Integration of Deep Learning Radiomics and Counts of Circulating Tumor Cells Improves Prediction of Outcomes of Early Stage NSCLC Patients Treated With SBRT

Zhicheng Jiao, PhD,* Hongming Li, PhD,* Ying Xiao, PhD,[#] Jay Dorsey, MD, PhD,[#] Charles B. Simone, II, MD,[$] Steven Feigenberg, MD,[#] Gary Kao, MD, PhD,[#] and Yong Fan, PhD*

*Department of Radiology, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, Pennsylvania;

[#]Department of Radiation Oncology, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, Pennsylvania;

[$]New York Proton Center, New York, New York; and Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York

Supplementary Data

Demographic and summary of clinical measures

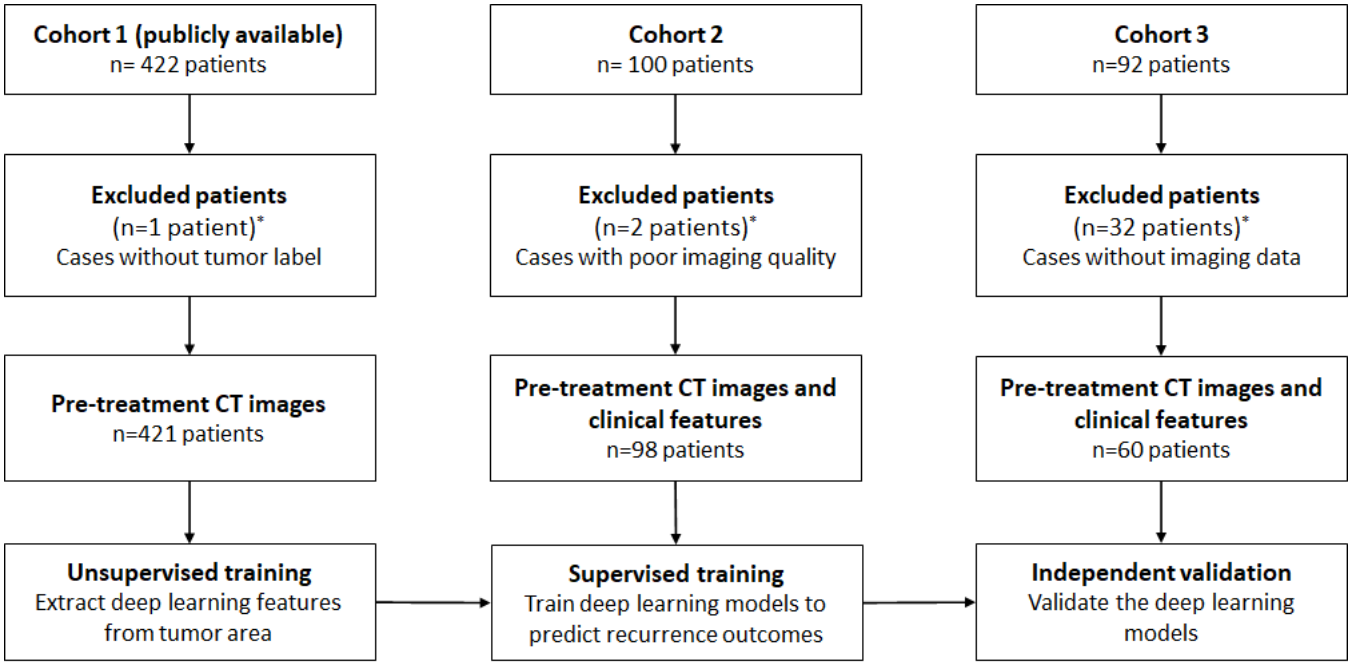
Supplementary Table 1. Characteristics of patients in the first cohort.

Characteristics	The first cohort (n=421)
Age (Years)	68.0±10.1
Sex (M/F)	69%/31%
Deceased	58%
Mean tumor size ± Standard Deviation, cm	3.17±1.52
Mean, median follow-up time (Month)	17.9, 13.5
Range of follow-up time (Month)	[0.3, 72.2]

Supplementary Table 2. Characteristics of patients in the second and third cohorts.

Characteristics	The 2nd cohort (n=98)	The 3rd cohort (n=60)	p value
Age (Years)	70.4±11.8	71.7±9.1	0.483
Sex (M/F)	48%/52%	38%/62%	0.240
Race (White/Others)	68%/32%	75%/25%	0.376
T stage (1a/1b/2a)	59%/30%/11%	65%/22%/13%	0.749
Current or former smoker	96%	97%	0.813
BMI ± Standard Deviation	27.3±6.4	26.5±5.8	0.393
Deceased	39%	22%	< 0.001
Local Failure	7%	18%	0.312
Nodal Failure	17%	18%	0.317
Distant Failure	9%	17%	0.939
Mean tumor size ± Standard Deviation, cm^3	1.98±0.88	1.77±0.80	0.131
Mean, median follow-up time (Month)	19.1, 16.6	28.6, 25.0	< 0.001
Range of follow-up time (Month)	[2.1, 47.8]	[7, 54.4]	
Median of prescription dose (Gy)	50	50	
BED10 (Gy)	100/112.5 (5/4 fractions)	100/112.5 (5/4 fractions)	

Data retrieval and exclusion



Supplementary Figure 1. Flow chart of study data retrieval and exclusion. One patient was excluded from Cohort 1 and the remaining patients were used to build a deep learning model for learning features; two patients were excluded from Cohort 2 and the remaining patients were used to train deep learning models for predicting recurrence, and 32 patients were excluded from Cohort 3 and the remaining patients were used to validate the deep learning models.

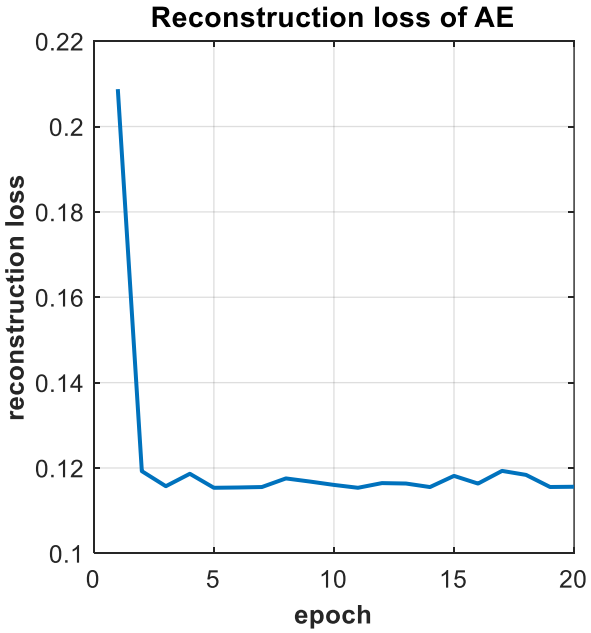
Data preprocessing and parameter settings of our deep learning (DL) models

CT images and tumor masks (GTVs) in all datasets were resampled into isotropic voxels of unit dimension (1 mm) to ensure comparability, where 1 voxel corresponds to a cubic of 1 mm^3 using linear and nearest neighbor interpolations, respectively. CT image intensities were then discretized using equally spaced bins with a bin-width of 25 Hounsfield Units and normalized to a 0-1 range. This discretization step not only reduces image noise, but also normalizes intensities across all patients, allowing for a direct comparison of all calculated textural features between patients. CT image patches (zero outside the GTV) with a size of 50 × 50 × 50 were extracted from CT scans around manually labeled GTV and were normalized to have values in range of 0 to 1 as input to our DL model. The image patch size was large enough to cover GTV of all tumors in the present study. Our DL models are based on the toolkit of PyTorch (version 1.3.0).

Supplementary Table 3. Parameters of our AE model.

Layer	*Num of filters*	*Kernel size*
Conv1	32	3x3x3
Conv2	64	3x3x3
Conv3	128	3x3x3
Conv4	64	3x3x3
Conv5	32	3x3x3
Conv6	1	3x3x3

The convolutional autoencoder (AE) model contained four 3D convolutional (Conv) layers to encode 3D CT patches to deep features followed by three 3D convolutional layers with up-sampling blocks to reconstruct the input patches from deep features. Loss function of this AE model was mean square error (MSE). A dropout layer with a drop rate of 0.5 following each 3D convolutional layer was adopted as a regularization method to reduce overfitting and improve generalization of the convolutional AE. The AE model's network parameters are summarized in Supplementary Table 3. The AE network was trained via Adam optimizer with batch size set to 4, learning rate set to 1e-03, and the number of epochs set to 20. The model obtained at the last epoch was used in all our experiments. Training loss of the AE model is shown as a function of the number of epochs in Supplementary Figure 2.

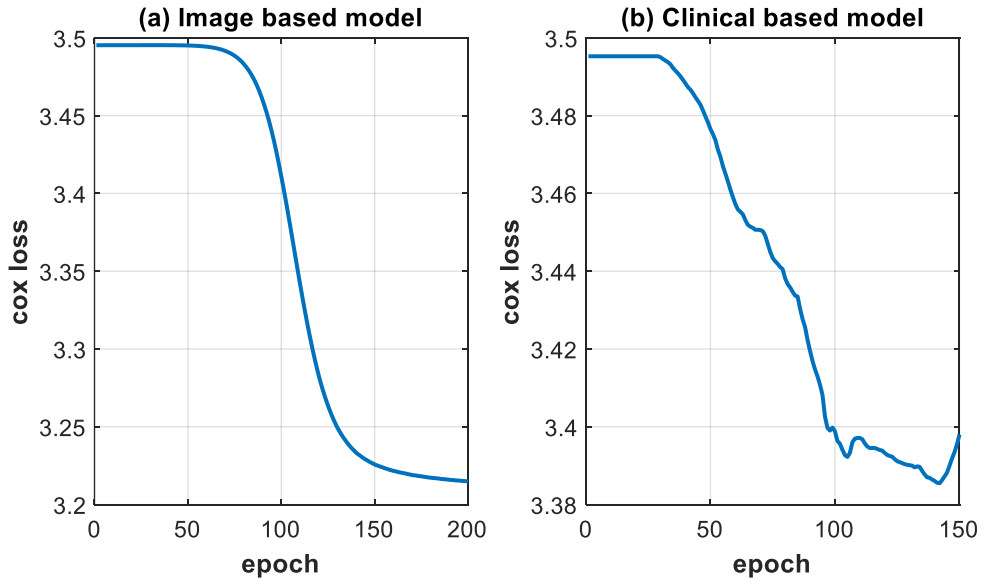


Supplementary Figure 2. Reconstruction loss of AE as a function of the number epochs.

Supplementary Table 4. Parameters of our recurrence prediction model (image based).

Layer	*Num of filters*	*Kernel size*
Conv1	32	3x3x3
Conv2	64	3x3x3
Conv3	128	3x3x3
FC1	32	NA
FC2	32	NA
FC3	1	NA

The DL models for predicting recurrences based on tumor images consisted of the same encoder layers (Conv1 to Conv3) of the AE model, an average pooling layer to obtain the visual features, and three additional fully connected (FC) layers, with their parameters specified in Supplementary Table 4. Weights of the encoder layers were fixed to have the same weights as the trained AE model to extract image features in a transfer learning setting, and the FC layers were trained to predict the recurrence risk scores by optimizing a cox proportional hazards loss (1,2). Again, a dropout layer with a drop rate of 0.5 was adopted following each FC layer to further improve the robustness. The recurrence prediction model was trained via Adam optimizer, with batch size set to the scale of 98, learning rate set to 1e-04, and the number of epochs set to 200. The model obtained at the epoch of minimized training loss was used in all our experiments. Training loss of the image-based recurrence prediction model is shown as a function of the number of epochs in Supplementary Figure 3 (a).



Supplementary Figure 3. Training loss of the DL models for predicting recurrences based on tumor images and clinical features.

Supplementary Table 5. Parameters of our recurrence prediction models (clinical based).

Layer	*Num of filters*
FC1	16
FC2	16
FC3	1

The DL model for predicting recurrences based on clinical measures consisted of three FC layers, with its parameters specified in Supplementary Table 5. The DL model was trained to predict the recurrence risk scores by optimizing a cox proportional hazards loss (1,2). The recurrence prediction model was trained via Adam optimizer,

with batch size set to the scale of 98, learning rate set to 1e-02, and the number of epochs set to 150. The model obtained at the epoch of minimized training loss was used in all our experiments. Training loss of the clinical feature based DL model is shown as a function of the number of epochs in Supplementary Figure 3 (b).

We also fit Cox regression models (3) based on different combinations of our DL based recurrence risks, CTC measures, and clinical measures to evaluate the recurrence prediction performance on the third cohort. Specifically, the DL based recurrence risks are the combined prediction of image-based and clinical based model, the CTC measures are binary values of negative and positive while the clinical measures are the same as that used in the clinical based DL models. These models were fit and evaluated on the 60-patient set and the recurrence prediction performance is summarized as Supplementary Table 6.

Supplementary Table 6. Prediction performance (c-index) of recurrence prediction models of image-based DL recurrence risks (DL), clinical measures (Clin), and CTC status and the results of cox regression model based on the combination of them (DL + Clin + CTC).

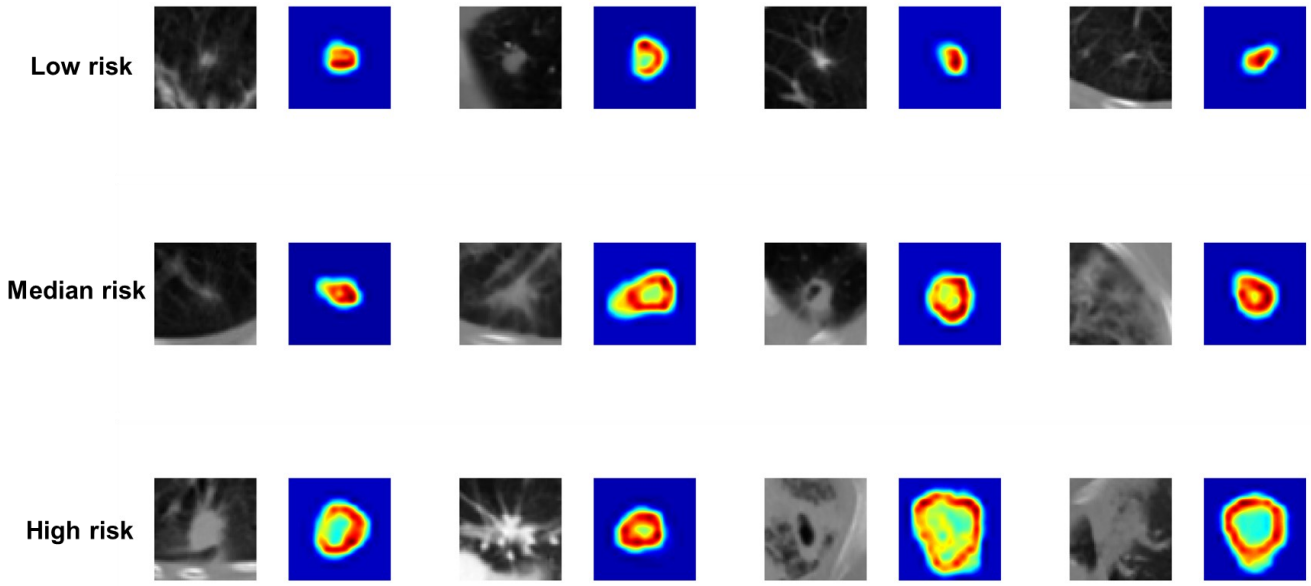
Included measures	*c-index*
DL	0.852
Clin	0.767
CTC	0.591
DL + Clin + CTC	0.870

We also compared the pooled hazard rate with each group-specific hazard rate (4) on the third cohort for investigating the association between the recurrence time of patients and a set of predictor variables, including DL model based recurrence risk scores, tumor size, clinical measures, SUV, and CTC measures. Particularly, the recurrence risk scores based on image-based DL model (DL risks) was obtained by applying the DL recurrence prediction models to individual patients of the third cohort. The λ^2 and p values of all these included are shown in Supplementary Tables 7.

Supplementary Table 7. Test statistic λ^2 and p value of each included measure in stratifying recurrence groups.

Variables	***λ^2***	***p value***
DL risks	15.1457	9.9524e-05
Tumor size	14.6205	0.0001
Age	0.1921	0.6612
Sex	0.0190	0.8904
Smoking status	0.6866	0.4073
BMI	3.4748	0.0623
SUV	0.5530	0.4571
CTC	4.0284	0.0447

Inspired by the widely used deep learning visualization method of class activation map (CAM) (5), for our image based DL model, we applied the weighted sum of the presence of tumors' visual patterns at different spatial locations by using the learned recurence network weights to form the visual attention maps of patients. Values in these heatmaps represent the levels of attention that our DL model payed to different tumor regions. Some instances of visual attention maps of low, median, and high recurrence risk patients are shown as Supplementary Figure 4.



Supplementary Figure 4. Attention visualization of our image-based DL model. Images in the rows from top to bottom are tumor images and the related attention maps of low, median, and high risk groups.

References:

1. Li H, Boimel P, Janopaul-Naylor J, et al. Deep convolutional neural networks for imaging data based survival analysis of rectal cancer. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* 2019:846-849.
2. Li H, Zhong H, Boimel PJ, et al. Deep convolutional neural networks for imaging based survival analysis of rectal cancer patients. *International Journal of Radiation Oncology, Biology, Physics* 2017;99:S183.
3. Breslow N. Covariance analysis of censored survival data. *Biometrics* 1974;30:89-99.
4. Fleming TR, Harrington DP. A class of hypothesis tests for one and two sample censored survival data. *Communications in Statistics - Theory and Methods* 1981;10:763-794.
5. Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. pp. 2921-2929.