF420 (0.0074), DF505 (0.0000)	{'learning _rate': 0.01, 'loss': 'square', 'n_estima tors': 10}	DF298 (0.2480), DF427 (0.1489), DF114 (0.0636), DF459 (0.0429), DF480 (0.0391), DF420 (0.0183), DF461 (0.0161), DF291 (0.0157), DF456 (0.0144), DF505 (0.0090)	{'learning _rate': 0.01, 'loss': 'linear', 'n_estimat ors': 10}

<i>CT_DRF</i>	RR	SVR	DF1012 (0.0000), DF884 (0.0000),	{'C': 0.01,	DF505 (0.1062), DF459 (0.0769), DF179	{'C': 1,
<i>+CT_HR</i>			DF756 (0.0000), DF684 (0.0000), DF663	'degree':	(0.0707), DF684 (0.0631), DF510	'degree':
<i>F</i>			(0.0000), DF628 (0.0000), DF535	2,	(0.0596), DF502 (0.0448), DF480	2, 'kernel':
			(0.0000), DF510 (0.0000), DF507	'kernel':	(0.0446), DF473 (0.0445), DF246	'linear'}
			(0.0000), DF505 (0.0000)	'linear'}	(0.0233), DF462 (0.0195)	

HRF: handcrafted Radiomics features, DRF: deep Radiomics features, DRF_CT: DRFs extracted from segmented CT, DRF_PET: DRFs extracted from segmented PET, HRF_CT: HFRs extracted from segmented CT, HRF_PET: HFRs extracted from segmented PET, FFCV: Five-Fold Cross-Validation, SSL: Semi-supervised Learning, SL: Supervised Learning, MAE: Mean Absolute Error, STD: Standard Deviation, ABR: AdaBoost Regressor, RFR: Random Forest Regressor, SVR: Support Vector Machine Regressor, KNNR: K-Nearest Neighbor Regressor, FSA: Feature selection Algorithm, RR: r_regression, FR: f_regression.

Supplemental Table S4. A list of the most important radiomics features used in this study, detailing their abbreviated names, full descriptions, and corresponding categories

<i>Abbreviation</i>	<i>Complete Form</i>	<i>Category</i>	<i>Description</i>
<i>ngt_strength_3D</i>	Neighbouring Grey Tone Difference Matrix (NGTDM) Strength	Texture	Measures the primitiveness of the texture. High strength means high contrast.
<i>ngt_coarseness_3D</i>	NGTDM Coarseness	Texture	Measures the spatial rate of change in intensity. Lower values mean "finer" texture.

<i>dzm_sdhge_3D</i>	Dependent Zone Matrix (DZM) Small Dependence High Grey Level Emphasis	Texture	Measures the joint distribution of small dependence zones with high grey levels.
<i>dzm_zdnu_norm_3D</i>	DZM Zone Distance Non- Uniformity Normalized	Texture	Measures the variability of zone distances in the image, normalized.
<i>dzm_hgze_3D</i>	DZM High Grey Level Zone Emphasis	Texture	Measures the distribution of high grey-level values.
<i>szm_szhge_3D</i>	Size Zone Matrix (SZM) Small Zone High Grey Level Emphasis	Texture	Measures the joint distribution of small size zones with high grey levels.
<i>szm_zsnu_norm_3D</i>	SZM Zone Size Non-Uniformity Normalized	Texture	Measures the variability of zone sizes in the image, normalized.
<i>szm_size_3D</i>	SZM Small Zone Emphasis	Texture	Measures the distribution of small size zones.
<i>cm_joint_max_3D_avg</i>	Co-occurrence Matrix (GLCM) Joint Maximum	Texture	Measures the most predominant pair of neighbouring intensity values.
<i>cm_joint ENTR_3D_avg</i>	Co-occurrence Matrix (GLCM) Joint Entropy	Texture	Measures the randomness/complexity of the intensity distribution.
<i>cm_energy_3D_avg</i>	Co-occurrence Matrix (GLCM) Energy	Texture	Measures the textural uniformity (also called Angular Second Moment).

<i>ivh_diff_v10_v90</i>	Intensity-Volume Histogram (IVH) Difference between V10 and V90	Intensity	The range of intensity values containing the central 80% of the tumor volume.
<i>ivh_v10</i>	IVH Intensity at 10th Volume Percentile	Intensity	The intensity level below which 10% of the tumor volume exists.
<i>morph_geary_c</i>	Morphological Geary's C	Shape	Measures spatial autocorrelation; how similar a voxel is to its neighbors.
<i>morph_av</i>	Morphological Feature (Average)	Shape	This is not a standard feature name. 'morph' refers to shape, and 'av' likely means an average of some property, but the specific feature is unclear.

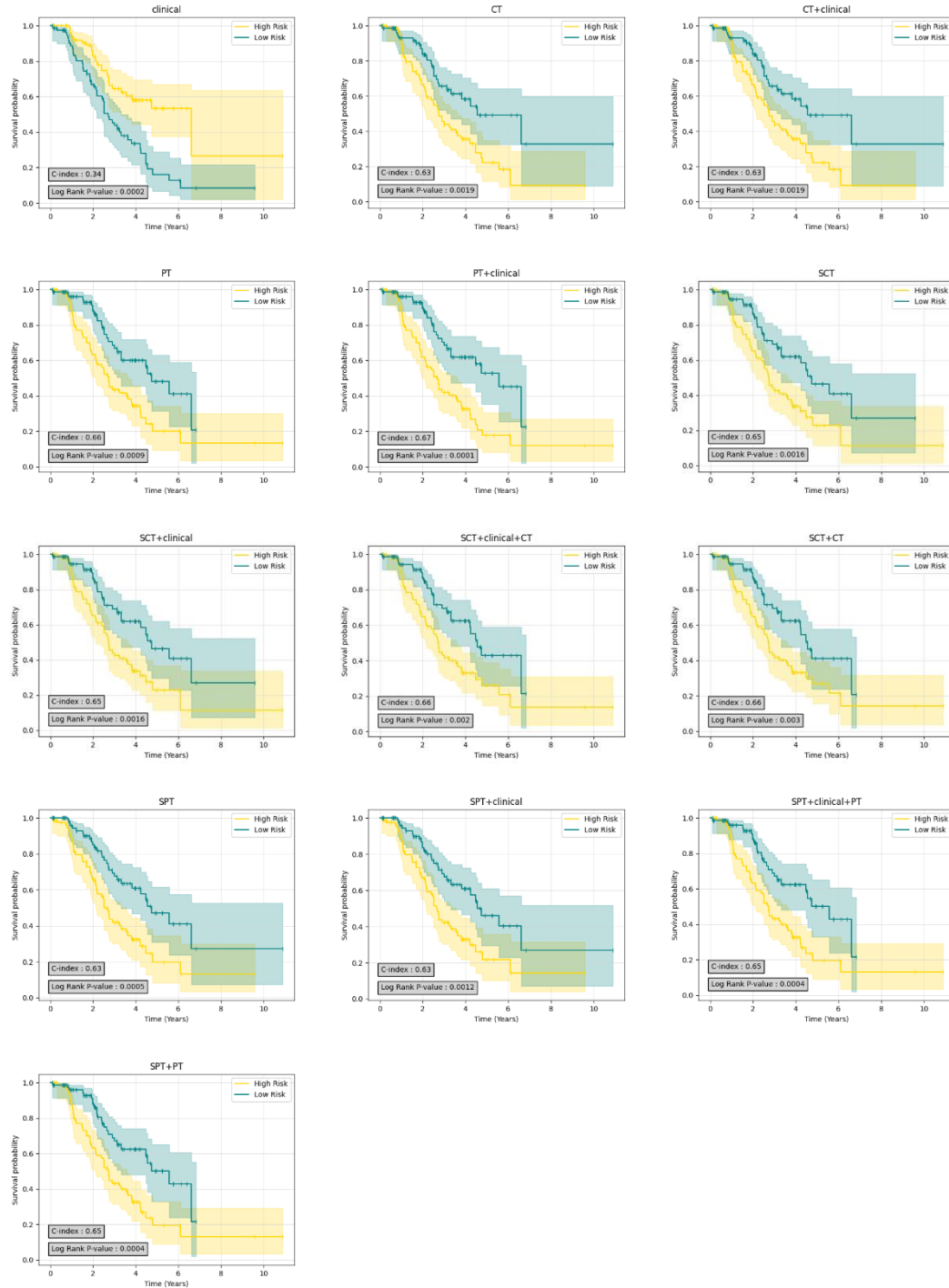
Supplemental Table S5. Hazard ratio survival analysis (HRSA) results

<i>Dataset</i>	<i>PCA+ HRSA</i>	<i>FFCV c-index ± STD</i>	<i>FFCV p- value</i>	<i>External Test c-index</i>	<i>External Test p-value</i>
<i>CF</i>	CWGBSA	0.34±0.04	0	0.564	0.822
<i>CT_HRF</i>	CWGBSA	0.63±0.08	0.002	0.656	0.019
<i>CT_HRF+CF</i>	CWGBSA	0.63±0.08	0.002	0.653	0.118
<i>PET_HRF</i>	CWGBSA	0.66±0.06	0.001	0.549	0.239
<i>PET_HRF+CF</i>	CWGBSA	0.67±0.07	0	0.553	0.052
<i>S_CT_DRF</i>	CWGBSA	0.65±0.07	0.002	0.547	0.021

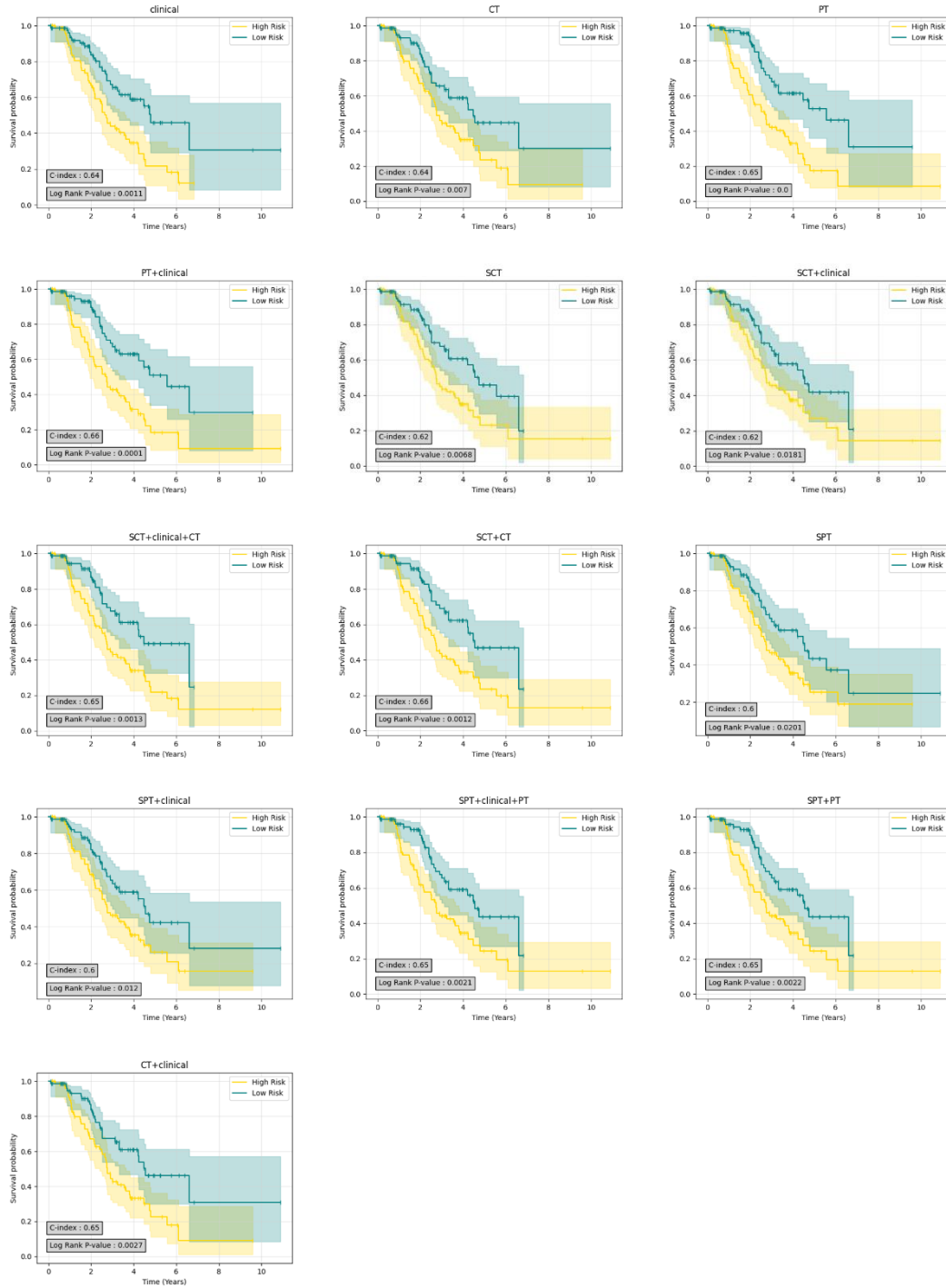
<i>S_CT_DRF+CF</i>	CWGBSA	0.65±0.06	0.002	0.551	0.085
<i>S_CT_DRF+CT_HRF</i>	CWGBSA	0.66±0.06	0.003	0.598	0.632
<i>S_CT_DRF+CT_HRF+CF</i>	CWGBSA	0.66±0.06	0.002	0.598	0.632
<i>S_PET_DRF</i>	CWGBSA	0.63±0.06	0.001	0.46	0.611
<i>S_PET_DRF+CF</i>	CWGBSA	0.63±0.06	0.001	0.471	0.452
<i>S_PET_DRF+PET_HRF</i>	CWGBSA	0.65±0.06	0	0.538	0.921
<i>S_PET_DRF+PET_HRF+CF</i>	CWGBSA	0.65±0.06	0	0.533	0.755
<i>CF</i>	FSSVM	0.64±0.02	0.001	0.633	0.449
<i>CT_HRF</i>	FSSVM	0.64±0.08	0.007	0.582	0.182
<i>CT_HRF+CF</i>	FSSVM	0.65±0.08	0.003	0.598	0.141
<i>PET_HRF</i>	FSSVM	0.65±0.07	0	0.56	0.209
<i>PET_HRF+CF</i>	FSSVM	0.66±0.07	0	0.598	0.06
<i>S_CT_DRF</i>	FSSVM	0.62±0.04	0.007	0.529	0.178
<i>S_CT_DRF+CF</i>	FSSVM	0.62±0.04	0.018	0.527	0.178
<i>S_CT_DRF+CT_HRF</i>	FSSVM	0.66±0.05	0.001	0.578	0.238
<i>S_CT_DRF+CT_HRF+CF</i>	FSSVM	0.65±0.05	0.001	0.582	0.166
<i>S_PET_DRF</i>	FSSVM	0.60±0.06	0.02	0.482	0.021
<i>S_PET_DRF+CF</i>	FSSVM	0.6±0.05	0.012	0.48	0.01
<i>S_PET_DRF+PET_HRF</i>	FSSVM	0.65±0.07	0.002	0.553	0.115

<i>S_PET_DRF+PET_HRF+CF</i>	FSSVM	0.65±0.06	0.002	0.553	0.115
<i>CF</i>	RSF	0.71±0.05	0	0.613	0.263
<i>CT_HRF</i>	RSF	0.62±0.05	0.052	0.521	0.742
<i>CT_HRF+CF</i>	RSF	0.64±0.12	0.001	0.523	0.886
<i>PET_HRF</i>	RSF	0.65±0.06	0.002	0.573	0.628
<i>PET_HRF+CF</i>	RSF	0.65±0.06	0	0.569	0.107
<i>S_CT_DRF</i>	RSF	0.61±0.05	0.014	0.553	0.075
<i>S_CT_DRF+CF</i>	RSF	0.61±0.02	0.025	0.556	0.027
<i>S_CT_DRF+CT_HRF</i>	RSF	0.66±0.05	0	0.632	0.585
<i>S_CT_DRF+CT_HRF+CF</i>	RSF	0.62±0.05	0.004	0.598	0.335
<i>S_PET_DRF</i>	RSF	0.59±0.03	0.229	0.509	0.068
<i>S_PET_DRF+CF</i>	RSF	0.58±0.05	0.021	0.513	0.371
<i>S_PET_DRF+PET_HRF</i>	RSF	0.63±0.06	0.02	0.587	0.207
<i>S_PET_DRF+PET_HRF+CF</i>	RSF	0.62±0.09	0.173	0.56	0.004

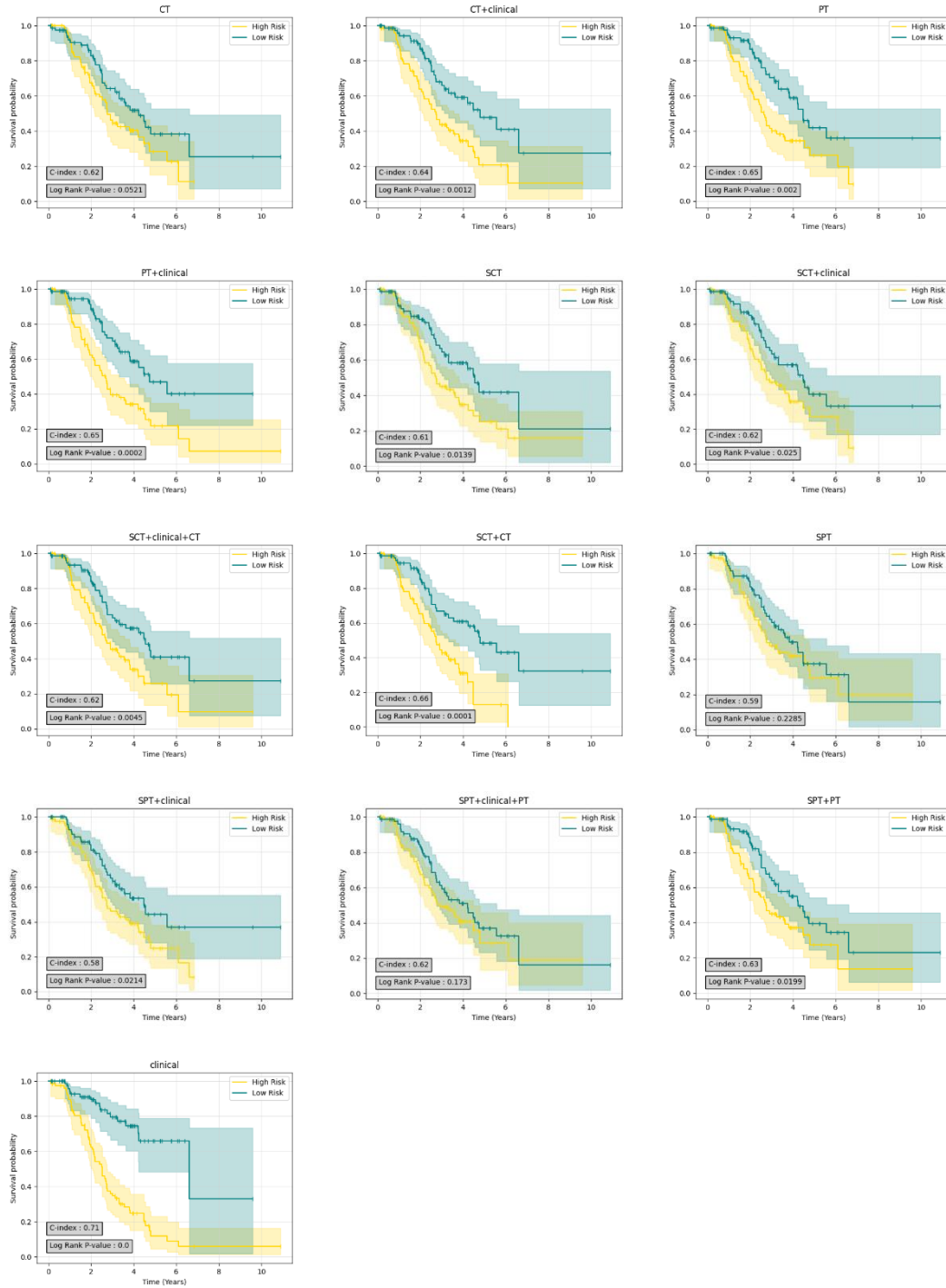
HRF: handcrafted Radiomics features, DRF: deep Radiomics features, DRF_CT: DRFs extracted from segmented CT, DRF_PET: DRFs extracted from segmented PET, HRF_CT: HRFs extracted from segmented CT, HRF_PET: HRFs extracted from segmented PET, FSSVM: Fast Survival Support Vector Machines, CWGBSA: Component-wise Gradient Boosting Survival Analysis, RSF: Random Survival Forest, FFCV: Five-Fold Cross-Validation, HRSA: Hazard Ratio Survival Analysis. PCA: Principal Component Analysis.



Supplemental Figure S2. Kaplan-Meier survival curves generated for Component-wise Gradient Boosting Survival Analysis (CWGBSA) Results; SCT: Deep Radiomics Features extracted from CT, CT: Handcrafted Radiomics Features Extracted From CT, SPET: Deep Radiomics Features extracted from PET, PET: Handcrafted Radiomics Features Extracted From PET



Supplemental Figure S3. Kaplan-Meier survival curves generated for Fast Survival Support Vector Machines (FSSVM) Results; SCT: Deep Radiomics Features extracted from CT, CT: Handcrafted Radiomics Features Extracted From CT, SPET: Deep Radiomics Features extracted from PET, PET: Handcrafted Radiomics Features Extracted From PET



Supplemental Figure S4. Kaplan-Meier survival curves generated for Random Survival Forest (RSF) Results; SCT: Deep Radiomics Features extracted from CT, CT: Handcrafted Radiomics Features Extracted From CT, SPET: Deep Radiomics Features extracted from PET, PET: Handcrafted Radiomics Features Extracted From PET

References

- [1] D. I. Garcia and C. Catana, "MR Imaging-Guided Attenuation Correction of PET Data in PET/MR Imaging," *PET Clin*, vol. 11, no. 2, pp. 129-49, 2016.
- [2] M. C. Adams, T. G. Turkington, J. M. Wilson and T. Z. Wong, "A systematic review of the factors affecting accuracy of SUV measurements.," *American Journal of Roentgenology*, vol. 195, no. 2, pp. 310-320, 2010.
- [3] J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, San Francisco, CA, USA: Elsevier, 2011.
- [4] A. Jain, R. P. W. Duin and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4-37, 2000.
- [5] M. R. Salmanpour, I. Shiri and et al, "ViSERA: Visualized & Standardized Environment for Radiomics Analysis - A Shareable, Executable, and Reproducible Workflow Generator," in *2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD)*, Vancouver, 2023.
- [6] The scikit-learn developers, "sklearn.feature_selection.r_regression," scikit-learn, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.r_regression.html. [Accessed 22 August 2024].
- [7] The scikit-learn developers, "sklearn.feature_selection.f_regression," scikit-learn, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_regression.html. [Accessed 22 August 2024].
- [8] M. Salmanpour, M. Bakhtiyari and et al, "Application of novel hybrid machine learning systems and radiomics features for non-motor outcome prediction in Parkinson's disease," *Physics in Medicine & Biology*, vol. 68, no. 3, p. 035004, 2023.
- [9] M. J. Crowther, P. Royston and M. Clements, "A flexible parametric accelerated failure time model and the extension to time-dependent acceleration factors," *Biostatistics*, vol. 24, no. 3, pp. 811-831, 2023 .

- [10] N. Ghorbani, J. Yazdani-Charati and et al, "Application of the Weibull Accelerated Failure Time Model in the Determination of Disease-Free Survival Rate of Patients with Breast Cancer," *Iranian Journal Of Health Sciences*, vol. 4, no. 2, pp. 11-18, 2016.
- [11] R. K. Halder, M. N. Uddin and et al, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 1, p. 113, 2024.
- [12] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, p. 119–139, 1997.
- [13] T. K. Ho, "Random Decision Forests," in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, August 1995.
- [14] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control Signal Systems*, vol. 303–314, 1989.
- [15] A. M. Ahmed, A. Rizaner and A. H. Ulusoy, "A Decision Tree Algorithm Combined with Linear Regression," in *International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, Khartoum, Sudan, 2018.
- [16] X. Yan, "Linear Regression Analysis: Theory and Computing," *World Scientific*, Vols. pp. 1–2, ISBN 9789812834119, 2009.
- [17] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning.*, vol. 20 (3);273–297, 1995.
- [18] S. Pölsterl, "A Library for Time-to-Event Analysis Built on Top of scikit-learn," *Journal of Machine Learning Research*, vol. 21, no. 212, p. 1–6, 2020.
- [19] S. Pölsterl, "Gradient Boosted Models," 2020. [Online]. Available: https://scikit-survival.readthedocs.io/en/stable/user_guide/boosting.html. [Accessed 9 January 2024].
- [20] S. Pölsterl, "Random survival forests," 2020. [Online]. Available: https://scikit-survival.readthedocs.io/en/stable/user_guide/random-survival-forest.html. [Accessed 9 January 2024].

- [21] P. Cascante-Bonilla, F. Tan and et al, "Curriculum Labeling: Revisiting Pseudo-Labeling for Semi-Supervised Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 8, pp. 6912-6920, 2021.