Iowa Initiative for Artificial Intelligence

Final Report

Project title:	**Validation of predicted tumor recurrence lesions of deep learning-segmentation based prognostication (DESEP) model**
Principal Investigator:	Yusung Kim, Ph.D (PI),
Prepared by (IIAI):	Through our collaborative group (Drs. Stephen Baek, Xiaodong Wu and Yusune He)
Other investigators:	Xiaodong Wu, Ph.D., Stephen Baek, Ph.D., John Buatti, M.D., Bryan Allen, M.D.,Ph.D., Brian Smith, Ph.D., Jordan Gainey, M.D.
Date:	1/22/2020
Were specific aims fulfilled:	Y
Readiness for extramural proposal?	N (but in preparation)
If yes … Planned submission date	June, 2021
Funding agency	NIH
Grant mechanism	U01
If no … Why not? What went wrong?	A grant proposal is under-preparation with the aim of submitting it at June, 2021. Nothing went wrong but we need more data based upon what we found through this pilot study.

Brief summary of accomplished results: All 3D high-risk map prediction results were validated for tumor recurrence lesions when a deep learning segmentation-based prognostication (DESEP) model is used on FDG PET/CT images for NSCLC patients treated with stereotactic body radiotherapy (SBRT). Total 36 recurrent patients were identified, and a total of 61 recurrent tumors have been analyzed. We found the 3D predicted recurrent region matched to real recurrent tumor region with only 20% - 29% accuracy even though we had found a very promising case before this pilot study. Developing the current AI algorithm (DESEP) model is beyond the

scope of this pilot grant, requiring much more commitment that could not be covered by the allowed budget. Thus, we have focused on the 1-D prediction of DESEP based upon what we found from the validation and analysis of all 61 recurrent tumors. We thoroughly analyzed the capability of DESEP prediction power in 1-D (i.e., yes or no) for progressive disease, local progression free survival (LPFS), overall (OS), and disease specific survival (DSS) when compared to current tumor response assessment standard (RECIST v1.1). We found there was a high concordance between DESEP-predicted LPFS risk category and manually calculated RECISTv1.1 (φ=0.544, p=0.001). Neither the auto-segmentation based volumetric RECIST nor the computer-based unidimensional RECIST correlated with RECISTv1.1 (p=0.081 and p=0.144, respectively). While RECISTv1.1 correlated with LPFS (HR=6.97,3.51-13.85,p<0.001), it could not provide insight regarding DSS (p=0.942) or OS (p=0.662). In contrast, the DESEP-predicted LPFS methods were predictive of LPFS (HR=3.58, 1.66-7.18, c=0.90, p<0.001), OS (HR=6.31, 3.65-10.93, c=0.83, p<0.001) and DSS (HR=9.25, 4.50-19.02, c=0.76, p<0.001). Based upon what we discovered, we concluded deep-learning segmentation-based prognostication can predict LPFS as well as OS, and DSS after SBRT for NSCLC. It can be used in conjunction with RECISTv1.1 to provide additional insights regarding DSS and OS in NSCLC patients receiving SBRT. The discovery founded in this pilot study will be used as preliminary data for the upcoming NIH U01 grant proposal with the expected submission date of June, 2021.

Research report:

Aims (provided by PI): "Validate the 3D high-risk map prediction accuracy for tumor recurrence lesions when a deep learning segmentation-based prognostication (DESEP) model is used on FDG PET/CT images for NSCLC patients treated with stereotactic body radiotherapy (SBRT)"

Our group successfully completed the validation of all available datasets of recurrent tumors. After completing the aim of this study, we went even further steps of exploring the prediction power of DESEP model. What we discovered from performing the aim of this pilot study, the current high-risk map of DESEP is premature to accurately predict the recurrent regions in 3D space. In addition, we also discovered 3-D spatial prediction of tumor recurrence is very challenging problem for which extensive studies need to be performed. For instance, the prediction of DESEP model for the tumor size changes and progression needs to be validated first throughout all follow up images, and the DESEP model needs to be further developed

accordingly. Current DESEP model is only trained on the pre-treatment images. It is expected the prediction power of recurrent tumor will be improved when the DESEP model is trained throughout SBRT and follow up CT images after SBRT as well as pre-treatment images.

Data:

As patient characteristics (**Table 1**), a total of 108 subjects were analyzed retrospectively following approval from the University of Iowa Institutional Review Board (IRB: 200503706). All patients provided consent for the use of their clinical information and medical images and signed an informed consent form approved by the Institutional Review Board. All data collection and experimental procedures are in accordance with relevant guidelines and regulations. All patients underwent SBRT for NSCLC with treatments ranging from July 2006 to October 2018. Target volumes were delineated by radiation oncologists using both CT and PET imaging, and contouring was completed using Velocity AI (Varian Medical System, Inc., Palo Alto, CA). Following SBRT, patients were followed with surveillance CT images at approximately 2 months following SBRT then every 3 months thereafter. There were a total of 51 male and 57 female patients represented in this study. There were 55 patients with adenocarcinoma, 41 with squamous cell carcinoma, 12 adenosquamous, 1 with metastasis from previous NSCLC, and 9 without a biopsy. The patients' prognostic stage varied and included 67 patients with stage I, 6 patients with stage II, 21 patients with stage III, and 14 patients with stage IV disease. Overall survival, disease specific survival, and local progression free survival were defined from the start of SBRT. By the end of the study, 72 patients had experienced local progression or death, 58 patients had died, 40 of those deaths were cancer-related.

Patient Population	
Age at SBRT	72.0 years (± 9.7 years)
Mean Overall Survival	1.97 years (± 1.44 years)
Mean Disease Specific Survival	1.82 years (±1.37 years)

Prior Treatment	
Previous Radiation Therapy	33 (30.6%)
Previous Surgery	31 (28.7%)
Previous Chemotherapy	41 (38.0%)

Mean Local Progression Free Survival	1.58 years (± 1.33 years)
Male	51 (47.2%)
Female	57 (52.8%)
Survival	
Alive	50 (46.3%)
Dead	58 (53.7%)
Cause of Death	
Alive	50 (46.3%)
Cancer-related	40 (37.0%)
Not Cancer-related	11 (10.2%)
Unknown	7 (6.5%)
Local Progression Free Survival	
Survived without Progression	36 (33.3%)
Progression or Death	72 (66.7%)
2 Year Overall Survival	
Survived	50 (46.3%)
Died	42 (38.9%)
Insufficient Follow-up	16 (14.8%)
2 Year Disease Specific Survival	
Survived	57 (52.8%)
Died	35 (32.4%)
Insufficient Follow-up	16 (14.8%)
2 Year Local Progression Free Survival	
Survived	28 (25.9%)
Died/Progressed	64 (59.3%)
Insufficient Follow-up	16 (14.8%)
Previous Immunotherapy	2 (1.9%)
Stage	
IA	51 (47.2%)
IB	16 (14.8%)
IIA	3 (2.8%)
IIB	3 (2.8%)
IIIA	5 (4.6%)
IIIB	16 (14.8%)
IV	14 (13.0%)
Karnofsky Performance Score	
100	5 (4.6%)
90	25 (23.1%)
80	40 (37.0%)
70	30 (27.8%)
60	6 (5.6%)
<60	2 (1.9%)
Histology	
Adenocarcinoma	55 (50.9%)
Squamous Cell Carcinoma	41 (38.0%)
Adenosquamous	2 (1.9%)
Metastasis from Prior NSCLC	1 (0.9%)
Clinical Diagnosis	9 (8.3%)

Table 1: Patient demographics and clinical characteristics. Continuous data are presented in the form mean (±standard deviation), discrete data are presented in the form number (percentage).

AI/ML Approach:

In this study, we used a 3D segmentation algorithm using a U-Net architecture that previously developed. The architecture has an “hourglass” structure which extracts imaging features at varying levels of granularity. The input of the CT Segmentation U-Net is a cropped 3D CT image measuring 96 x 96 x 48 mm^3 with the tumor located in the center of the image. The target output of the U-Net is a segmentation mask trained on the ground truth of our study which is a binary mask map defined by three radiation oncologists’ contours of the gross tumor volume aggregated by the STAPLE algorithm. Through training the segmentation U-Net, we had achieved over 75% of segmentation accuracy measured by dice similarity coefficient in the previous study that was performed before this pilot study. As the U-Net segments the tumor region, it has also encoded a large amount of image features (textural, geometric and radiomic features) at the “bottleneck” layer which are critical to predict a binary segmentation map. These encoded features contain rich information about the tumor shape and texture that may be correlated with survival, cancer progression, as well as tumor recurrence. We performed an unsupervised feature selection by applying the k-medoids clustering method to cluster the U-Net features into a reduced number of representative features (i.e. medoids of the clusters). Then, we use least absolute shrinkage and selection operator to identify features exhibiting strong correlations with the survival outcomes. Using these DESEP features, we were able to generate predictions associated with a low or high risk for overall survival (DESEP-predicted OS), disease-specific survival (DESEP-predicted DSS), and local progression free survival (DESEP-predicted LPFS).

To visualize the high-risk regions, an activation maximization scheme was employed to visualize the LASSO-selected U-Net features. For a trained U-Net encoder X = q(·|W, b), neurons at the bottleneck layer corresponding to the LASSO-selected features were denoted as qi. Then, the equation was solved for each individual neuron via gradient ascent: X + = X + γ ∇q (X) where X(k) is the current solution at k-th iteration and γ (k) is a step length. We set γ (k) as 1/σ (k) where σ (k) denotes the standard deviation of the gradients. The gradient ∇qi was computed using the standard backpropagation algorithm. The initial image X(0) was initialized with random voxel values following the Gaussian distribution. We visualized a risk map by evaluating each voxel's contribution to the prediction of survival. We employed a guided backpropagation approach. For each voxel in the input image, with marginal change of the survival probability with respect to the voxel's intensity, defined as $\frac{\partial P}{\partial x_{i,j,k}}$, where P is the probability of death and xi,j,k is a voxel value at the position (i, j, k). In the guided backpropagation process, we rectified the gradient by dropping the negative gradient values to focus on the "risk". This was achieved by applying

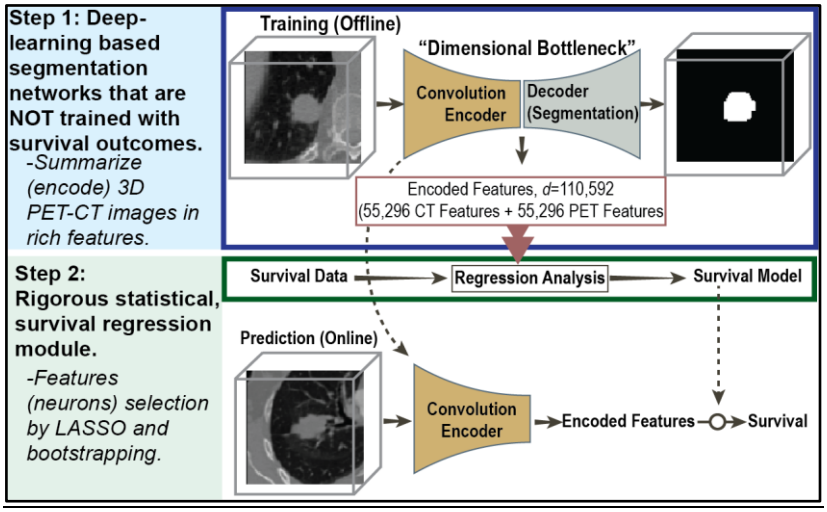


Figure 1. Deep learning-based segmentation model consists of two-independent-steps. Deep learning-based segmentation that is not trained with survival outcomes (blue box) and statistical, survival regression model (green box) where the latent radiomic features are analyzed for survival prediction.

rectified linear unit (ReLU) activation when the values were backpropagated from node to node:

$$\alpha^{(m)} = \max\left(\frac{\partial(P)}{\partial A_{i,j,k}^{(m)}}, 0\right)$$ where A(m) denotes the activation map corresponding to the m-th convolutional kernel at the bottleneck encoding. Note that only the LASSO-selected features were involved in the survival model P such that $\frac{\partial P}{\partial x_{i,j,k}}$ is zero most of the time. Finally, the risk map $\mathcal{R}$ was defined as a linear combination of all activation maps at the bottleneck layer with the coefficients α(m) obtained from the above:

$$\mathcal{R}(\mathbf{X}) = \sum_m \alpha^{(m)} A^{(m)}(\mathbf{X})$$

Experimental methods, validation approach:

To validate the accuracy of the 3D high-risk map of DESEP model, we followed the same experimental methods and validation approach that developed through the previous study (**Figure 1**). The key innovative approach of the DESEP model resides in the unique two-step architecture as demonstrated in **Figure 1**: Step 1 focuses on learning purely geometric characteristics, developed through the automated tumor segmentation task while Step 2 focuses on learning to correlate those independently-learned geometric features to clinical endpoints. To visualize which lesions in the patient images predicted high risk in terms of 2-year OS (**Figure 1**), we employed a guided gradient backpropagation approach that detailed in the section of AI/ML approach. **Figure 1** presents the gradient which named as high-risk maps. Heated regions (red) are the areas that lowered the probability of 2-year whereas the other area (blue) are the ones that had negligible effect on the survival. Throughout this pilot study, the rigorous quantitative validation and characterization of high-risk maps of DESEP model over full 108 NSCLC SBRT patient datasets has been performed. The preliminary 108 patient data include three (33%) or four (67%) follow up post-CT within 2 year after SBRT treatment. The prediction performance of the high-risk maps was fully characterized over the actual recurrent lesions. The series of high-risk maps (images) on pre-treatment CT have been generated for all 96 patients per their clinical end points (2-year OS). In this study, each high-risk map for each 2-year OS was registered to the post-CT image where actual tumor recurrence was identified by radiation oncologists based upon radiologists' medical records and follow-up CT images. The accuracy of predicted high-risk maps

over actual recurrent lesions as well as pre-treatment tumor volumes were measured. The optimal visualization threshold of high-risk map was investigated with balancing the sensitivity and specificity in its prediction performance. However, due to the nature of pilot grant with limited budget, full development and optimization of high-risk maps was beyond the scope of this study.

RECISTv1.1 criteria were utilized to categorize treatment response on follow-up CT imaging. Measurements were taken of the target lesion along the largest tumor diameter. Progression of disease (PD) was determined based on a 20% or greater increase in the diameter relative to the smallest of previously measured diameters with a minimum absolute increase of at least 5mm. A complete response (CR) was defined as a disappearance of the target lesion. A partial response (PR) was defined as a 30% or greater decrease in target lesion summed diameters relative to its baseline pre-treatment measurement. A lesion was categorized as stable disease (SD) if it did not meet any of the previous criteria.

The deep-learning based auto-segmentation model was trained to segment the tumor volume on each follow-up CT scan. From this segmentation, the total volume and largest tumor diameters were calculated. Each follow-up scan was assigned a category ranging from complete response to progression of disease based on the calculated tumor volume (auto-segmentation based volumetric RECIST) using ellipsoid volumetric thresholds and the calculated tumor diameter (computer-based unidimensional RECIST) using standard thresholds. Using this method, the patient's final categorization was defined as the worst category received on any one follow-up CT image which were obtained 2 months after completion of SBRT then every 3 months thereafter.

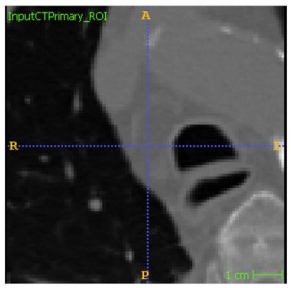
All statistical analyses were performed using SPSS Statistics, Version 26.0 (IBM Corp. Armonk, NY) with a two-sided $\alpha=0.05$ used to establish statistical significance. The primary endpoints utilized in this study were local progression free survival (LPFS), disease-specific survival (DSS), and overall survival (OS) which were all defined from the start of SBRT. Local progression was defined as having a RECISTv1.1 categorization of progressive disease at the location of the treated target lesion as measured by the physician using the largest tumor diameter. Data for RECISTv1.1 categorization were dichotomized with a distinction drawn between progression of disease versus any other category indicating non-progressive disease. Data for survival prediction were produced as a continuous probability ranging from zero to one which was then dichotomized into a low-risk group and high-risk group based on a cut-off at a 50% predicted probability of an event within 2 years after SBRT.

Correlation between dichotomous variables was established using Cramer's Phi which can be interpreted similarly to a correlation coefficient with a value of one indicating a perfect agreement between two variables.18 Survival curves within RECIST v1.1 and DESEP prediction-based categories were estimated with the method of Kaplan-Meier and compared statistically with log-rank tests.19 Survival differences between categories were estimated with hazard ratios (HR) obtained from Cox regression. For surviving patients, their information was censored at the date of last follow-up. Model predictions were evaluated at a time point of 2 years after SBRT and analyzed using a receiver operating characteristic curve to calculate the c-statistic

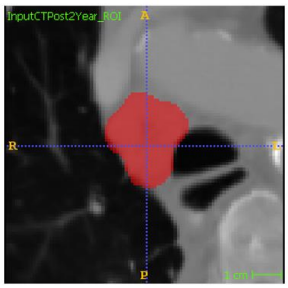
Results:

The accuracy of tumor recurrence prediction was 20 %. The below images are some of selected results. We tested different window levels of high-risk maps, optimizing the window levels. However, the window levels are all patient-specific. We could not improve the prediction accuracy of risk-map by changing window levels. We concluded the current DESEP model is limited to predict the recurrence in 3D space (**Figure 2**).

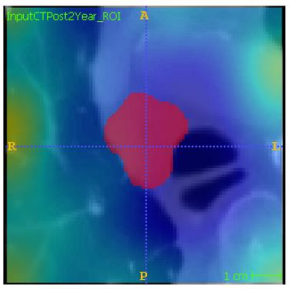
No Primary tumor



Initial CT ROI

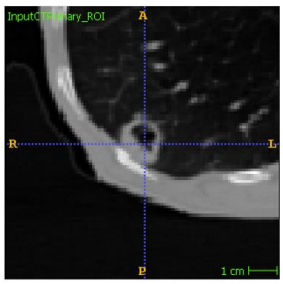


Recurrent CT ROI

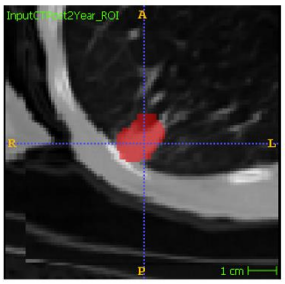


Overlay with Recurrence CT

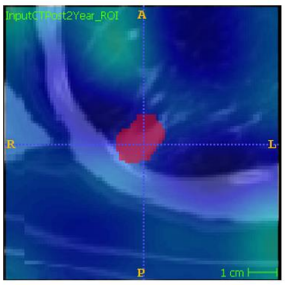
No Primary tumor



Initial CT ROI

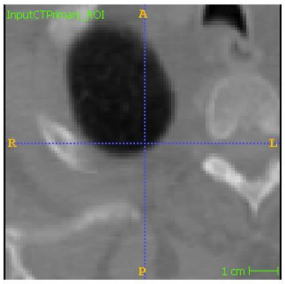


Recurrent CT ROI

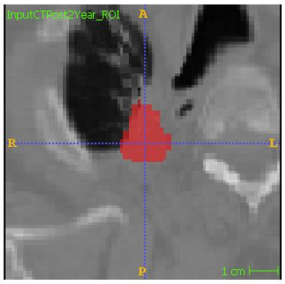


Overlay with Recurrence CT

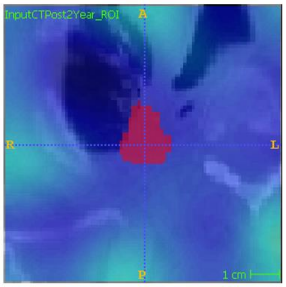
No Primary tumor



Initial CT ROI

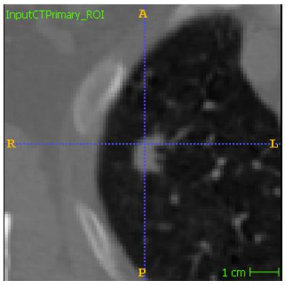


Recurrent CT ROI

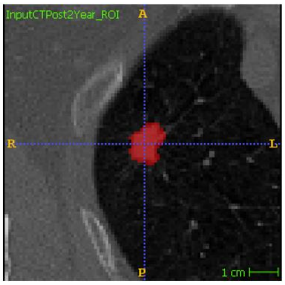


Overlay with Recurrence CT

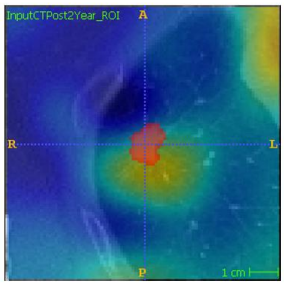
No Primary tumor



Initial CT ROI

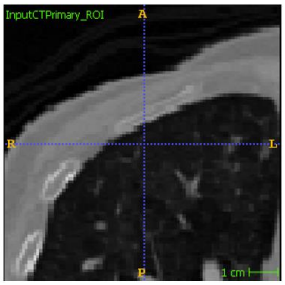


Recurrent CT ROI

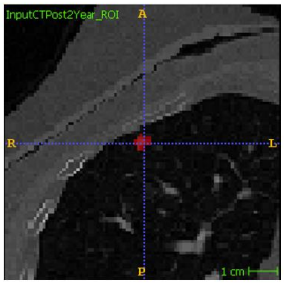


Overlay with Recurrence CT

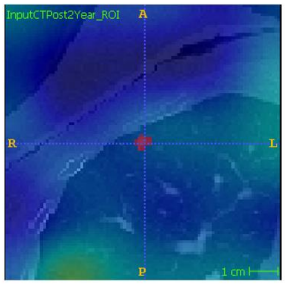
No Primary tumor



Initial CT ROI

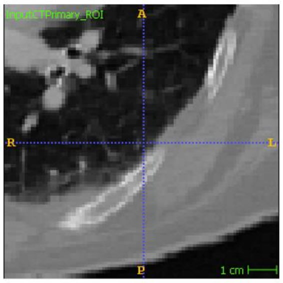


Recurrent CT ROI

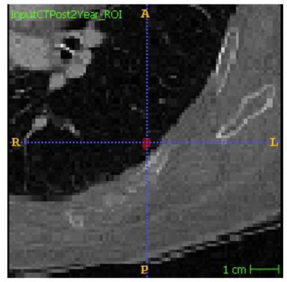


Overlay with Recurrence CT

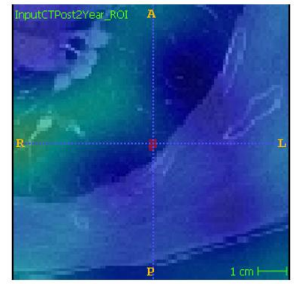
No Primary tumor



Initial CT ROI

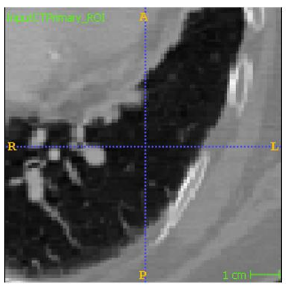


Recurrent CT ROI

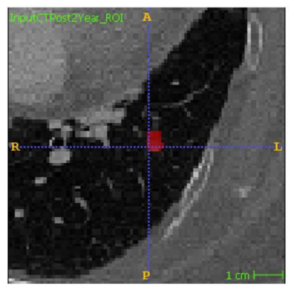


Overlay with Recurrence CT

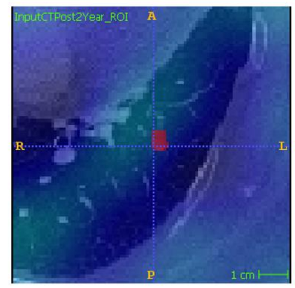
No Primary tumor



Initial CT ROI

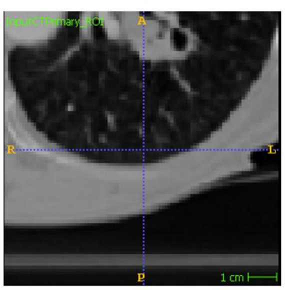


Recurrent CT ROI

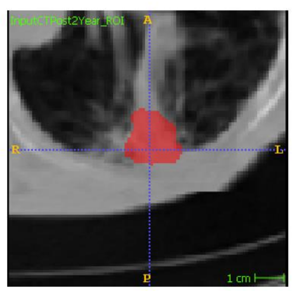


Overlay with Recurrence CT

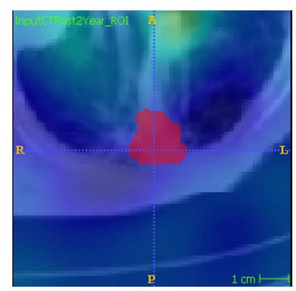
No Primary tumor



Initial CT ROI

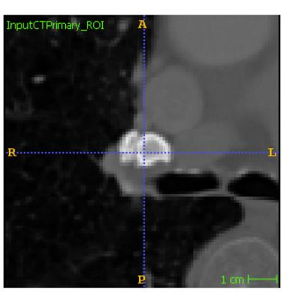


Recurrent CT ROI

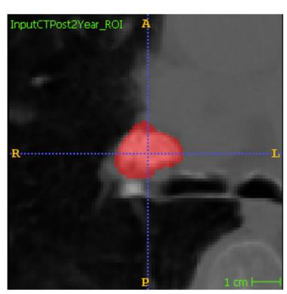


Overlay with Recurrence CT

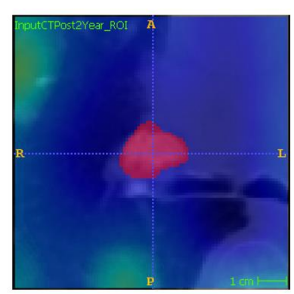
No Primary tumor



Initial CT ROI



Recurrent CT ROI



Overlay with Recurrence CT

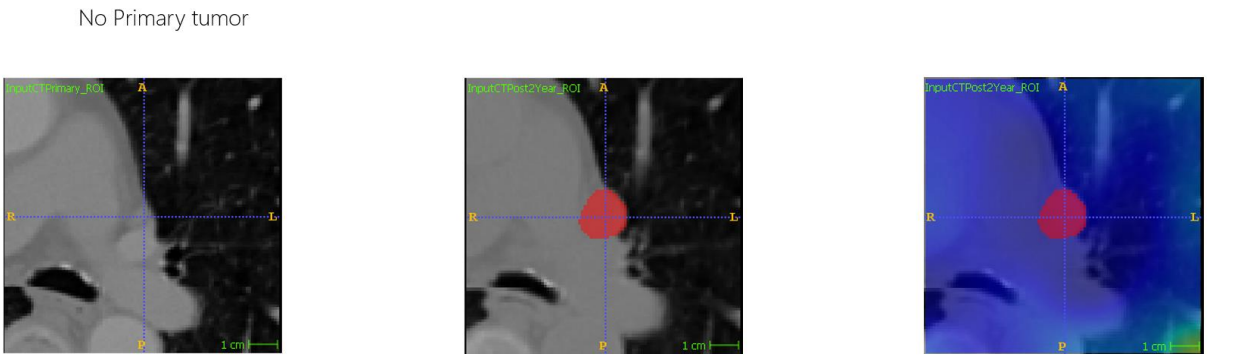


Figure 2. The comparison of high-risk map (right) over the recurrent tumor (center) for the selected cases.

This is the overall accuracy results. No overlapping cases (not generating high-risk signals on the recurrent tumors) were recorded as 60%. About 20% cases present partial overlapping, while 20% cases presented the high-risk signals (orange or higher color) were found within the recurrent tumors. Based upon this 3D prediction, we revisited the prediction power of progression. We found its accuracy was 0.755, sensitivity 0.793, and specificity 0.701 which are considerably lower than our findings on its prediction power for OS and DSS. Progression disease means the tumor is not responding to the high level radiation, and keeps growing. However, tumors typically shrink down first and grow later, thus, it seems we need to validate our conclusion through well-controlled datasets which have enough follow ups such as longer than 2 years. Also, our limited dataset size (108 cases) can be part of our poor prognostic result.

Overall overlapping ratio:

- No: 60%
- Partial: 20%
- Yes: 20%

The poor prediction power of local progression was confirmed when we generated a K-M curve by using overall survival.

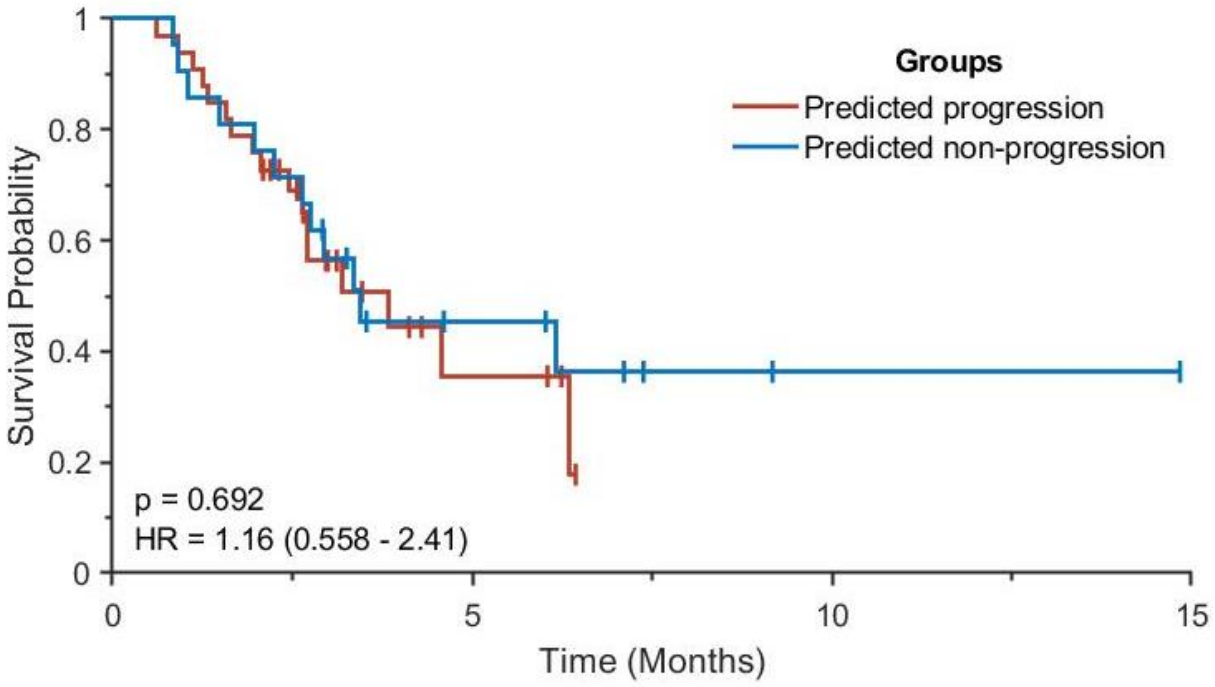


Figure 3. The Kaplan-Meier curve for the two groups that classified by DESEP model; one is predicted local progression disease group and another is not. No statistical difference was found which means the power of differentiating two groups was not efficient.

We also validated the prediction accuracy for the recurrence (1-D, yes or no), since we had difficulties in predicting 3D location. We obtained its prediction accuracy of 0.75, sensitivity 0.86, and specificity 0.61. It was obtained trough data augmentation (i.e. total 92 datasets were increased to 1012). In addition, the prediction power for the progressive disease was found as the accuracy of 0.71, the sensitivity 0.83, and specificity 0.53.

After completion of the aim of this pilot study, our group further investigated the prediction power of LPFS, OS, and DSS. The method which had the highest agreement with the manually measured RECISTv1.1 was the DESEP-predicted LPFS method which extracted features associated with worse local progression (φ=0.544, p=0.001). There was a reduced agreement with RECISTv1.1 categorizations when only utilizing the auto-segmentation based volumetric RECIST method (φ=0.227, p=0.081) or when using the computer-based unidimensional RECIST method (φ=0.184, p=0.144).

Kaplan-Meier curves were generated to estimate differences in LPFS, these curves are presented in **Figure 4**. Having progression of disease by RECISTv1.1 was associated with worse LPFS, (HR=6.97, 3.51-13.85, p<0.001). Similarly, having a DESEP-predicted high risk for local

progression (DESEP-predicted LPFS) was associated with a worse LPFS, (HR=3.58, 1.66-7.18, p<0.001). Utilizing the auto-segmentation model to simply calculate the pre-treatment tumor volume (HR=1.35, 0.79-2.32, p=0.271) or tumor diameter (HR=0.95, 0.50-1.81, p=0.772) did not show a statistically significant association with LPFS. When evaluating the predictive power at a time point of 2 years after completion of SBRT using a receiver operating characteristic curve, DESEP-predicted LPFS demonstrated a c-statistic of c=0.90 (0.83–0.97).

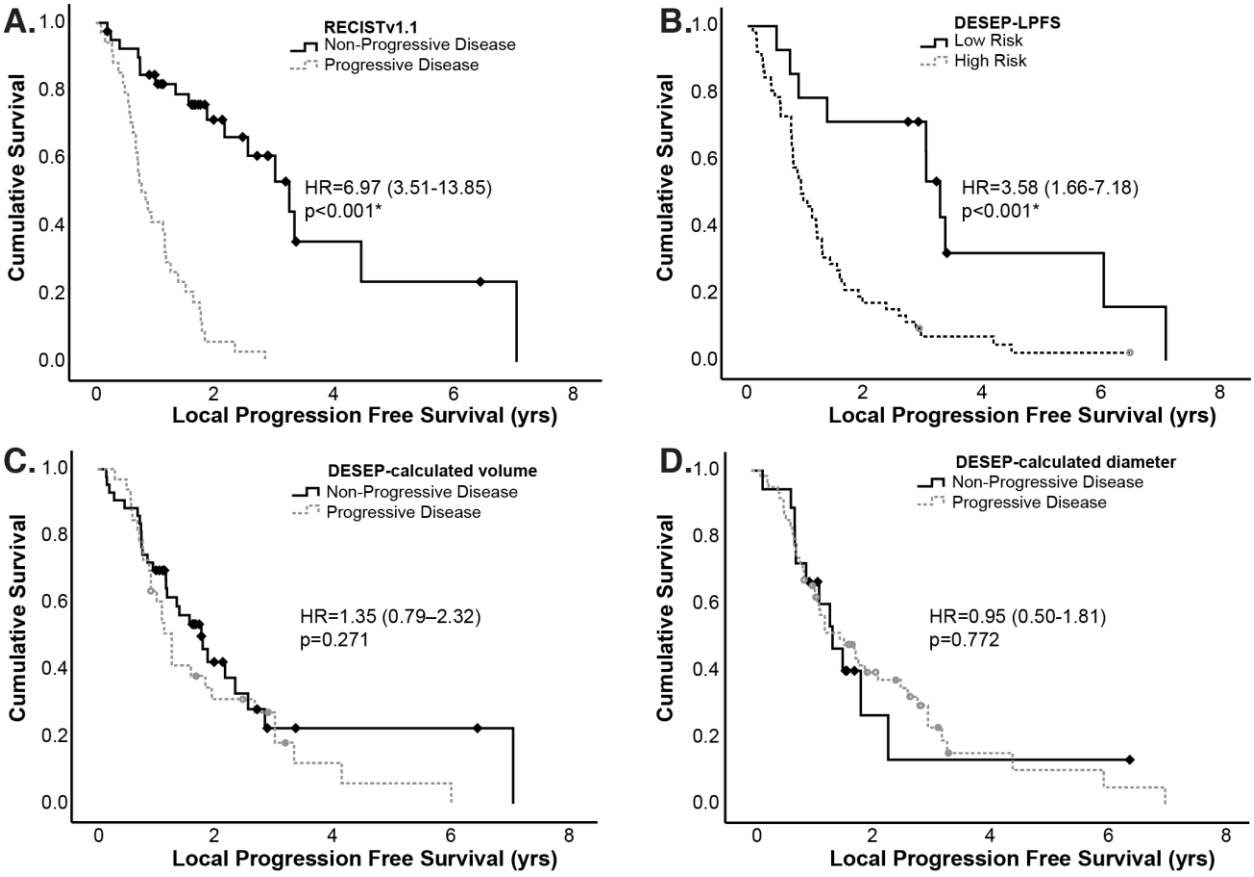


Figure 4: Kaplan-Meier curves generated examining local progression free survival of high and low risk groups identified using (A) RECISTv1.1, (B) DESEP-predicted LPFS, (C) auto-segmentation-based volumetric RECIST (DESEP-calculated volume), and (D) computer-based unidimensional RECIST (DESEP-calculated diameter). RECISTv1.1, auto-segmentation-based volumetric RECIST, and computer-based unidimensional RECIST made serial measurements on multiple surveillance images to categorize patients as having progression of disease vs. non-progressive disease. DESEP-predicted LPFS extracts radiomic features associated with local progression of disease. Comparisons between groups were made using a log-rank test.

Kaplan-Meier curves were generated using both the dichotomized RECISTv1.1 and the DESEP predictions to estimate differences in OS and DSS; these curves are presented in **Figure 5** and **Table 2**. RECISTv1.1 was unable to discriminate patients on the basis of OS (HR=1.16, 0.60-2.26, p=0.662) or on the basis of DSS (HR=0.97, 0.41-2.29, p=0.942). DESEP-predicted OS performed well when discriminating OS (HR=6.31, 3.65-10.93, p<0.001). The mean OS time was 3.60 years (±0.33 years) in the group with a predicted low risk for death compared to 1.03 years (±0.18 years) in the high-risk group. DESEP-predicted DSS performed similarly well with DSS predictions (HR=9.25, 4.50-19.02, p<0.001). The mean disease specific survival time was 4.15 years (±0.41 years) compared to 0.84 years (±0.11 years) in the low risk and high risk groups respectively. DESEP-predicted OS and DESEP-predicted DSS demonstrated a c-statistic of c=0.83 (0.73–0.93) and c=0.76 (0.64–0.88) respectively when evaluated at a time point of two years after completion of SBRT.

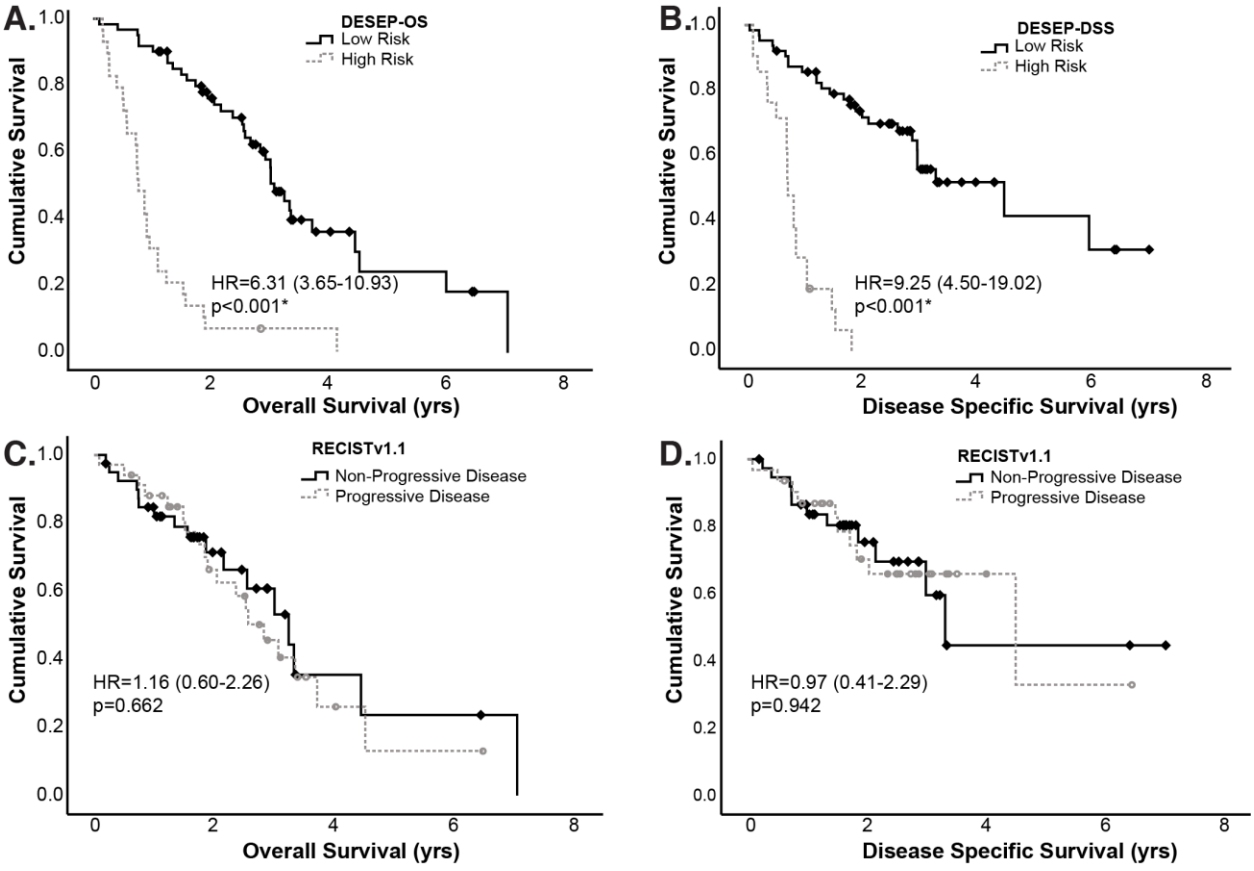


Figure 5: Kaplan-Meier curves examining the predictive power of DESEP-predicted categorizations for (A) overall survival (OS) and (B) disease specific survival (DSS). This is presented in comparison to RECISTv1.1 for (c) OS and (D) DSS. Both the DESEP-predicted OS

and DESEP-predicted DSS methods extracted radiomic features associated with OS and DSS, respectively. RECISTv1.1 method made serial measurements on multiple surveillance images to categorize patients as having progression of disease vs. non-progressive disease, Comparisons between groups were made using a log-rank test.

Overall Survival					
Method	Low Risk Group (Mean±SD)	High Risk Group (Mean±SD)	Hazard Ratio (95%CI)	C-statistic (95%CI)	p-value
DESEP-predicted OS	3.60 yrs (±0.33 yrs)	1.03 yrs (±0.18 yrs)	HR=6.31 (3.65-10.93)	0.83 (0.73-0.93)	<0.001*
RECISTv1.1	3.52 yrs (±0.55 yrs)	3.02 yrs (±0.39 yrs)	HR=1.16 (0.60-2.26)	N/A	0.662
Disease Specific Survival					
Method	Low Risk Group (Mean±SD)	High Risk Group (Mean±SD)	Hazard Ratio (95%CI)	C-statistic (95%CI)	p-value
DESEP-predicted DSS	4.15 yrs (±0.41 yrs)	0.84 yrs (±0.11 yrs)	HR=9.25 (4.50-19.02)	0.76 (0.64-0.88)	<0.001*
RECISTv1.1	4.30 yrs (±0.68 yrs)	4.04 yrs (±0.60 yrs)	HR=0.97 (0.41-2.29)	N/A	0.942
Local Progression Free Survival					
Method	Low Risk Group (Mean±SD)	High Risk Group (Mean±SD)	Hazard Ratio (95%CI)	C-statistic (95%CI)	p-value
DESEP-predicted LPFS	3.57 yrs (±0.69 yrs)	1.33 yrs (±0.18 yrs)	HR=3.58 (1.66-7.18)	0.90 (0.83-0.97)	<0.001*
RECISTv1.1	3.51 yrs (±0.55ys)	0.99 yrs (±0.11yrs)	HR=6.97 (3.51-13.85)	N/A	<0.001*

Table 2: Comparison of mean survival time of high risk and low risk groups for three primary endpoints of overall survival, disease specific survival, and local progression free survival. Using RECIST categorization, the high risk group was defined as having progression of disease while the low risk group was any other category indicating non-progressive disease. Comparisons between groups was performed using a log-rank test.

The current study provides evidence of the prognostic power of a deep-learning segmentation based model in patients with NSCLC treated with SBRT. The DESEP model may

be used with RECISTv1.1 criterion to determine a patient's risk for disease progression, overall survival, and disease specific survival.

Ideas/aims for future extramural project: As a multi-disciplinary research team, we (Drs. Xiaodong Wu, Stephen Baek, and myself at UI and Dr. Sanjay Anita at Yale) have been preparing the proposal of NIH U01 that is planned to be submitted in June, 2021. As co-investigators, Drs John Buatti, Bryan Allen, and Brian Smith will still contribute this research. Its specific aims are still under development. The current versions of the three specific aims are; 1) Optimize and validate weakly-supervised learning strategies to train a deep radiomics model. 2) Develop a novel federated learning method to minimize the domain shift problem. 3) Validate the efficacy of interpretability methods and Bayesian inference in clinical decision making. The underlying scientific premise for this grant proposal is that the current challenges in deep-learning based cancer research can be addressed by an interpretable, Bayesian deep learning model trained with weakly-supervised learning objectives optimized via federated learning. Weak supervision is an emerging idea allowing deep learning models to be trained from incomplete, inexact, and/or missing supervision, which evidently can reduce data annotation costs and thus increase the size of training data. Federated learning trains a model at each local data source in a distributed manner, which can effectively eliminate the impediments of data sharing. We hypothesize that a substantial synergy will be produced when weakly-supervised learning is combined with federated learning, by means of increased amount of qualified data, reduced overhead to produce local federated updates, and consequently, lowered barrier to join a federated learning network, ultimately leading to a more popular and broader use. Furthermore, the complexity and the low tractability of federated learning can be compensated by an interpretable, Bayesian network that has visually interpretable features, an explainable reasoning process, and quantifiable uncertainty. Our vision is to build a multi-institutional radiomics network, called RadiomiQ, building upon these three concepts, implemented in an easily scalable Python framework. Multiple institutions including University of Iowa, Yale, and Stanford, representing different geographical locations across the United States will participate. The study teams in these institutions have been collaborating through multiple capacities, including the NSF-funded federated learning project which will complement the current project.

Publications resulting from project: A manuscript is about to be submitted. The final draft has been reviewed by coauthors and has been updated. It is expected to be submitted within a month into the New England Journal of Medicine (impact factor 74.699 that is ten times higher than the most prestigious journal in Radiation Oncology field).