Supplementary II:

Section 1. Study population and overall experimental design

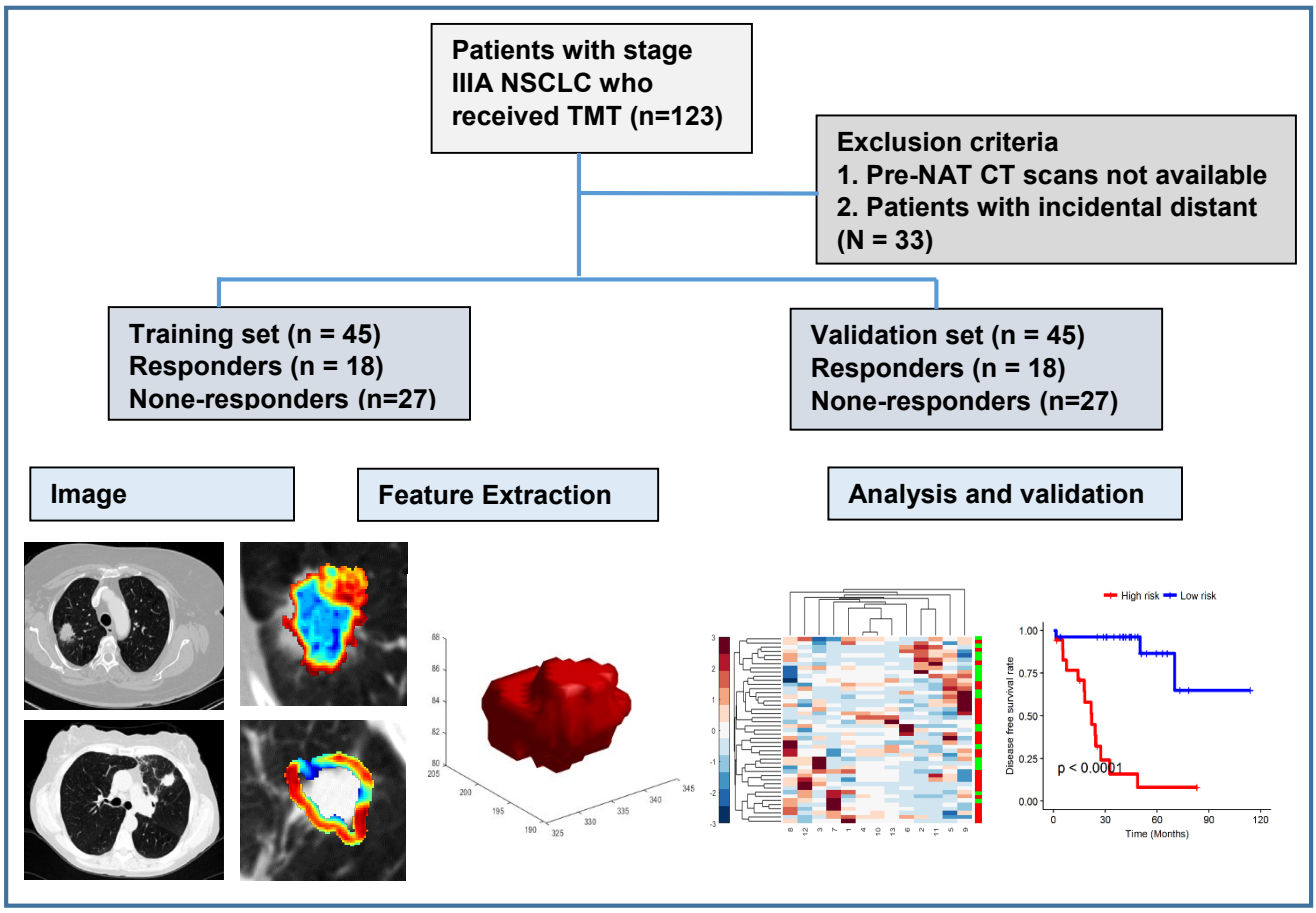


Figure S.1. Flowchart of study population with patient inclusion and exclusion criteria

Section 2. Image segmentation

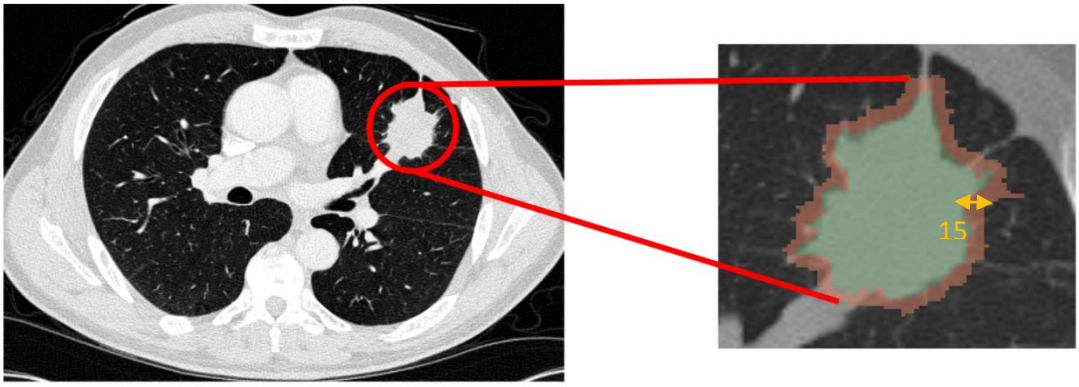


Figure S.2. The delineation of the intra-tumoral (green) and peritumoral (red) compartments of the nodule.

Section 3. Comparing peritumoral mask nearby normal and cancerous regions

We did peritumoral calculation mask nearby normal structures for central tumors. The nodule-adjacent peritumoral values were significantly different ($p < 0.05$) between the responder and non-responder tumor. Interestingly, there was no significant difference between the peritumoral values of the non-tumoral normal regions taken from responder patients (normal_R) and normal regions taken from non-responder patients (normal_N). The p-value of the t-test significance analysis between cancerous-responders and normal_R was 0.0012 and it was 0.0064 between cancerous-non-responders and normal_N. Figure S.3 illustrates a nodule (center panel), a 2D peritumoral of nodule (right panel) as well as 2D peritumoral in a corresponding normal region (left panel). None of the normal peritumoral features could discriminate responder from non-responder patients in the training and testing sets in conjunction with multivariable logistic regression (MLR) classifier.

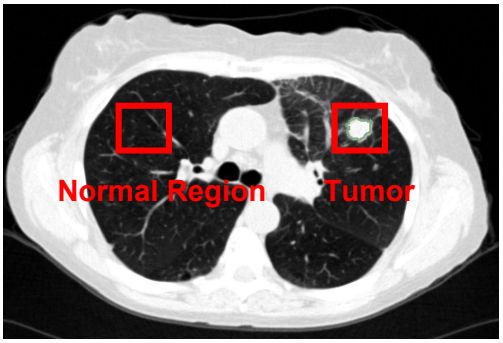


Figure S.3. A tumoral region (right panel including a nodule) and corresponding non-tumoral normal region (left panel without a nodule).

Section 4. The Dice index statistic to evaluate the nodule contour segmentation agreement between the two radiologists

The Dice values between two readers are averaged and then plotted.

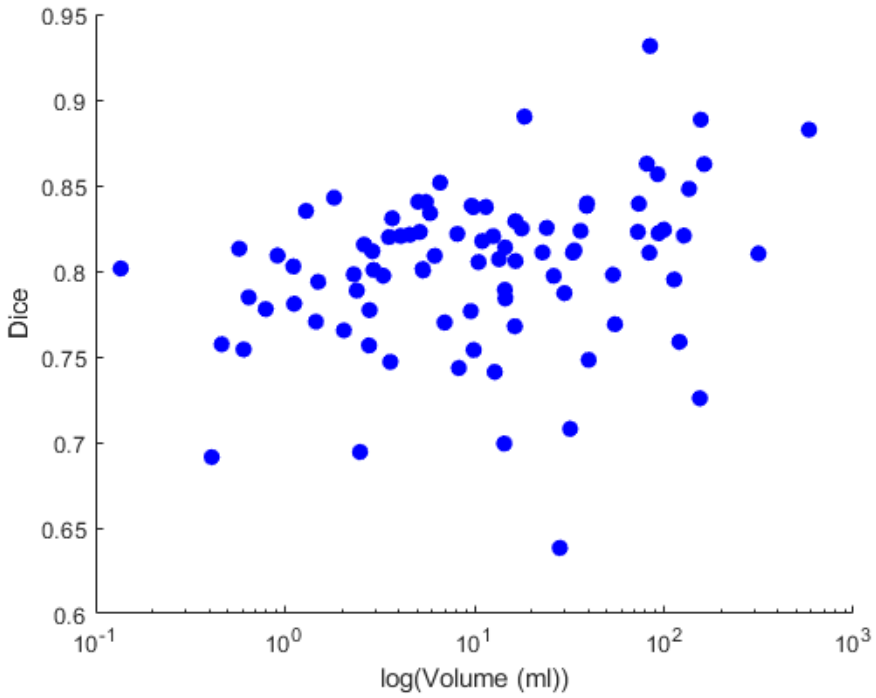
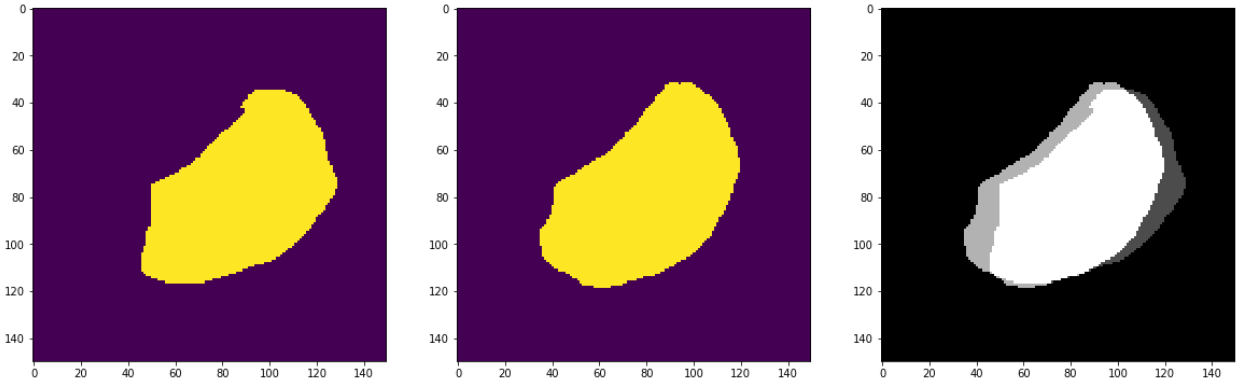


Figure S.4.A The inter-readers Dice values vs. logarithmic scale of volume for 90 studies.

Moreover, the average surface distance was used to compare different segmentation methods according to the pixel surface distance.

$$ASD = \frac{1}{(|S(A)| + |S(B)|)}\left(\sum_{S_A \in S(A)} d^2(S_A, S(B) + \sum_{S_B \in S(B)} d^2(S_B, S(A)\right)$$



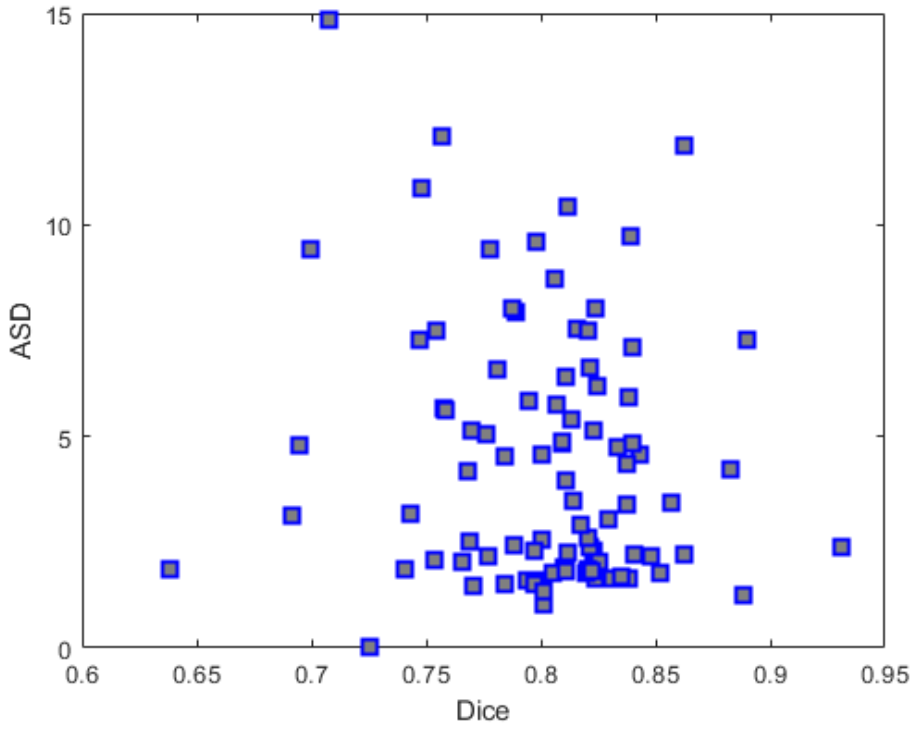
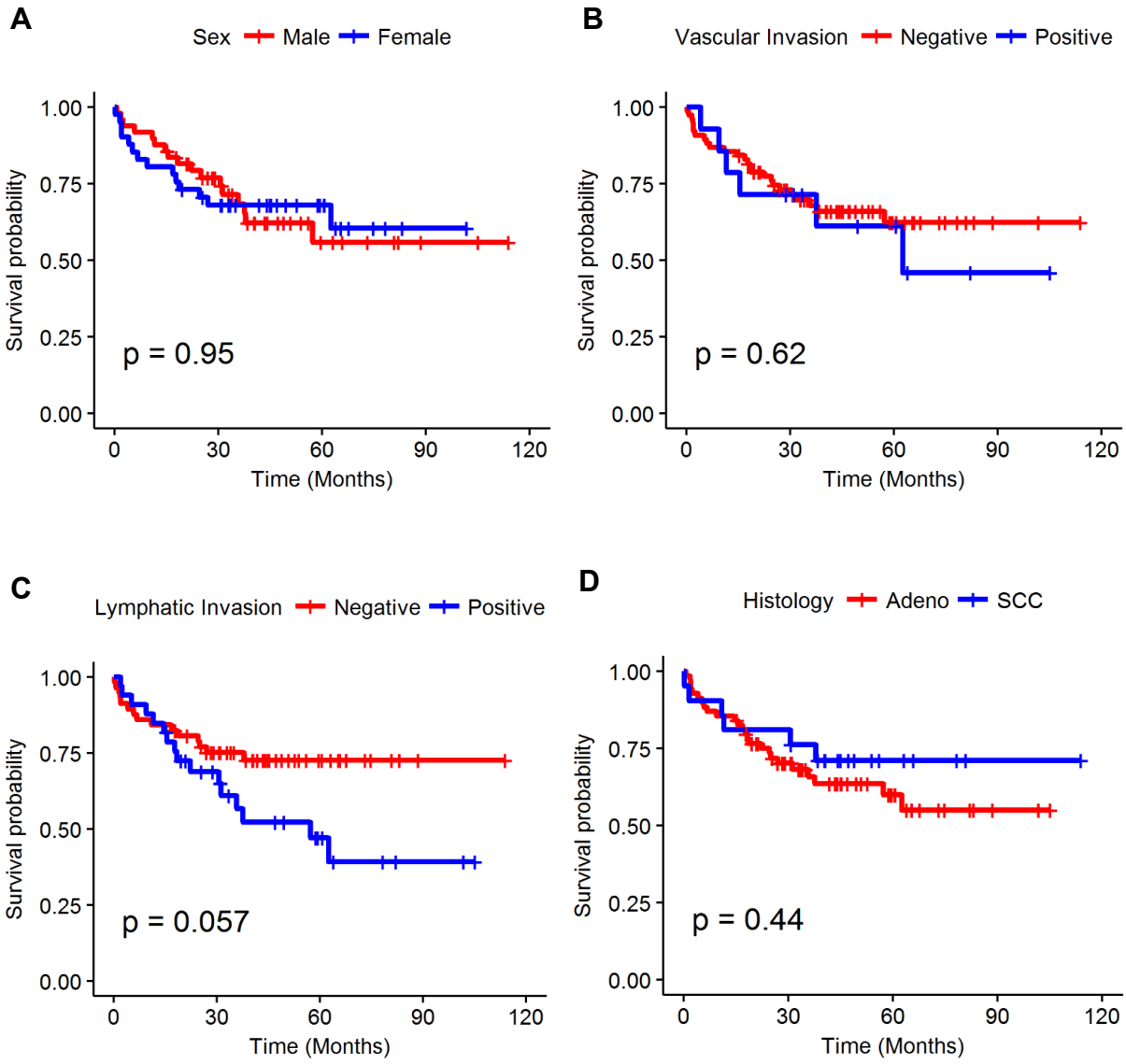


Figure S.4.B The inter-readers average surface distance vs. Dice values for 90 studies.

Section 5. Univariate Analysis Results: Kaplan-Meier curves for the clinical outcomes (overall survival) are presented. None of the clinical-pathologic features were significant for the Kaplan Meier (Log-rank test).



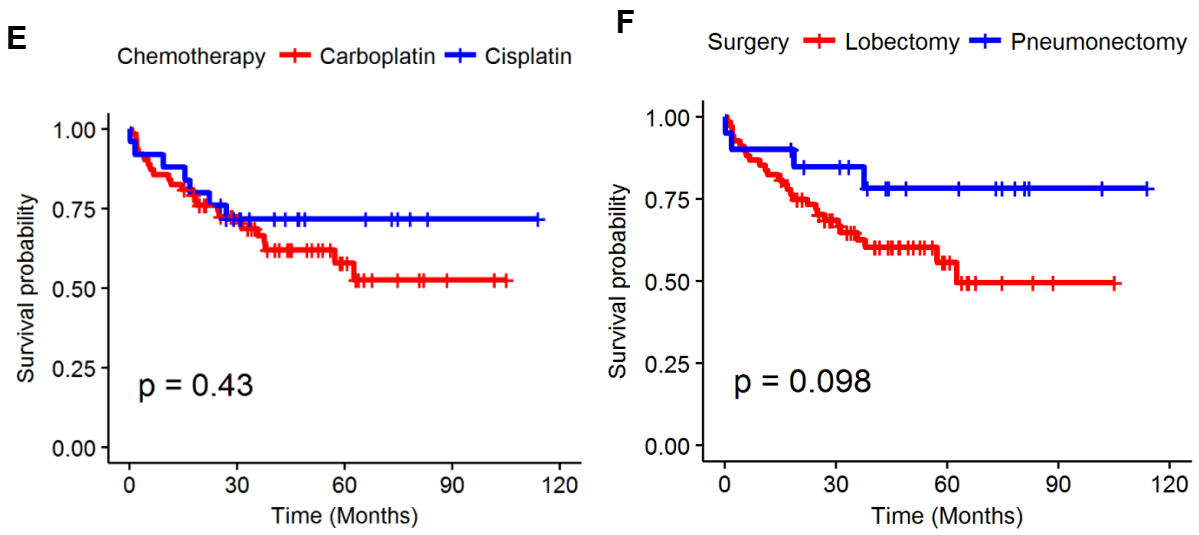


Figure S.5. Kaplan-Meier curves for (A) Sex (male vs. female; p-value = 0.95) (B) Vascular invasion (negative vs. positive; p-value = 0.62) (C) Lymphatic invasion (negative vs. positive; p-value = 0.057) (D) Histology (adenocarcinoma vs. squamous; p-value = 0.44) (E) Chemotherapy (Carboplatin vs. Cisplatin; p-value = 0.43) and (D) surgery type (Lobectomy vs. Pneumonectomy; p-value = 0.098).

Section 6. The effect of image reconstruction convolution filters

In this section, classifier performance, OS and DFS rate as a function of different convolution filters on CT image reconstruction will be evaluated. First, we identified that from 90 cases, 71 cases had two CT scans with different convolution kernels (sharp and standard). Basically, what was done in the main manuscript was based on standard convolution kernel which provide more informative radiomics features. Here, we explored a new experiment to see the effect of sharp convolution kernels on distinguishing responders and non-responders as well as OS and DFS. Standard convolution kernels operate like a low pass filter on image reconstruction, smooth the image and cancel noise and provide more informative radiomics features and increase the performance of classifier. We call this set of patients with sharp convolution kernels as D2 set.

In order to test our hypothesis, we extracted intratumoral and peritumoral texture features from CT images with sharp convolution kernel. The MLR classifier trained with the previous training set (standard filters) could discriminate responders and non-responders with AUC of 0.78 in D2 data set. The radiomic risk-score was also generated for OS and DFS by using Cox regression backward elimination. The radiomics signature was associated with OS (HR: 9.6, 95% CI: 3.74 – 38.2, p-value = 0.012) and DFS (HR = 3.24, 95% CI = 1.14, 3.56; p-value = 0.0088) in the D2 data set. The median split of radiomic risk-score was also used to stratify patients to low and high risk groups. Figure S.6.A shows the Kaplan Meier curve for patients in D2 with low and high risk groups of OS based on median risk-score (log-rank test, p-value < 0.0001, D2

= 71) and Figure S.6.B shows the Kaplan Meier curve for those patients for disease free survival rate based (p-value < 0.0001, D2 = 71)

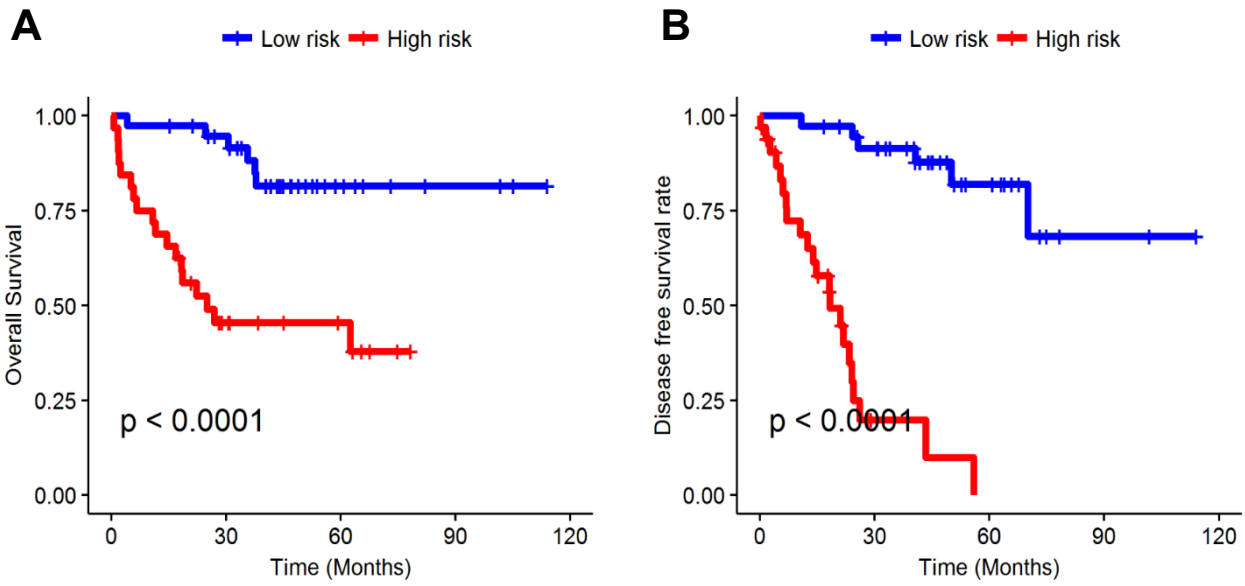


Figure S.6. (A) OS and (B) DFS rate for patients with low and high risk (D2 = 71).