Title: (<200 characters including spaces)

Deep Learning Serial CT Response Score Predicts Overall Survival in Advanced NSCLC Treated with PD-(L)1 Immune Checkpoint Inhibitors

Authors:

Chiharu Sako, Ph.D.[1], Taly Gilat Schmidt, Ph.D. [1], Arpan A. Patel, M.D.[3], Dwight H. Owen, M.D.[4], Arya Amini, M.D.[5], Brendan D. Curti, M.D.[6], Ronan J. Kelly, M.D.[19], Ray D. Page D.O., Ph.D.[7], Aurelie Swalduz, M.D.[13], Jean-Paul Beregi, M.D.[14], Jan Chrusciel, M.D.[15], Stephane Sanchez, M.D., Ph.D[15], Richard Rosenberg, M.D. [16,17], Eric Snyder, B.S.[3], Jakob Weiss, M.D.[8], Andrew L. Wentland, M.D., Ph.D.[9], Anish Kirpalani, M.D.[10], An Liu, Ph.D.[5], Olivier Gevaert Ph.D.[11], George R. Simon M.D.[12,+], Hugo J.W.L. Aerts, Ph.D.[1] , Ravi B. Parikh, M.D., M.P.P.[18], Petr Jordan, Ph.D.[1]

[1]Onc.AI, [3]*University of Rochester Medical Center,* [4]*Ohio State University Comprehensive Cancer Center,* [5]*City of Hope,* [6]*Providence Cancer Institute,* [7]*The Center for Cancer and Blood Disorders,*[8]*Mass General Brigham,* [9]*University of Wisconsin School of Medicine and Public Health,* [10]*University of Toronto,* [11]*Stanford University,* [12]*Moffitt Cancer Center,* [13]*Centre Léon Bérard,* [14]*CHU de Nîmes,* [15]Centre Hospitalier de Troyes, [16]Medicom Technologies, [17]Arizona Oncology, [18]*University of Pennsylvania,* [19]*Baylor University Medical Center*

Abstract: (<400 words before references)

Background

Early identification of patients likely to derive long-term benefit from Immune checkpoint inhibitors (ICIs) is crucial for treatment planning. Radiomic features beyond tumor size may assess early response [1,2]. We developed and externally validated a deep learning serial imaging biomarker using routine CT scans to predict Overall Survival (OS) for advanced non-small cell lung cancer (NSCLC), including an analysis of patients with early stable disease and heterogeneous outcomes.

Methods

We curated a real-world dataset (RWD) of 1,538 patients with advanced NSCLC treated with PD-(L)1 ICIs. Using a discovery subset of 14,186 CT series from 1,053 patients, we developed a pipeline consisting of image quality control, preprocessing, deep learning feature extraction, and a survival model. This pipeline generates a serial CT response score (serialCTRS), defined as the probability of 12-month OS. Input to the inference pipeline included paired pretreatment and three-month follow-up CT scans without additional clinical variables or manual annotations.

SerialCTRS was validated on a hold-out RWD (n=485), comparing OS Hazard Ratios (HRs) to tumor volume change derived from manual volumetric segmentations and Response Evaluation Criteria in Solid Tumors 1.1 (RECIST) categories of PR/CR, SD, and PD using the same input scans. SerialCTRS and volume change were each categorized as high, medium, or low response with counts matching corresponding

RECIST categories. Furthermore, we evaluated whether serialCTRS could identify patients with OS benefit within the SD group.

Results

SerialCTRS performance, compared to RECIST and volume change, is shown in the Kaplan-Meier plots in Figure 1. The serialCTRS high and low groups showed HRs of 0.33 (95% CI 0.22-0.48) and 4.01 (2.97-5.41), respectively. In comparison, the RECIST assessment group of PR/CR and PD had HRs of 0.41 (0.28-0.59) and 2.32 (1.71-3.15), respectively. The volume change groups showed HRs of 0.31 (0.21-0.47) and 2.86 (2.12-3.87). SerialCTRS remained a significant predictor of OS after multivariate adjustment with other known predictors, including volume change, PD-L1 TPS, age, sex, and line of therapy (Figure 2). Within the SD group, the 12-month OS ROC-AUC was 0.74 (0.65-0.82) and 0.62 (0.52-0.72) for serialCTRS and volume change, respectively, demonstrating improved prognostication for serialCTRS.

Conclusions

We developed and validated a deep learning serial imaging biomarker that identified patients who derived OS benefit with an improved HR compared to RECIST and tumor volume change. This platform, requiring no manual annotations, can be easily integrated into treatment planning and drug development workflows. Additional external validation using clinical trial data is underway.

References:

[1] Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, I., Mak, R. H., & Aerts, H. J. W. L. (2019). Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging. Clinical Cancer Research, 25(11), 3266–3275.

[2] Dercle, L., Zhao, B., Gönen, M., Moskowitz, C. S., Firas, A., Beylergil, V., Connors, D. E., Yang, H., Lu, L., Fojo, T., Carvajal, R., Karovic, S., Maitland, M. L., Goldmacher, G. v., Oxnard, G. R., Postow, M. A., & Schwartz, L. H. (2022). Early Readout on Overall Survival of Patients With Melanoma Treated With Immunotherapy Using a Novel Imaging Analysis. JAMA Oncology.

Ethics Approval:

Ethics approval for US data:

The study was conducted under IRB-approved procedures using de-identified data for patients diagnosed with Stage-IV NSCLC and treated between Jan. 1, 2017 and December 31, 2021. All records were de-identified per HIPAA guidelines at the institution level. Upon transfer, the data was quarantined and then re-inspected by authorized personnel prior to ingestion to ensure compliance and that no PHI was present in the records.

Ethics approval for EU data:

The study was conducted under IRB-approved procedures using de-identified data for patients diagnosed with Stage-IV NSCLC and treated between Jan. 1, 2017 and December 31, 2021. All records were de-identified per GDPR requirements at the institution level. The patients were also notified that their de-identified data would be part of a study and were given the required time and opportunity to respond if they had any objection. Upon transfer, the data was quarantined and then re-inspected by authorized personnel prior to processing to ensure compliance and that no PHI was present in the records.

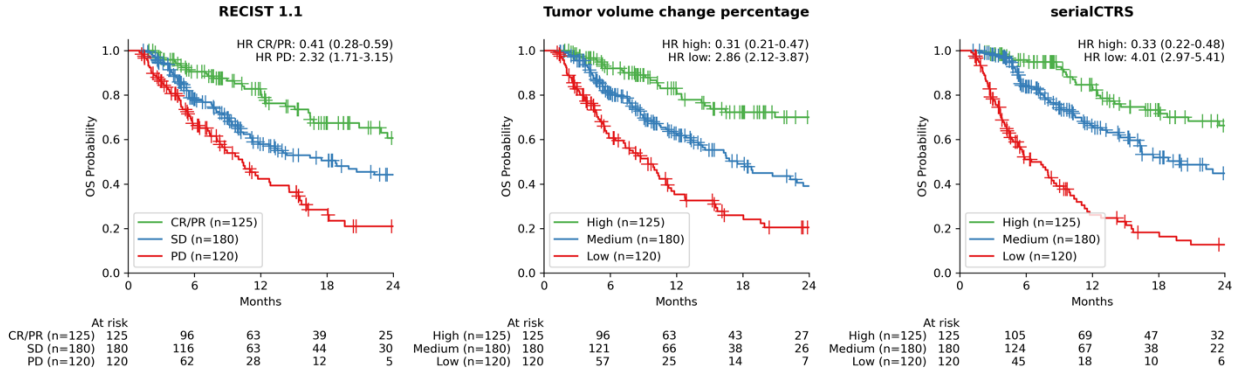

Figure 1. Kaplan-Meier OS plots for the hold-out RWD validation, where groups are compared between three-month RECIST 1.1 assessment (left), three-month tumor volume change percentage (middle), and serialCTRS (right).

Variable	Univariate HR (95% CI)	Adjusted HR (95% CI)	Adjusted HR P
serialCTRS (+10%)	0.67 (0.63 - 0.72)	0.70 (0.65 - 0.75)	<0.0001
Tumor Volume Change %	2.42 (1.85 - 3.16)	1.45 (0.96 - 2.17)	0.075
Sex (Male)	1.51 (1.09 - 2.11)	1.49 (1.07 - 2.08)	0.02
Age (≥65 years)	1.09 (0.79 - 1.51)	1.20 (0.87 - 1.66)	0.27
PD-L1 TPS (≥50%)	0.53 (0.38 - 0.74)	0.58 (0.41 - 0.82)	0.0018
ICI in 2L+	1.64 (1.19 - 2.25)	1.27 (0.90 - 1.77)	0.17

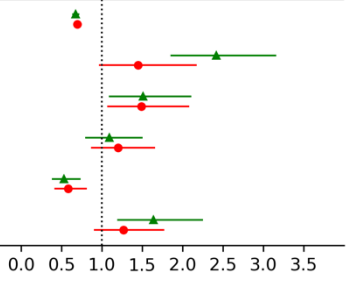

Figure 2. OS HRs for serialCTRS and covariates after multivariate adjustment with a Cox proportional hazard regression model (circle), shown together with univariate HRs (triangle), for the hold-out RWD. HR for serialCTRS was normalized for 10% higher score.