Survival prediction of NLST lung cancer patients from radiomics and clinical data

Anh Leu
University of Illinois Urbana-Champaign
Vaishnavi Subramanian
University of Illinois Urbana-Champaign
Mary Dasso
University of Illinois Urbana-Champaign
Minh Do
University of Illinois Urbana-Champaign
Martha Menchaca (mmench1@uic.edu)
University of Illinois at Chicago

Research Article

Keywords: survival analysis, lung cancer, machine learning

Posted Date: March 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2600770/v1

Additional Declarations: No competing interests reported.

RESEARCH

Survival prediction of NLST lung cancer patients from radiomics and clinical data

Anh H. Leu[2]
, Vaishnavi Subramanian[2]
, Mary K. Dasso[2]
, Minh N. Do[2]
and Martha G. Menchaca[1*]

[*]Correspondence:
mmench1@uic.edu
[1]Department of Radiology, University of Illinois Chicago, Chicago, USA
Full list of author information is available at the end of the article

Abstract

Background: It is critical to accurately predict the survival likelihood of cancer patients to allow the best care and treatment. Publicly available datasets have emerged recently, such as the National Lung Screening Trial (NLST) data with low-dose computed tomography (LDCT) scans of at-risk populations. Recent research focuses primarily on improving survival prediction performance by developing more complex models without sufficient focus on interpretability. In contrast, this study focuses on identifying and analyzing the importance of different radiology features and clinical variables in survival prediction.

Methods: This research used the NLST data with two widely-used predictive models - the Cox proportional hazards (CPH) model and the random survival forests (RSF). The first step was to generate semi-automated primary nodule annotations for an NLST subset of adenocarcinoma patients with the close help of a radiologist and extract commonly-used radiomic features to characterize these primary nodules. By coupling the radiomic features with the patient's clinical data, the models predict death by lung cancer from the first LDCT scan. The next step is to construct smaller subsets of influential features and demonstrates that these subsets preserve survival performance. Additionally, this study investigated the potential of combining radiomic features and clinical data in survival prediction. Lastly, to make similar studies on the NLST more feasible, the semi-automated nodule segmentations for the NLST subset used in this study were provided for public use along with the code for the experiments at `https://github.com/hleu/survival_nlst`.

Results: The best result of 67.06 C-index and a mean time-dependent area under the receiver operating characteristic (TD-AUC) of 71.27 was obtained by using CPH models with a combination of a subset of clinical features and shape-based radiomic features.

Conclusions: The first contribution is the nodule annotation, segmentation, and exact feature extraction from the NLST dataset. This annotated dataset increases the model performance for lung cancer prognosis. Secondly, this study applied two different survival analysis methods to radiology and clinical features and compared the results obtained from different techniques. The models with a combination of features can outperform the model with only radiomics or clinical features.

Keywords: survival analysis; lung cancer; machine learning

1 Background

There were over 19 million new cancer cases worldwide, with over 2.2 million being lung cancer in 2020 [1]. Lung cancer was the leading cause of cancer-related deaths in both men and women, accounting for 22% of all cancer-related deaths in 2021 in the US according to the American Cancer Society [2]. Generally, lung cancer patients face a poor prognosis, with a five-year survival rate of only 21% [3]. To improve patient outcomes by identifying cancer early, the US Preventative Services Task Force recommends high-risk adults over 50 years old to undergo yearly screening with low dose computed tomography (LDCT) [4]. Survival analysis, or the estimation of time until death, from a lung cancer patient's LDCT scans can help plan patient care and management. However, this has remained a challenging task since it is hard to quantify the difference in appearance between benign and malignant nodules in the early stages of lung cancer.

To perform lung cancer prognosis, doctors usually follow the American Joint Committee on Cancer's (AJCC) TNM staging. The TNM staging captures the size and extent of the main tumor (T), and any spread of cancer to nearby lymph nodes (N) or to other parts of the body, metastastis (M). The TNM staging, although widely used, relies largely on the size and location of the tumor. The American College of Radiology's LungRADS classification system is also widely used to aid findings in low-dost CT screening examinations for lung cancer. The LungRADS scheme scores nodules based on size and visual appearance and recommends different follow-up times for different scores. Both of these commonly used schemes rely primarily on a manual evaluation of the nodules, which is prone to missed/incorrect cancer diagnosis [5].

In recent times, computer-aided diagnosis (CAD) systems have been developed to aid clinical decision making [6]. Such systems can help identify and quantify abnormalities on images to assist radiologists in quick decision-making. In particular, the extraction of radiomic features [7, 8, 9, 10, 11] has the potential to capture area, size and textural properties of the nodules from images, which can be missed by a visual examination with the human eye. been shown to improve lung cancer prognosis predictions [12, 13]. Complex radiomic features are also being captured using deep learning modules for the identification and computation of advanced radiomic features [14, 15, 16, 17, 18]. The computed radiomics features can be utilized for predicting patient survival outcomes using a range of prediction models.

Medical researchers employ survival prediction models to evaluate the significance of different prognostic variables, such as radiomic features and clinical variables. Standard survival analysis methods such as the Cox proportional hazards (CPH) [19] have been used extensively in lung cancer analysis. The CPH model is semi-parametric and estimates patients risk of death via inferred hazard functions. There have been a large number of works that apply machine learning (ML) methods to improve the estimation by using more complex prediction functions. One of the most commonly used ML-based survival analysis methods is the random survival forests (RSF) [20]. Deep-learning based ML survival analysis methods include DeepSurv [21] and DeepHit [22].

In addition to radiology imaging, modalities such as clinical factors, including age, gender, and smoking status, and genomics have been shown to impact individualized prognosis prediction [23, 24, 25, 26]. Detailed analysis concluded that medical

imaging modalities such as computed tomography (CT) offer patient-specific and tumor-specific information that could complement clinical information for better lung cancer prognosis [27].

Despite the growing interest in computer-aided survival prediction using ML tools, few works dive deeper into the interpretability of the models or the importance of features for prediction. Some works which target interpretable predictions in cancer include [28, 29, 30, 31]. These works focus on interpretability of imaging (and genomics) features from advanced ML prediction models. The understanding of feature importance for simpler prediction models such as CPH and RSF are still largely unexplored, especially in the context of lung cancer survival prediction. There is also limited work conducting a systematic study evaluating the potential of combining radiomic features and clinical variables. Our work aims to fill these gaps.

Specifically, we evaluate the radiomic and clinical features for survival prediction in cancer positive patients using the CPH and RSF models in a systematic study. We work with the largest publicly available lung cancer dataset - the National Lung Screening Trial (NLST) which provides low dose CT, clinical and survival information for the screened participants. We employ Cox proportional hazards (CPH) and random survival forests (RSF) as predictors. We utilize radiomic features and clinical features independently for the prediction. We quantify the importance of the different input features for the prediction to identify subsets of features that are most predictive of survival. We then investigate the predictive performance of these subsets. To test the use of multiple modalities in prediction, we also evaluate the predictive power of combined radiomic-clinical features. Further, to assist comprehensive studies on the NLST data, we publish our curated data of nodule segmentations and the extracted radiomic features for public use.

Our work is compared to previous works in the next section - Section 2. The survival methods and feature importance computation are detailed in Section 3. The NLST data and feature extraction schemes are discussed in Section 4 followed by the experiments (Section 5) and a discussion on the results (Section 6).

2 Related work

Several recent research efforts use machine learning and deep learning models to predict lung cancer properties from CT scans. The work by Huang et al. [32] predicts the likelihood of lung cancer incidence at regular time intervals using a two-layer multi-layer perceptron on LDCTs. Ardila et al. [33] built an end-to-end deep model to predict malignancy from a combination of CT scans over multiple time points and regions of interest (ROI) detected in those scans. Gao et al. [34] developed a multi-task loss to predict a cancer-free progression time using a 3D attention-based neural network, while Hawkins et al. [35] predicted malignancy of nodules from CT scans using random forests. In comparison to the above works, we tackle the the problem of survival prediction. Pérez-Morales et al. [36] characterized lung cancer nodule structural heterogeneity using computed tomography texture analysis (CTTA) [37] and identified key radiomic features. We extend this idea further to investigate clinical variables for the prediction.

Modalities other than LDCTs have also been utilized in lung cancer prediction models. Gao et al. [38] integrated clinical information with CT images in a multi-modal neural network co-learning framework for cancer risk estimation on the NLST. Subramanian et al. [26] combined LDCTs with genomics to predict lung cancer recurrence, while Yao et al. [29] investigated deep learning based survival prediction on the NLST and The Cancer Genome Archive (TCGA) datasets using histology imaging. In contrast to these works, we focus on simpler and more interpretable methods for prediction. We also quantify the importance of different variables and features for the prediction, which aids in interpretability.

3 Survival Prediction

Time-to-event analysis is a type of statistical analysis that, from a series of observations, attempts to estimate the time it takes for an event of interest to occur. When the event of interest is the risk of death, this analysis is referred to as *survival analysis* or *survival prediction*. Survival analysis examines the relationship of the death/survival variable distribution to the feature variables from observations, by assigning a time-varying likelihood of event (risk of death). It makes use of the time of first observation (for example, confirmed lung cancer), time of event of interest (death), and the duration between the first observation and event. For samples which do not observe the event of interest (the patient survives in the observation window), the observation duration is recorded instead.

3.1 Survival and hazard functions

The main goal of survival analysis is to estimate and interpret survival and hazard functions from survival data. The survival function $S(t)$ is the probability that the time of death T of a subject will be at least t, and is given by

$$S(t) = \mathbb{P}[t \leq T].$$

$S(t)$ is between 0 and 1, and is a non-increasing function of t. In theory, the survival function is smooth. However, in practice, the events are observed on a discrete timeline (days), making the graph of the survival function a step function. The hazard function $\lambda(t)$ is the probability of the death event occurring at time t, given that the subject did not experience the death event until time t, and is expressed as

$$\lambda(t) = \lim_{\delta t \to 0} \frac{\mathbb{P}[t \leq T < t + \delta t \mid t \leq T]}{\delta t}.$$

The hazard function describes the instantaneous potential per unit time for the death event to occur at time t, and usually changes over time. The survival function and hazard function are related. The hazard function $\lambda(t) = -S'(t)/S(t)$ where the survival $S(t)$ at time t is

$$S(t) = \exp\left(-H(t)\right), \quad H(t) = \int_0^t \lambda(z)dz,$$

where $H(t)$ as the cumulative hazard function (CHF).

In this work, we focus on two models of survival prediction: the semi-parametric Cox proportional hazards (CPH) model, and the machine-learning based random survival forests (RSF) model. These models are simple, yet powerful, and have been widely used in survival prediction tasks.

3.2 Cox proportional hazards model (CPH)

The CPH model [19] is a regression model used often in multivariate survival statistics to describe the relationship between the survival distribution and the covariates. The hazard function of a given data x is

$$\lambda(t|x) = \beta_0(t) \exp\left(\sum_{i=1}^{n} \beta_i x_i\right),$$

where $\sum_{i=1}^{n} \beta_i x_i$ is the log-partial hazard with the regression coefficient β_i weighing individual's observed covariates/features x_i, and the baseline hazard β_0 that can change over time t. The risk score, or the partial hazard is given by $\exp\left(\sum_{i=1}^{n} \beta_i(x_i)\right)$.

This is a semi-parametric model, with the baseline hazard function accounting for the time component, while the the regression coefficients β, remains constant throughout time. We can estimate these regression coefficients β_is that make up the partial hazard by maximizing the partial likelihood of the observed data. The regression coefficients can be interpreted as follows. A positive sign for the regression coefficient means that the corresponding feature is positively correlated to cancer risk, and the opposite for a negative regression coefficient. One of the biggest advantages of the CPH model is that we can interpret the coefficients associated with each feature used in the model, which can offer valuable insights.

3.3 Random survival forests (RSF)

Random survival forests [20] estimate the hazard functions differently, using a machine learning (ML) tree-based approach. This model is slightly more complex than the CPH, while still being easy to interpret. The main idea of RSFs is to grow trees to construct an ensemble estimate for the cumulative hazard function (CHF) $H(t)$. To do so, RSF first creates bootstrap samples from the original data, excluding on average 37% of the data called out-of-bag data (OOB data). Then, similar to Classification and Regression Tree Analysis (CART) [39], RSF grows a binary *survival tree* for each set of bootstrap samples. The algorithm recursively splits the tree nodes by maximizing survival difference using a predetermined survival criterion until each node in the tree becomes homogeneous with similar survival.

For each tree in the ensemble, all bootstrap samples in the same terminal (leaf) node h are considered for estimating the CHF corresponding to covariate $\mathbf{x}_i$ in node h at time t. Let $t_{(1,h)} < t_{(2,h)} < \cdots < t_{(N(h),h)}$ be the $N(h)$ distinct event times corresponding to the leaf node h. Then,

$$H(t|\mathbf{x}_i) = \hat{H}_h(t) = \sum_{t_{l,h} \leq t} \frac{d_{l,h}}{Y_{l,h}}, \quad \text{if } \mathbf{x}_i \in h,$$

where $d_{l,h}$ and $Y_{l,h}$ are the number of deaths and individuals at risk at time $t_{l,h}$ respectively. An ensemble CHF is then computed by averaging over all the survival trees to obtain the final hazards.

3.4 Comparison metrics

The performance of different survival analysis methods is usually compared using two metrics - Harrell's concordance index, and the time-dependent/cumulative area under ROC curve. The Harrell's concordance index, or the C-index, reflects how well the model predicts the ordering of patient risks - higher risks should be assigned to patients with smaller survival times. A value of 0.5 for C-index indicates a random model, and 1 indicates a perfect ranking of death times.

The cumulative/dynamic AUC ROC or time-dependent AUC (TD-AUC) measures the effectiveness of the model prediction as a function of time. The regular receiver operating characteristic curve (ROC curve) shows a classification model's performance at a range of classification thresholds by plotting the true positive rate (TPR) and the false positive rate (FPR) of a model. The AUC measures the area underneath the ROC curve. This is essentially a measure of the trade-off between TPR and FPR and ranges from 0 to 1, with 1 being the perfect classification. Since a patient's clinical status changes over time, sensitivity and specificity become time-dependent measures. The TD-AUC metric measures the area under the ROC curve as a function of time for a binary classification problem - distinguish patients who die by time t from patients who have not yet died at time t.

3.5 Feature selection

Once the predictive models are fitted, the feature importance of the different features can be computed to provide insights into the value of different features in the survival prediction. Knowing which features are significant enables us to select features that are important to the model and evaluate their predictive potential.

For the CPH model, we examine the regression coefficients of every input feature from the model. Based on the coefficient signs and values, we can understand the effect of different features on the hazard function. We also look at the confidence intervals (shown by the p-value) for each feature evaluated while fitting the CPH model to identify the important features.

For the RSF, we compute the feature importances by looking at the permutation feature importance weights from the model. The idea is to quantify how the score of the model decreases when a feature is not available, for example by replacing a chosen feature with noise drawn from the same distribution as original feature values. The easiest way to get that is to shuffle values for a feature. That way, the feature still belongs to the same distribution as the original feature values, but it will no longer carry helpful information. The feature importance is assigned as the reduction of the prediction score of the model when the feature is not available. We can thus compute the importance of the different features in the RSF prediction.

4 Data and feature extraction

In this work, we train prediction models on data from the NLST trial. This data is the largest publicly available dataset on lung cancer. The National Lung Cancer

Screening Trial (NLST) was a randomized and controlled trial between August 2002 and April 2004 [40]. It was initially a study that compared two ways of detecting lung cancer: LDCT and standard chest X-ray. The study found that patients who received LDCT scans had a lower risk of dying from lung cancer than those with chest X-rays. For this work, we are interested in the clinical information and the radiomics features obtained from the lung LDCT images. The lung LDCT images are available for the participants from up to three time points of initial and follow-up screenings. In this study, we focus on the first scan of the patients.

4.1 Generating train-test splits

The NLST dataset is a large dataset with 53,452 participants. We looked into a subset of 15,000 patients, of which 928 were cancer-positive patients, and 649 cancer-positive patients belonged to the CT arm. We focused on lung adenocarcinoma, the most common form of lung cancer, to improve the quality of our analysis and derive meaningful takeaways. Of the 928 cancer-positive patients, 318 patients had adenocarcinoma. Eliminating all CT scans that have technical problems or a slice thickness higher than 3 mm, we successfully collected the radiomic features from 284 patients first lung CT scan. We further combine the results with clinical information and further exclude patients with lots of missing fields. leaving 263 patients. These patients information are randomly divided according to a train/test ratio of 80/20, resulting in the train set of 208 patients and a test set of 55 patients. For each patient, we use death by cancer as the event of interest and calculate the time duration between the first scan until death (by cancer or other means), or the last follow-up date.

4.2 Clinical features

The NLST dataset provides demographic and clinical characteristics information, such as age, education, ethnicity, gender, race, personal smoking information, personal diseases history, personal cancer history, and family cancer history. Here, we used the demographic and clinical characteristics information as the clinical data.

To pre-process the clinical data, we followed the outline in Fig. 1. The clinical data includes three different types of data - Boolean, numerical, and categorical - which we process via a pre-processing pipeline in `scikit-learn`. For Boolean data, we filled the `N/A` values with `False`. For numerical data, we used SimpleImputer to fill `N/A` values with the feature column's mean, then applied StandardScaler to scale the column to be between 0 and 1. We used OneHotEncoder conversion method to map all the categories of each feature column to a separate Boolean feature column for categorical data. Here, the `N/A` values are considered a separate category for each feature column.

Some of those columns may have very little variation, thus increasing the model bias, making it necessary to combine columns. For example, out of 208 patients, we have 189 patients belonging to the White race and a total of 19 patients of all other races combined (Black or African-American, Asian, American Indian, or Alaskan Native, Native Hawaiian or Other Pacific Islander, and others). This discrepancy is a significant bias in the data. To alleviate this data bias, we combined all the other races into one feature. A similar behavior is observed with ethnicity category, where

those columns have very low variance when conditioned on death event present or not, and we decided to drop this category. The specific clinical features distribution is in Table 1.

Table 1 Clinical feature distributions information for the NLST data split used. Values reported are mean (SD) for the first numerical data or number n (%) in selected category for the following categorical data

Clinical features	Training (n=208)	Testing (n=55)
Age	63.75 (5.37)	63.78 (4.68)
BMI	26.22 (4.01)	27.53 (4.22)
Pack years	65.79 (27.39)	62.18 (29.51)
Smoking start age	16.28 (3.55)	16.07 (3.41)
Cigarettes per day	30.62 (12.44)	28.0 (11.5)
Number of smoking years	43.44 (7.12)	44.75 (7.32)
High school graduate	50 (24.04)	15 (27.78)
Post HS training	26 (12.50)	4 (7.41)
Associate degree	55 (26.44)	13 (24.07)
Bachelors degree	31 (14.90)	10 (18.52)
Graduate school	24 (11.54)	9 (16.67)
Non-white	19 (9.13)	32 (59.26)
Female	101 (48.56)	18 (33.33)
Smoking at the start of trial	115 (55.29)	6 (11.11)
Lived with smoker	182 (87.50)	46 (85.19)
Worked with smoker	188 (90.38)	49 (90.74)
Cancer prior to trial	20 (9.62)	6 (11.11)
Family member had cancer	61 (29.33)	14 (25.93)

4.3 Radiomic features

We followed the outline in Fig. 2 to obtain the radiomic features from the NLST dataset. For this, we make use of the radiological characteristics information, such as information about CT screening abnormalities and size of nodules, alongside the LDCTs.

The NLST provides the approximate location of the primary nodule in the lung, the approximate sizes of the nodules, and the approximate locations of all nodules in the last time point. With the radiologist's guidance, we identified the suspected primary nodule in the last available CT time point of the patient. By tracing the same nodule back to earlier time points, we marked the nodule's approximate center. We annotated the bounding box for 462 primary nodules of 310 cancer patients at up to 3 CT screening time points, 8 patients scans were not readable. Examples of nodule bounding boxes are in Fig. 3.

Using the bounding box's center point, we can segment out a cube size of 90 mm^3. We picked 90 because the maximum diameter of a nodule in the dataset is 80 mm. Then, we resampled all the nodule cubes that we got to the same slice thickness of 1.7 mm and pixel spacing of 0.7 mm. These values are picked based on the approximate median of slice thickness and pixel spacing of all CT images. We then used a semi-automatic tool, 3D Slicer [41], to generate a tighter segmentation of the nodule from the nodule bounding boxes. Examples of images and the corresponding segmentations are shown in Fig. 4. The CT images restricted to the segmented areas serve as the regions of interest (ROIs) in our experiments.

From the nodule segmentation, we employed PyRadiomics [42] to extract radiomic features from the nodules segmentation. We were able to only extract radiomics features from 284 patients.

We extracted 106 nodule features, separated into eight classes: 14 shape-based (both 2D and 3D) features, 17 first-order statistics (intensity) features, 24 gray-level covariance matrix (GLCM) features, 14 gray-level dependence matrix (GLDM) features, 16 gray-level run-length matrix (GLRLM) features, 16 gray-level size zone matrix (GLSZM) features, and 5 neighboring gray-tone difference matrix (NGTDM) features.

Since all the radiomic features are numerical, it suffices to preprocess radiomic features similar to how we handle numerical clinical features, using SimpleImputer to fill `N/A` values with the feature column's mean, then applying StandardScaler. Doing this allows the data to have zero mean and standard deviation. Additionally, to input features into the CPH model, we check the correlation matrix of features to make sure not to have input features with high correlation. For this, we identify the variables with high correlation and remove one of the highly correlated variable.

5 Experiments and results

5.1 Experimental Setup

We first experiment using only clinical information as input features and only radiomic features as input to both the prediction models - the Cox proportional hazards (CPH) model and the random survival forest (RSF). For each prediction model, we evaluate the importance of the different features. We pick the subset of the most important features and predict survival using this subset alone to understand the predictive ability of the subset. Finally, to evaluate the potential of aggregating information from the clinical and radiology modalities, we combine the subsets of representative features from clinical and radiology data and predict survival. To summarize, we study the survival prediction performance using:

- Radiomics and clinical features independently
- Subsets of important radiomics and clinical features independently
- Combined subsets of important radiomics and clinical features.

We utilized the survival analysis package `lifelines` [43] package for CPH predictions and the `scikit-survival` [44] for the RSFs. We tuned the hyper-parameters using GridSearch on the training set. Throughout the experiments we fix a random seed for consistency across runs. To evaluate the models, we looked at the C-index from the model to check how well the model performs in risk estimation. We also looked at the TD-AUC of the whole model to quantify how well the model predicts over time. The mean TD-AUC of the model here is the mean throughout all time points of the TD-AUC every 30 days (one month) up to the end of the study. All results are reported on the testing set. The data and code for experiments can be found at `https://github.com/hleu/survival_nlst`.

5.2 Clinical features

We input the pre-processed clinical variables data to both the CPH model and the RSF model. The C-index and mean TD-AUC results are reported in Table 2 and the TD-AUC curve is plotted in Fig. 5.

From Fig. 5, we see that the RSF model is the most effective in predicting death in the early term, within the first 1000 days (roughly around three years). In contrast, the CPH model seems to be better at predicting death by cancer in the middle and late term, after around 1000 days (roughly around three years).

Table 2 C-index and mean TD-AUC (MTD-AUC) using clinical variables with CPH and RSF on the testing set.

Features	Model	C-index	MTD-AUC
Full set	CPH	54.81	50.51
	RSF	48.35	53.32
Subset	CPH	60.21	61.68
	RSF	53.03	54.14

We then examine which features are important for the prediction of the two models. For the CPH, we examined the regression coefficients of every input feature from the model. For the RSF, we acquired the weight table by looking at the permutation feature importance weights from the model. From there, we checked the significance of a specific feature. The importance weights of the different clinical features are documented in Table 3.

Table 3 Clinical feature importance for CPH and RSF models.

	CPH		RSF
	Coefficient	p-value	Importance Weight
Age	0.0480	0.8247	0.0442 ± 0.0115
Associate-degree	0.2474	0.6397	0.0049 ± 0.0034
Bachelors-degree	0.0330	0.9551	0.0041 ± 0.0035
BMI	0.0843	0.5480	0.0911 ± 0.0173
Cancer-prior-to-trial	0.6341	0.1235	0.0074 ± 0.0027
Cigarettes-per-day	-0.5407	0.4824	0.0244 ± 0.0070
Family-member-had-cancer	0.1164	0.6901	0.0071 ± 0.0033
Female	-0.5367	0.0613	0.0248 ± 0.0067
Graduate-school	0.0470	0.9388	0.0038 ± 0.0014
High-school-graduate	0.4243	0.4150	0.0132 ± 0.0038
Lived-with-smoker	0.5899	0.2294	0.0046 ± 0.0018
Non-white	-0.4223	0.4424	0.0036 ± 0.0014
Number-of-smoking-years	-0.2417	0.5540	0.0422 ± 0.0131
Pack-year	0.4975	0.5151	0.0391 ± 0.0126
Post-HS-training	0.3389	0.5676	0.0153 ± 0.0066
Smoking-at-the-start-of-trial	0.1143	0.7579	0.0086 ± 0.0034
Smoking-start-age	0.0950	0.6233	0.0392 ± 0.0116
Worked-with-smoker	-0.3418	0.4599	0.0145 ± 0.0047

From Table 3, we identified the most important features for the prediction based on the lowest p-values for CPH and the highest importance weights for the RSF model. From this, the top six features which are important for both models are chosen: BMI, Pack-year, Smoking-start-age, Female, Worked-with-smoker, and Cancer-prior-to-trial. We run this subset of features as inputs to the CPH and RSF models and perform the earlier analysis again. The survival prediction results for these are summarized in Table 2 and Fig. 6.

Based on Fig. 6, we see that the overall AUC for both CPH and RSF models improved, showing the importance of selecting features, suggesting that both models benefit from some feature selection method. In addition, we see that while the CPH model still seems to be better at predicting death by cancer in the middle and late term, after around 1000 days (roughly around three years), CPH and RSF models perform around the same at the earlier stage.

5.3 Radiomic features

Next, we look into the predictive performance of the radiomic features. Since we only had 208 patients with possible radiomic features extracted for training while there are 106 radiomic features per patient and many features are correlated to each other, training the model with all features would not be practical. Hence,

we decided to separate all the radiomic features into smaller sets to see how each radiomic feature can affect the outcomes. Here, we will look at three different sets of radiomic features:

1 Radiomics-1: Shape-based features only. These features include descriptors of the 2D and 3D size and shape of the ROI.
2 Radiomics-2: First-order statistics features (Intensity) only. First-order statistics describe the distribution of voxel intensities of the image ROI defined by the mask through commonly used and basic metrics.
3 Radiomics-3: Features selected from a statistical analysis that significantly differentiates patients into risk groups [36]. The features include peritumoral radiomic features and an intratumoral radiomic feature. The peritumoral radiomic features are the average co-occurrence joint entropy, the average co-occurrence angular second moment, and the NGTDM busyness. The intratumoral radiomic feature is the statistical RMS.

Looking at Fig 7, we can see the TD-AUC for all radiomic features. Based on those graphs, we observe that overall, RSF model performs better for Radiomics-1 and Radiomics-3 group while CPH model performs better for Radiomics-2 group throughout all time periods.

Next, similar to with clinical features, we checked the feature importance of CPH and RSF models for Radiomics-1 and Radiomics-2. We exclude Radiomics-3 in this step, because there are only four features in this group. The feature importance for the two groups are depicted in Table 4 and Table 5 correspondingly. Based on this, we created a subset of features for each group and run these subsets of features as inputs to the CPH and RSF models again. Specifically, we selected six features that appears in the top important features for both CPH and RSF models. Thus, we picked for group Radiomics-1 are 6 features: Maximum-2D-Diameter-Column, Maximum-2D-Diameter-Row, Maximum-2D-Diameter-Slice, Maximum-3D-Diameter, Sphericity, and Voxel-Volume. Similarly, the features we picked for group Radiomics-2 are: First-Order-Energy, First-Order-Kurtosis, First-Order-Maximum, First-Order-Median, First-Order-90th-Percentile, and First-Order-Root-Mean-Squared.

Table 4 Radiomics 1 feature importance fåor CPH and RSF models

	CPH		RSF
	Coefficient	p-value	Importance Weight
Elongation	-0.1839	0.1973	0.0212 ± 0.0147
Flatness	0.0745	0.6149	0.0091 ± 0.0066
Least-Axis-Length	0.0158	0.9353	0.0072 ± 0.0120
Major-Axis-Length	0.0995	0.6038	0.0203 ± 0.0180
Maximum-2D-Diameter-Column	0.2512	0.1881	0.0151 ± 0.0127
Maximum-2D-Diameter-Row	0.1078	0.5731	0.0103 ± 0.0190
Maximum-2D-Diameter-Slice	-0.0970	0.6051	0.0107 ± 0.0072
Maximum-3D-Diameter	0.1639	0.4140	0.0065 ± 0.0104
Minor-Axis-Length	-0.0264	0.8958	0.0098 ± 0.0104
Sphericity	0.1697	0.2048	0.0318 ± 0.0163
Surface-Area	0.0723	0.7077	0.0027 ± 0.0061
Surface-Volume-Ratio	0.0166	0.9127	0.0308 ± 0.0212
Voxel-Volume	-0.1431	0.3484	0.0025 ± 0.0063

The C-index and mean TD-AUC for all the models are listed in Table 6. From that, we can compare the results using different models. We can see that, RSF model achieve better results for Radiomics-1 and Radiomics-3 group while CPH model is better for Radiomics-2 group. In addition, we can compare the results using different

Table 5 Radiomics 2 feature importance for CPH and RSF models

	CPH		RSF
	Coefficient	p-value	Importance Weight
First-Order-10th-Percentile	0.0000	0.9992	0.0099 ± 0.0028
First-Order-90th-Percentile	0.1222	0.4860	0.0114 ± 0.0038
First-Order-Energy	0.2898	0.0068	0.0186 ± 0.0074
First-Order-Entropy	0.0000	0.9994	0.0079 ± 0.0036
First-Order-Interquartile-Range	0.0000	1.0000	0.0174 ± 0.0105
First-Order-Kurtosis	-0.0899	0.5035	0.0219 ± 0.0070
First-Order-Maximum	-0.1003	0.4537	0.0438 ± 0.0126
First-Order-Median	0.0971	0.6299	0.0140 ± 0.0092
First-Order-Minimum	0.0729	0.5872	0.0098 ± 0.0033
First-Order-Root-Mean-Squared	-0.1331	0.5187	0.0117 ± 0.0062
First-Order-Skewness	-0.0228	0.9081	0.0158 ± 0.0090
First-Order-Total-Energy	0.0000	0.9993	0.0205 ± 0.0079
First-Order-Uniformity	0.0000	0.9992	0.0092 ± 0.0041
First-Order-Variance	0.0000	0.9995	0.0119 ± 0.0031

input features. Here, we see that using shape-based features as inputs achieves the best model outcome for both CPH and RSF models. Similar to clinical features, results of subsets of features are better than a full set, showing the importance of features selection. Finally, from Fig 8, we observe that CPH and RSF models perform relatively similar to each other over time except when CPH model performs slightly better than RSF model for Radiomics-2 subset.

Table 6 C-index and mean TD-AUC (MTD-AUC) using radiomics variables with CPH and RSF on the testing set.

Features	Model	C-index	MTD-AUC
Radiomics 1 - Full set	CPH	60.21	56.96
	RSF	64.10	62.36
Radiomics 1 - Subset	CPH	60.08	62.34
	RSF	63.77	64.78
Radiomics 2 - Full set	CPH	55.73	51.52
	RSF	49.93	44.07
Radiomics 2 - Subset	CPH	56.79	53.06
	RSF	52.44	47.38
Radiomics 3 - Full set	CPH	59.42	55.37
	RSF	62.45	63.61

5.4 Combination of radiomic and clinical features

Lastly, in addition to using radiomic features, we tested features as the combination of clinical, and shape-based features since shape-based features produce best results. This combination resulted in the best results for both CPH and RSF models, showed in Table 7. This improvement shows that a combination of clinical and radiomic features does make a difference in prognosis for lung cancer patients.

Table 7 C-index and mean TD-AUC (MTD-AUC) using combination with CPH and RSF on the testing set.

Model	C-index	MTD-AUC
CPH	67.06	71.27
RSF	64.69	66.45

Fig. 9 plots the TD-AUC graph for the best RSF model against the TD-AUC for the best CPH model for combination features. From the graph, we observe that the CPH model outperforms RSF throughout time. This is especially seen for early follow-up times (less than 500 days), where the TD-AUC for CPH model can reach up to roughly 0.8 or 80%.

Finally, comparing all model performances, we notice that CPH model with combination inputs is the best, with C-index value of 67.06% and mean TD-AUC of 71.27%. The highest TD-AUC value is around 80% around right before 500 days mark for CPH model with combination inputs. Overall, based on all TD-AUC graphs, the CPH model with combination inputs performs the best for follow up time before 500 days mark or after 1000 days, while the RSF model with Radiomics-3 group inputs performs the best for follow up time around between 500 days and 1000 days mark.

6 Conclusion

6.1 Significance

This work makes two main contributions. The first is the nodule annotation, segmentation, and feature extraction from the NLST dataset. The original NLST dataset only provided the approximate location and the slice with the nodule's largest diameter. However, that severely limits the number of tasks that we can perform using the dataset. Critical studies such as nodule segmentation, nodule classification, and survival analysis using lung nodules require the exact segmentation of nodules. However, it is expensive and time-consuming to annotate a dataset. This work provides an annotated dataset for the NLST. We demonstrated the importance of this annotated dataset using only a subset of the final total annotations. With only annotations about the suspected primary nodule for confirmed cancer patients at the first time point, we observed an increase in model performance for lung cancer prognosis.

Secondly, this work applied two different survival analysis methods to radiology and clinical features and compared the results obtained from different techniques. Using different methods, we looked closely at each feature's relationship and feature-to-model relationship. We consistently showed the significance of the radiomic features obtained from the annotated dataset both separately and with the clinical information from multiple experiments. We also examined the performance of each feature and model throughout time using the dynamic/cumulative AUC ROC. From our experiments, we can conclude that the models with a combination of features can outperform the model with only radiomics or clinical features.

6.2 Limitation

However, this work does suffer some limitations. The first is the small number of patients available. Because of the currently small dataset and many features per entry, the model easily overfits. We attempted to alleviate this problem by using a simpler model and separating radiomic features into smaller subsets. However, the RSF model still differs significantly between the train and test performance. With a larger dataset, the model should achieve better performance.

The second limitation is the feature selection process. This work selected basic clinical features based on the radiologist's general cancer study recommendations, not specifically for a prognostic model. However, some features such as race or ethnicity were significantly skewed in the dataset, making inference on those features impossible. For radiomic features, we separated features into different types and arbitrarily took out features based on the correlation matrix. We concatenated the

subset of clinical features with the shape-based radiomic features for the final combination model. Better feature selection methods could boost model performance.

With this work as the starting point, we can enhance the existing model using better feature selections and hyperparameter tuning techniques. Additionally, we can expand on the topic of lung cancer prognosis using different statistical models for a more rigorous evaluation. We can also investigate the use of deep learning models which extract complex features directly from the nodule segmentation to predict survival.

Acknowledgements
Not applicable.

Funding
No funding was received.

Abbreviations
AUC: Area under the ROC Curve
C-index: Concordance index
CART: Classification and Regression Tree Analysis
CHF: Cumulative hazard function
CPH: Cox proportional hazard
CT: Computed tomograph
CTTA: Computed tomography texture analysis
FPR: False positive rate
GLCM: Gray-level covariance matrix
GLDM: gray-level dependence matrix
GLRLM: gray-level run-length matrix
GLSZM: gray-level size zone matrix
LDCT: Low-dose computed tomography
ML: Machine Learning
NGTDM: neighboring gray-tone difference matrix
NLST: National Lung Screening Trial
ROC: Receiver operating characteristic
ROI: Region of interest
RSF: Random survival forests
RSM: Root mean square
TCGA: The Cancer Genome Archive
TD: Time-dependent
TPR: True positive rate

Availability of data and materials
The original NLST dataset analyzed during this study is available at: `https://cdas.cancer.gov/datasets/nlst/`. The semi-annotated nodule segmentations for the NLST subset used in this study are available at: `https://github.com/hleu/survival_nlst`

Ethics approval and consent to participate
The use of this public dataset did not need ethical approval.

Competing interests
The authors declare no competing interests.

Consent for publication
Not applicable.

Authors' contributions
All authors gave approval for the submitted version and have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work. AHL: Conceptualization, Methodology, Data curation, Formal analysis, Supervision of the paper writing, Writing-Original draft, and final manuscript. VS: Conceptualization, Methodology, Data curation, Formal analysis, Supervision of the paper writing, Writing - Review & Editing. MKD: Data curation, Writing - Review & Editing. MND: Conceptualization, Methodology, Supervision of the paper writing, MGM: Conceptualization, Data curation, Supervision of the paper writing, Writing - Review & Editing.

Authors' information
Author details
[1]Department of Radiology, University of Illinois Chicago, Chicago, USA. [2]Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, USA.

References

1. Sung, H., et al.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians (2021)
2. American Cancer Society, A.: Cancer Facts and Figures 2021. American Cancer Society (2021)
3. Howlader, N., Noone, A.M., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., Lewis, D.R., Chen, H.S., Feuer, E.J.: SEER cancer statistics review, 1975-2017. National Cancer Institute, Bethesda, MD (2020)
4. Force, U.P.S.T.: Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. JAMA **325**(10), 962–970 (2021). doi:10.1001/jama.2021.1117
5. Del Ciello, A., Franchi, P., Contegiacomo, A., Cicchetti, G., Bonomo, L., Larici, A.R.: Missed lung cancer: when, where, and why? Diagnostic and interventional radiology **23**(2), 118 (2017)
6. Chan, H.-P., Hadjiiski, L.M., Samala, R.K.: Computer-aided diagnosis in the era of deep learning. Medical physics **47**(5), 218–227 (2020)
7. Gillies, R.J., Kinahan, P.E., Hricak, H.: Radiomics: Images are more than pictures, they are data. Radiology **278**, 2 (2016)
8. Rizzo, S., Botta, F., Raimondi, S., Origgi, D., Fanciullo, C., Morganti, A.G., Bellomi, M.: Radiomics: the facts and the challenges of image analysis. European radiology experimental **2**(1), 1–8 (2018)
9. Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R.G., Granton, P., Zegers, C.M., Gillies, R., Boellard, R., Dekker, A., *et al.*: Radiomics: extracting more information from medical images using advanced feature analysis. European journal of cancer **48**(4), 441–446 (2012)
10. Avanzo, M., Stancanello, J., El Naqa, I.: Beyond imaging: the promise of radiomics. Physica Medica **38**, 122–139 (2017)
11. Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.-C., Pieper, S., Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. Cancer research **77**(21), 104–107 (2017)
12. Zhang, Y., Oikonomou, A., Wong, A., Haider, M.A., Khalvati, F.: Radiomics-based prognosis analysis for non-small cell lung cancer. Scientific Reports **7**(1), 46349 (2017). doi:10.1038/srep46349
13. He, B., Zhao, W., Pi, J.-Y., Han, D., Jiang, Y.-M., Zhang, Z.-G.: A biomarker basing on radiomics for the prediction of overall survival in non–small cell lung cancer patients. Respiratory Research **19**(1), 199 (2018). doi:10.1186/s12931-018-0887-8
14. Avanzo, M., Wei, L., Stancanello, J., Vallieres, M., Rao, A., Morin, O., Mattonen, S.A., El Naqa, I.: Machine and deep learning methods for radiomics. Medical physics **47**(5), 185–202 (2020)
15. Avanzo, M., Stancanello, J., Pirrone, G., Sartor, G.: Radiomics and deep learning in lung cancer. Strahlentherapie und Onkologie **196**(10), 879–887 (2020)
16. Hatt, M., Parmar, C., Qi, J., El Naqa, I.: Machine (deep) learning methods for image processing and radiomics. IEEE Transactions on Radiation and Plasma Medical Sciences **3**(2), 104–108 (2019)
17. Parekh, V.S., Jacobs, M.A.: Deep learning and radiomics in precision medicine. Expert review of precision medicine and drug development **4**(2), 59–72 (2019)
18. Afshar, P., Mohammadi, A., Plataniotis, K.N., Oikonomou, A., Benali, H.: From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. IEEE Signal Processing Magazine **36**(4), 132–160 (2019)
19. Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological) **34**(2), 187–220 (1972)
20. Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S.: Random survival forests. The Annals of Applied Statistics **2**(3), 841–860 (2008). doi:10.1214/08-AOAS169
21. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC medical research methodology **18**(1), 1–12 (2018)
22. Lee, C., Zame, W.R., Yoon, J., van der Schaar, M.: Deephit: A deep learning approach to survival analysis with competing risks. In: Thirty-second AAAI Conference on Artificial Intelligence (2018)
23. Liang, W., Zhang, L., Jiang, G., Wang, Q., Liu, L., Liu, D., Wang, Z., Zhu, Z., Deng, Q., Xiong, X., Shao, W., Shi, X., He, J.: Development and validation of a nomogram for predicting survival in patients with resected non–small-cell lung cancer. Journal of Clinical Oncology **33**(8), 861–869 (2015). doi:10.1200/JCO.2014.56.6661. PMID: 25624438. https://doi.org/10.1200/JCO.2014.56.6661
24. Warren, G.W., Kasza, K.A., Reid, M.E., Cummings, K.M., Marshall, J.R.: Smoking at diagnosis and survival in cancer patients. International Journal of Cancer **132**(2), 401–410 (2013)
25. Gevaert, O., Xu, J., Hoang, C.D., Leung, A.N., Xu, Y., Quon, A., Rubin, D.L., Napel, S., Plevritis, S.K.: Non–small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. Radiology **264**(2), 387–396 (2012)
26. Subramanian, V., Do, M.N., Syeda-Mahmood, T.: Multimodal fusion of imaging and genomics for lung cancer recurrence prediction. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 804–808 (2020). IEEE
27. Lu, M.T., Ivanov, A., Mayrhofer, T., Hosny, A., Aerts, H.J.W.L., Hoffmann, U.: Deep learning to assess long-term mortality from chest radiographs. JAMA Network Open **2**(7), 197416–197416 (2019). doi:10.1001/jamanetworkopen.2019.7416
28. Wulczyn, E., Steiner, D.F., Moran, M., Plass, M., Reihs, R., Tan, F., Flament-Auvigne, I., Brown, T., Regitnig, P., Chen, P.-H.C., *et al.*: Interpretable survival prediction for colorectal cancer using deep learning. NPJ digital medicine **4**(1), 1–13 (2021)
29. Yao, J., Wang, S., Zhu, X., Huang, J.: Imaging biomarker discovery for lung cancer survival prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 649–657 (2016). Springer

30. Oh, J.H., Choi, W., Ko, E., Kang, M., Tannenbaum, A., Deasy, J.O.: Pathcnn: interpretable convolutional neural networks for survival prediction and pathway analysis applied to glioblastoma. Bioinformatics **37**(Supplement_1), 443–450 (2021)
31. Smedley, N.F., Aberle, D.R., Hsu, W.: Using deep neural networks and interpretability methods to identify gene expression patterns that predict radiomic features and histology in non-small cell lung cancer. Journal of Medical Imaging **8**(3), 031906 (2021)
32. Huang, P., Lin, C.T., Li, Y., Tammemagi, M.C., Brock, M.V., Atkar-Khattra, S., Xu, Y., Hu, P., Mayo, J.R., Schmidt, H., Gingras, M., Pasian, S., Stewart, L., Tsai, S., Seely, J.M., Manos, D., Burrowes, P., Bhatia, R., Tsao, M.-S., Lam, S.: Prediction of lung cancer risk at follow-up screening with low-dose CT: A training and validation study of a deep learning method. The Lancet Digital Health **1**(7), 353–362 (2019). doi:10.1016/S2589-7500(19)30159-1
33. Ardila, D., Kiraly, A.P., Bharadwaj, S., Choi, B., Reicher, J.J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., Naidich, D.P., Shetty, S.: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine **25**(6), 954–961 (2019). doi:10.1038/s41591-019-0447-x
34. Gao, R., Li, L., Tang, Y., Antic, S.L., Paulson, A.B., Huo, Y., Sandler, K.L., Massion, P.P., Landman, B.A.: Deep multi-task prediction of lung cancer and cancer-free progression from censored heterogenous clinical imaging. In: Medical Imaging 2020: Image Processing, vol. 11313, p. 113130 (2020). International Society for Optics and Photonics
35. Hawkins, S., Wang, H., Liu, Y., Garcia, A., Stringfield, O., Krewer, H., Li, Q., Cherezov, D., Gatenby, R.A., Balagurunathan, Y., *et al.*: Predicting malignant nodules from screening ct scans. Journal of Thoracic Oncology **11**(12), 2120–2128 (2016)
36. Pérez-Morales, J., Tunali, I., Stringfield, O., Eschrich, S.A., Balagurunathan, Y., Gillies, R.J., Schabath, M.B.: Peritumoral and intratumoral radiomic features predict survival outcomes among patients diagnosed in lung cancer screening. Scientific Reports **10**(1), 10528 (2020). doi:10.1038/s41598-020-67378-8
37. Lubner, M.G., Smith, A.D., Sandrasegaran, K., Sahani, D.V., Pickhardt, P.J.: CT texture analysis: Definitions, applications, biologic correlates, and challenges. RadioGraphics **37**(5), 1483–1503 (2017). doi:10.1148/rg.2017170056. PMID: 28898189. https://doi.org/10.1148/rg.2017170056
38. Gao, R., Tang, Y., Khan, M.S., Xu, K., Paulson, A.B., Sullivan, S., Huo, Y., Deppen, S., Massion, P.P., Sandler, K.L., *et al.*: Cancer risk estimation combining lung screening ct with clinical data elements. Radiology: Artificial Intelligence **3**(6), 210032 (2021)
39. Breiman, L., Friedman, J., Olshen, R., Stone, C.: Classification and Regression Trees. Wadsworth and Brooks, Monterey, CA (1984)
40. National Lung Screening Trial Research Team: The national lung screening trial: overview and study design. Radiology **258**(1), 243–253 (2011)
41. Kikinis, R., Pieper, S.D., Vosburgh, K.G.: 3D Slicer: A platform for subject-specific image analysis, visualization, and clinical support. In: Jolesz, F.A. (ed.) Intraoperative Imaging and Image-Guided Therapy, pp. 277–289. Springer, New York, NY (2014)
42. van Griethuysen, J.J.M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G.H., Fillion-Robin, J.-C., Pieper, S., Aerts, H.J.W.L.: Computational radiomics system to decode the radiographic phenotype. Cancer Research **77**(21), 104–107 (2017). doi:10.1158/0008-5472.CAN-17-0339
43. Davidson-Pilon, C.: lifelines: survival analysis in python. Journal of Open Source Software **4**(40), 1317 (2019). doi:10.21105/joss.01317
44. Pölsterl, S.: scikit-survival: A library for time-to-event analysis built on top of scikit-learn. Journal of Machine Learning Research **21**(212), 1–6 (2020)

Figures

Figure 1 Pipeline for pre-processing clinical features

Figure 2 Pipeline for extracting radiomic features from low-dose CTs (LDCTs)

Figure 3 Examples of nodule bounding boxes
Including tight nodule bounding boxes (red) and bounding boxes that include surrounding areas (green) along with slice number and location of nodule reported in the NLST dataset from four different patients

Figure 4 Examples of nodule bounding boxes and nodule segmentations
Including nodule bounding boxes (left) and nodule segmentations extracted with 3D Slicer (right) for two patients

Figure 5 Time-dependent AUC curves with clinical data

Figure 6 Time-dependent AUC curves with a subset of clinical data

Figure 7 TD-AUC for radiomic features

Figure 8 TD-AUC for subsets of radiomic features

Figure 9 TD-AUC for subsets of combination of features

Figures

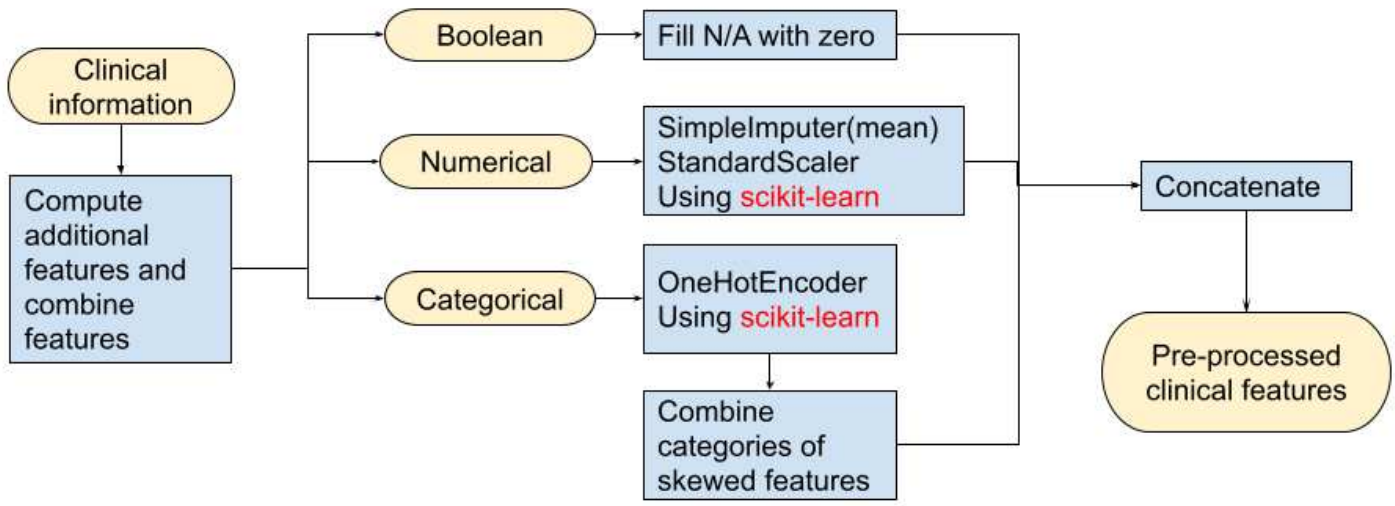

Figure 1

Pipeline for pre-processing clinical features

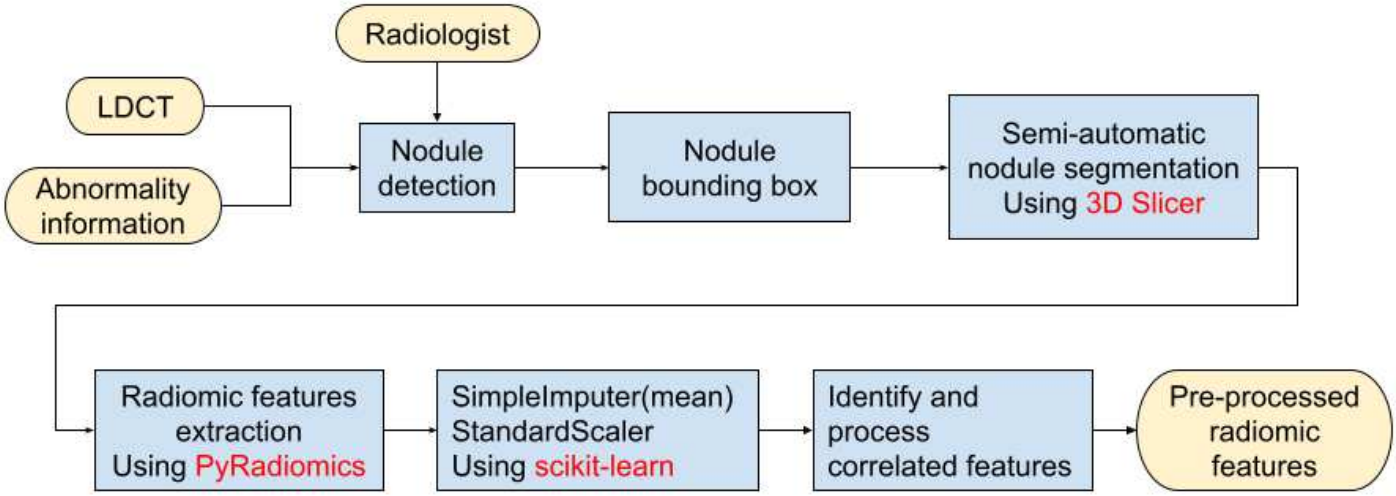

Figure 2

Pipeline for extracting radiomic features from low-dose CTs (LDCTs)

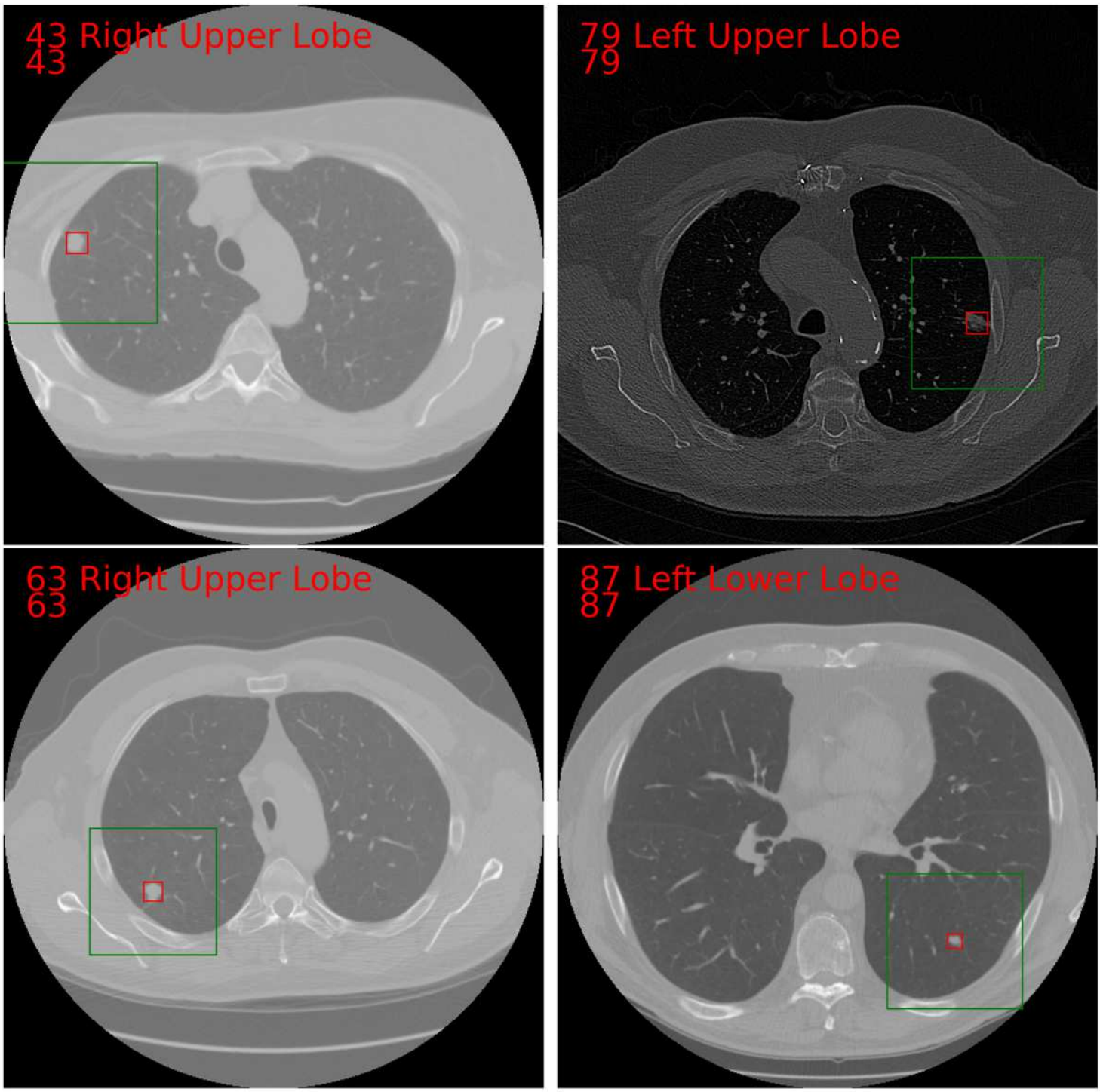

Figure 3

Examples of nodule bounding boxes

Including tight nodule bounding boxes (red) and bounding boxes that include surrounding areas (green) along with slice number and location of nodule reported in the NLST dataset from four different patients

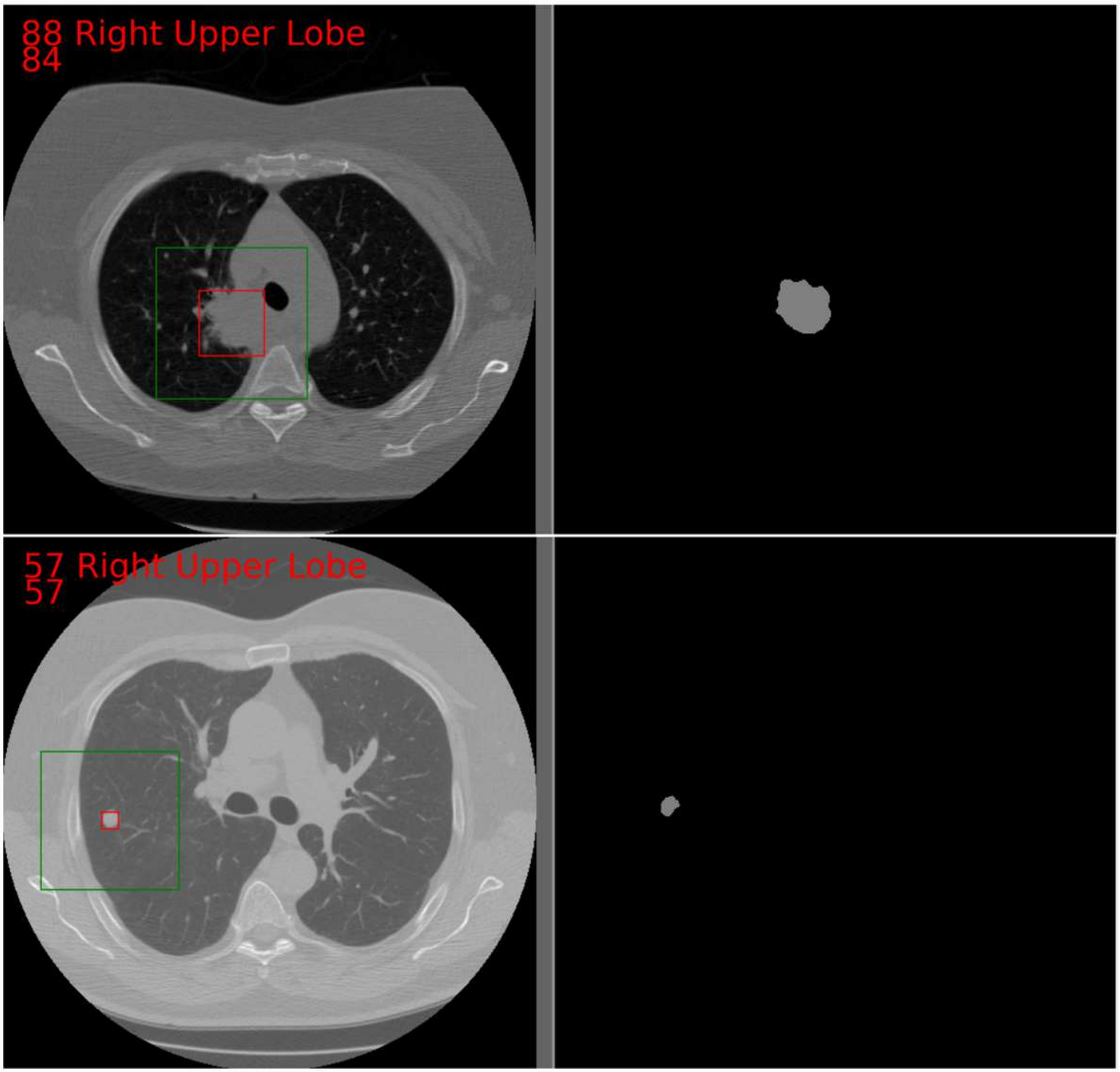

Figure 4

Examples of nodule bounding boxes and nodule segmentations

Including nodule bounding boxes (left) and nodule segmentations extracted with 3D Slicer (right) for two patients

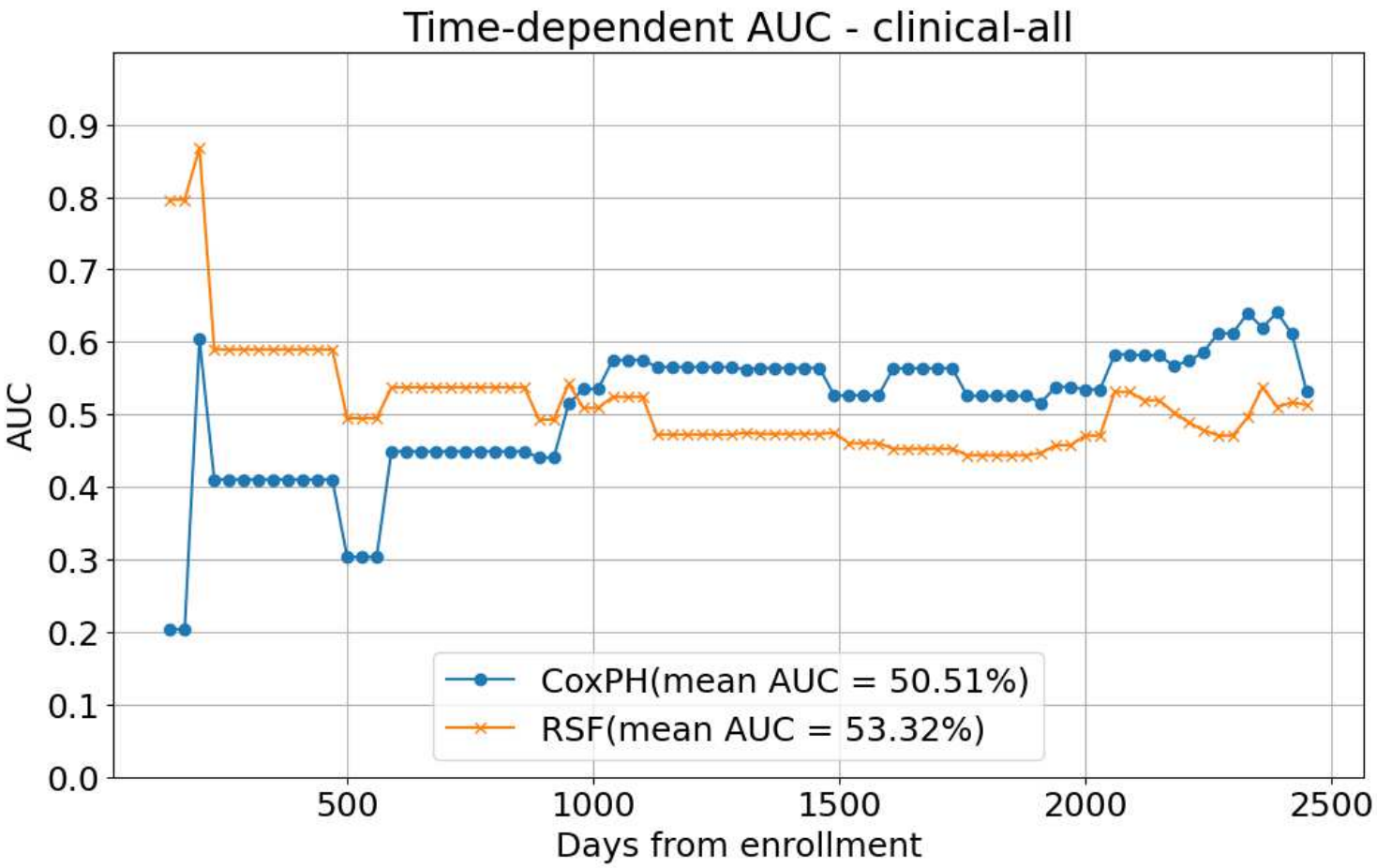

Figure 5

Time-dependent AUC curves with clinical data

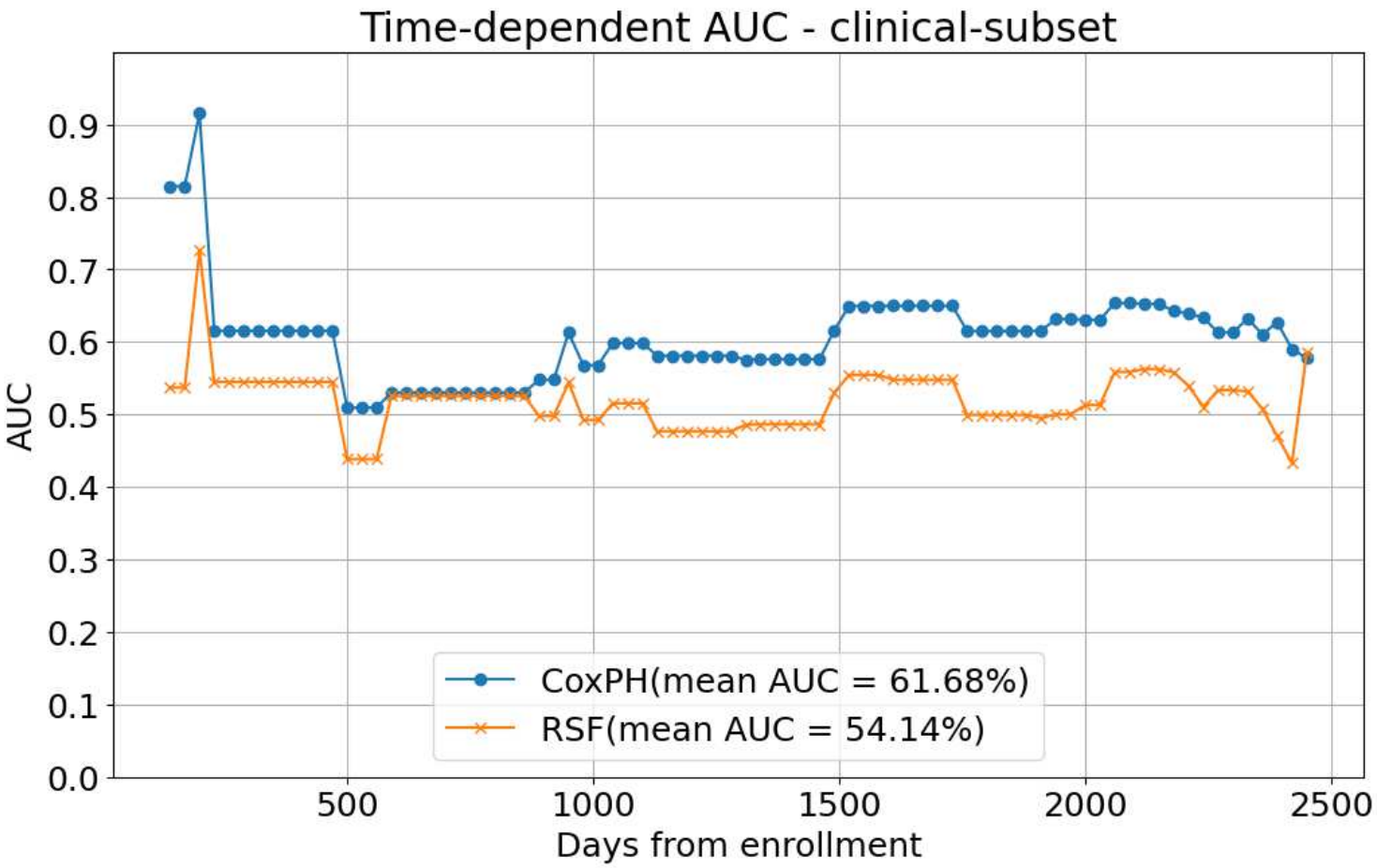

Figure 6

Time-dependent AUC curves with a subset of clinical data

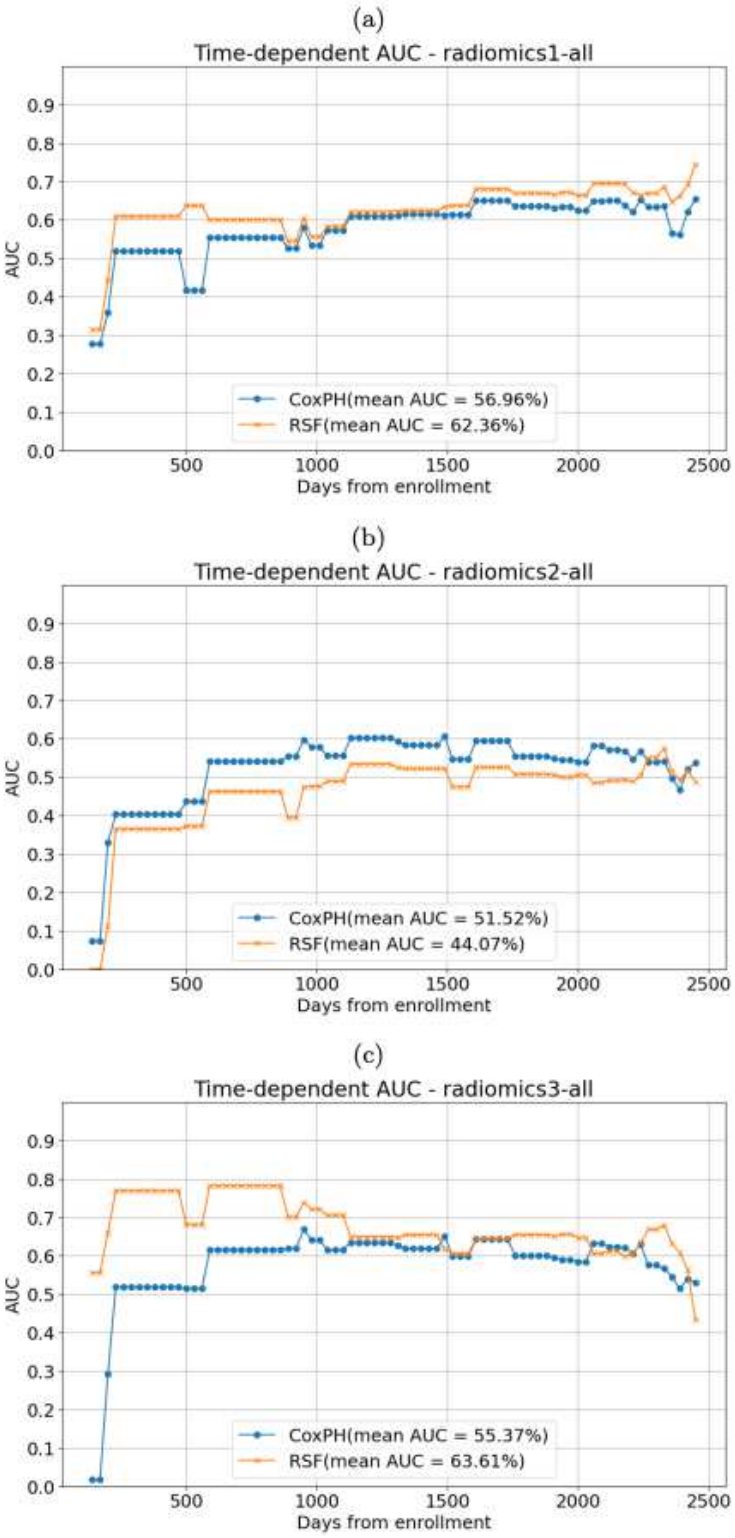

Figure 7

TD-AUC for radiomic features

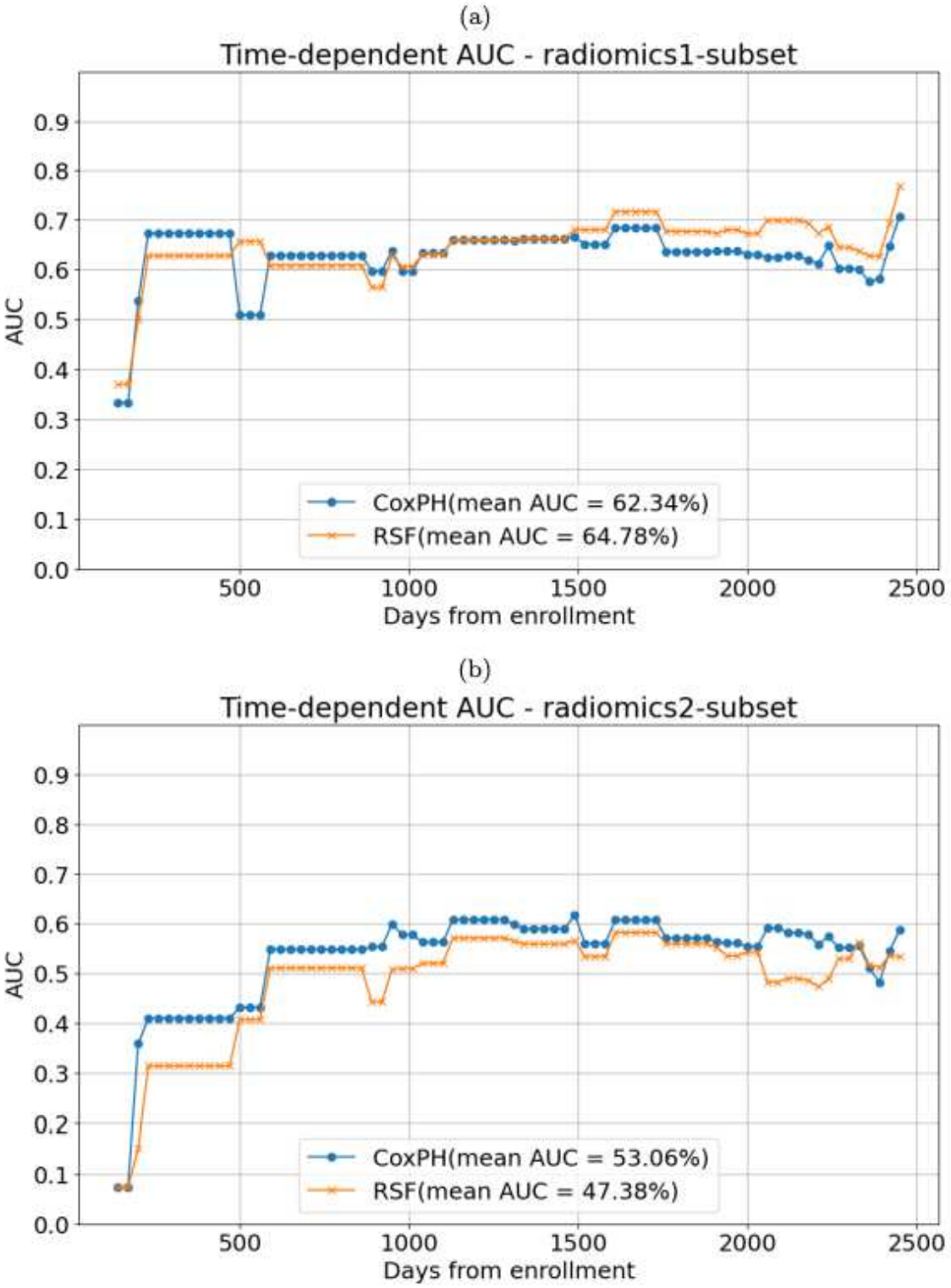

Figure 8

TD-AUC for subsets of radiomic features

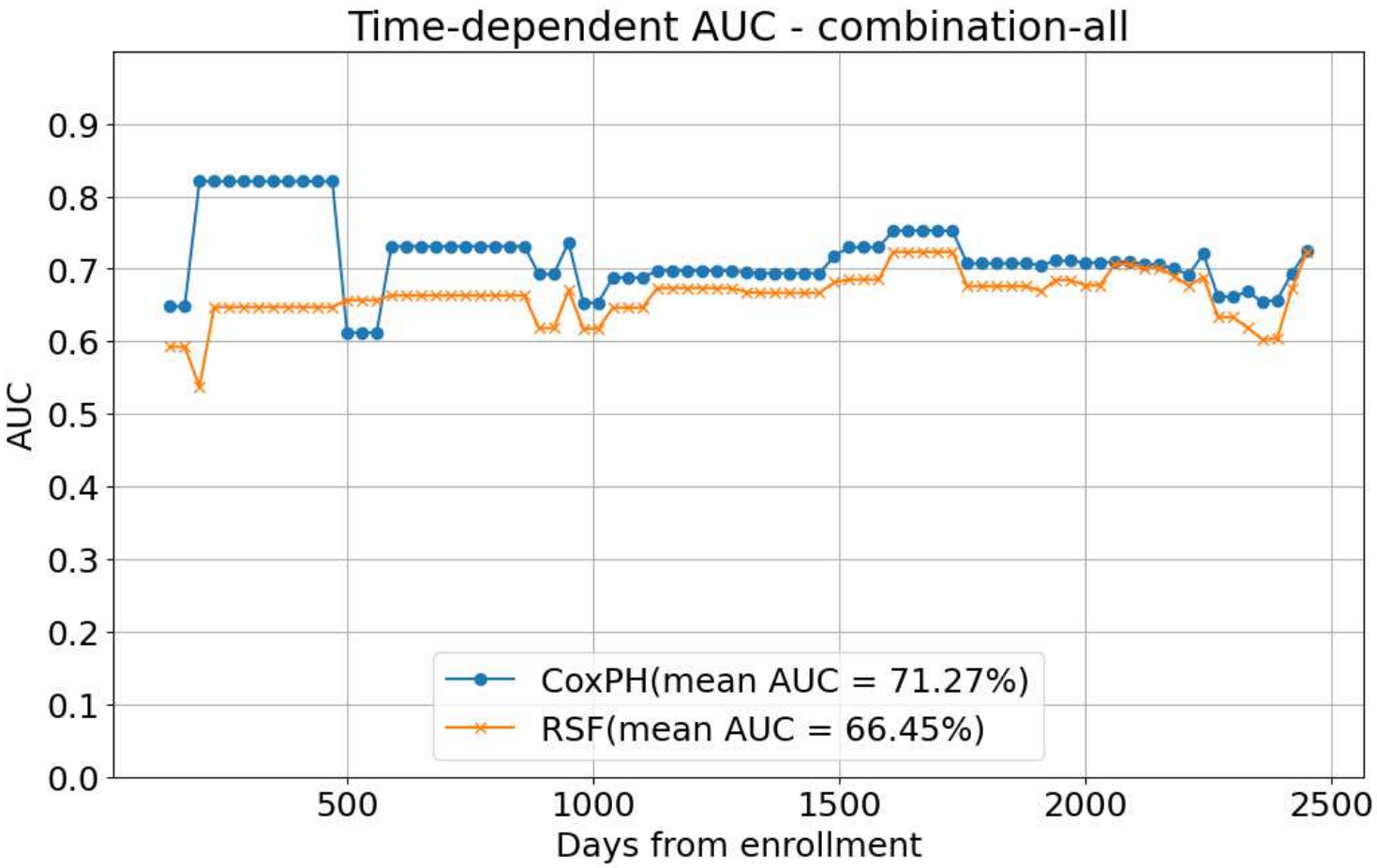

Figure 9

TD-AUC for subsets of combination of features