Deep learning to estimate RECIST in patients with NSCLC treated with PD-1 blockade

SUPPLEMENTARY TABLES AND FIGURES

Table of Contents

	Page
Supplementary Table 1. Baseline characteristics……	2
Supplementary Table 2. Rules for Determining Progression Free Survival by the "Ensemble Method"…..…….……	3
Supplementary Figure 1. Standardized RECIST coding sheet…….……	4
Supplementary Figure 2. Attention map and loss functions for the MSK training cohort ……..……..……...	5
Supplementary Figure 3. Performance of the DL model ……..……..……	6
Supplementary Figure 4. Performance of RECIST vs DL model PFS ……	7

Supplementary Table 1. Baseline characteristics

	Training (n=361)	Internal test (n=92)	External test (n=97)
Age, median [range] - years	67 [32-88]	65 [42-93]	66 [36-85]
Sex - no. (%)			
Female	177 (49%)	49 (53%)	45 (46%)
Male	184 (51%)	43 (47%)	52 (54%)
ECOG PS - no. (%)			
0-1	334 (93%)	86 (93%)	86 (89%)
>1	27 (7%)	6 (7%)	11 (11%)
Smoking history - no. (%)			
Former/ Current	320 (89%)	81 (88%)	83 (86%)
Never	41 (11%)	11 (12%)	14 (14%)
Histology - no. (%)			
Adenocarcinoma	272 (75%)	74 (80%)	68 (70%)
Squamous cell lung carcinoma	69 (19%)	12 (13%)	25 (26%)
Other (adenosquamous, LC-NE, NOS)	20 (6%)	6 (7%)	4 (4%)
Line of therapy - no. (%)			
1st line treatment	149 (41%)	43 (47%)	21 (22%)
>1st line treatment	212 (59%)	49 (53%)	76 (78%)
Immunotherapy type - no. (%)			
PD-1 blockade	269 (75%)	70 (76%)	81 (84%)
PD-1 blockade + CTLA-4 blockade	92 (25%)	22 (24%)	16 (16%)
RECIST BOR - no. (%)			
CR	5 (1%)	4 (4%)	4 (4%)
PR	77 (21%)	15 (16%)	25 (26%)
SD	132 (37%)	35 (38%)	28 (29%)
POD	147 (41%)	38 (41%)*	40 (41%)
Clinical Trial - no. (%)	151 (42%)	48 (52%)	56 (58%)
Standard of Care - no. (%)	210 (58%)	44 (48%)	41 (42%)

* Percentages may not add up to 100 due to rounding. ECOG PS = Eastern Cooperative Oncology Group Scale of Performance Status, LC-NE = large cell neuroendocrine, NOS = not otherwise specified, PD-1 = programmed cell death protein 1, CTLA-4 = cytotoxic T-lymphocyte-associated protein 4, RECIST BOR = response evaluation criteria in solid tumors best overall response, CR = complete response, PR = partial response, SD = stable disease, POD = progression of disease

Supplementary Table 2. Rules for Determining Progression Free Survival by the "Ensemble Method"

1	If both Method A and Method B determine progression, date of progression is defined as date predicted by Method B
2	If Method A determines progression but Method B determines NO progression, last recorded scan date is used as date of progression
3	If Method A determines no progression, the patient is treated as not having progression. Patient is censored at last recorded scan date.

Supplementary Figure 1

Tumor Measurements using RECIST v1.1
Patient Name:
Patient MRN:

INDEX LESIONS													
Index Lesion Code	Site	Scan Type	Screening Date:	Timepoint1 Date:	Timepoint2 Date:	Timepoint3 Date:	Timepoint4 Date:	Timepoint5 Date:	Timepoint6 Date:	Timepoint7 Date:	Timepoint8 Date:	Timepoint9 Date:	Timepoint10 Date:
	Long diameter (mm). Nodes use short axis diameter												
T01													
T02													
T03													
T04													
T05													
Sum of index lesions			0	0	0	0	0	0	0	0	0	0	0
% Change from Baseline													
% Change from Nadir													
OVERALL RESPONSE BY RECIST 1.1													

NON-INDEX LESIONS													
Non-Index Lesion Code	Site	Scan Type	Screening Date:	Timepoint1 Date:	Timepoint2 Date:	Timepoint3 Date:	Timepoint4 Date:	Timepoint5 Date:	Timepoint6 Date:	Timepoint7 Date:	Timepoint8 Date:	Timepoint9 Date:	Timepoint10 Date:
NT01													
NT02													
NT03													
NT04													
NT05													

NEW LESIONS													
New Lesion Code	Site	Scan Type	Screening Date:	Timepoint1 Date:	Timepoint2 Date:	Timepoint3 Date:	Timepoint4 Date:	Timepoint5 Date:	Timepoint6 Date:	Timepoint7 Date:	Timepoint8 Date:	Timepoint9 Date:	Timepoint10 Date:
1													
2													
3													
4													
5													

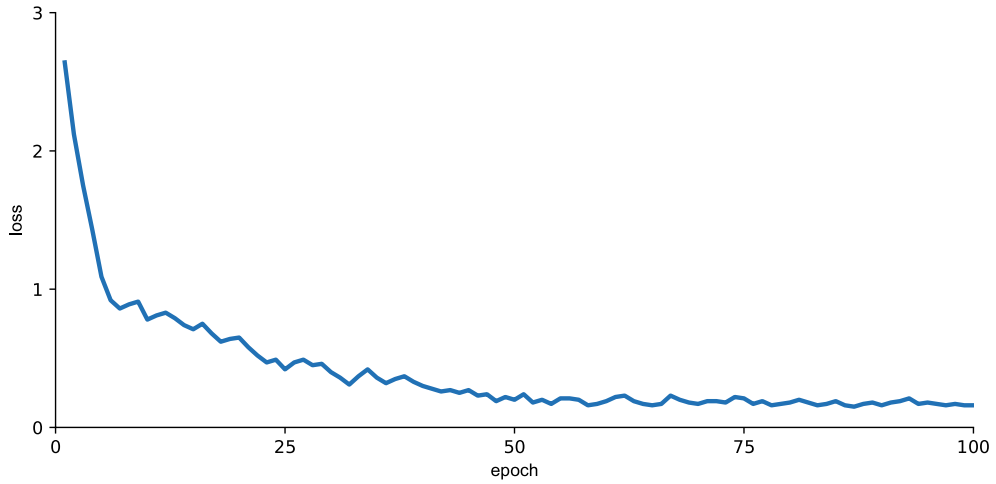
Supplementary Figure 1: Standardized RECIST coding sheet An example of the standardized form used by radiologists to record RECIST results

Supplementary Figure 2

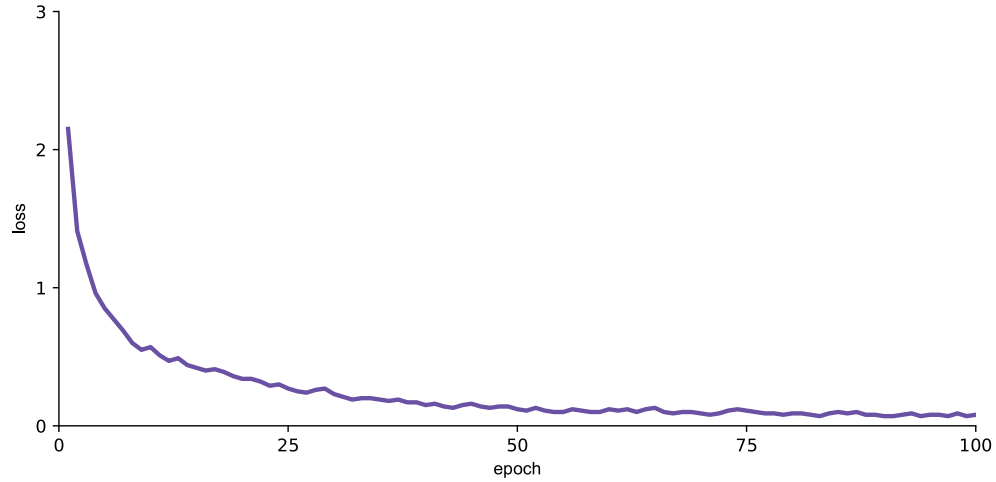
A

Left supraclavicular adenopathy is new and measures 1.3 x 1.0 cm. Right pericardial lymph adenopathy is new and measures up to 1.2 x 1.1 cm. A 1.6 x 1.1 cm in mediastinal lymph node previously measured 1.2 x 1.2 cm. 1.3 x 0.9 cm right retrocrural lymph node previously measured 1 x 0.8 cm. A 1.2 cm left hilar lymph node previously measured 1.0 cm. [redacted] Unremarkable SPLEEN: 0.8 cm low density splenic lesion is stable. PANCREAS: Unremarkable ADRENAL GLANDS: Unremarkable [redacted] too small [redacted] Clustered lymph nodes [redacted] mildly increased. [redacted] 2.0 x 1.6 cm peripancreatic lymph node previously measured 1.3 x 0.9 cm. A 0.7 cm celiac axis lymph node previously measured 0.5 cm. Subcentimeter lymph nodes persist [redacted] clustered lymph nodes in the mesentery [redacted] unchanged. [redacted] Unremarkable OTHER: Slight increase [redacted] enhancing mass along the subcutaneous right chest wall measuring 1.4 x 1.0 cm, previously 1.0 x 0.6 cm. [redacted]

B

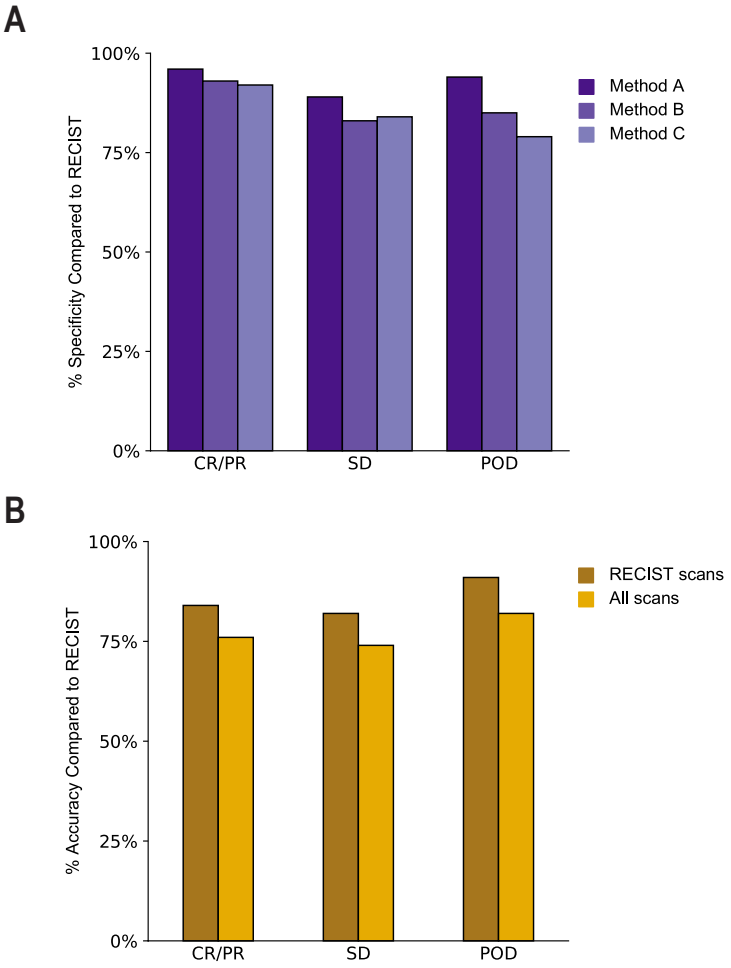


C



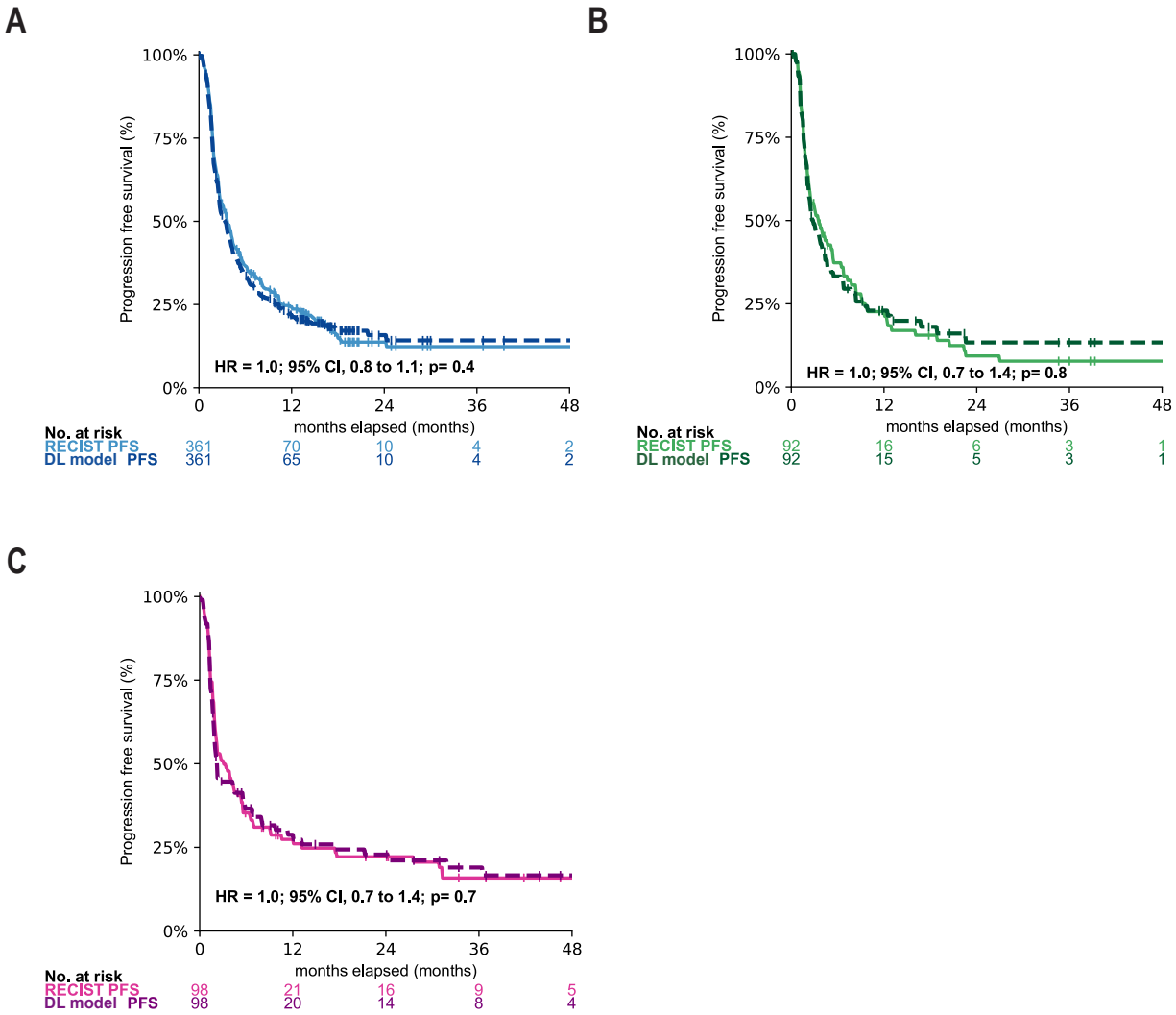
Supplementary Figure 2: Attention map and loss functions for the MSK training cohort A. Example attention map generated applying an attention-LSTM network to our model showing words and phrases the model organically identified as important during training. The intensity of the red highlighting denotes the relative importance of that word or phrase. Other text is redacted in black to avoid personal identifying information. **B.** Training loss for predicting the BOR task showing stability after epoch 60 and **C.** Training loss for predicting the PFS task showing stability after epoch 50.

Supplementary Figure 3



Supplementary Figure 3: Performance of the DL model A. Specificity of the DL model in the training cohort comparing Method A, B, and C (CR/PR: 96% vs 93% vs 92%, SD: 89% vs 83% vs 84%, POD: 94% vs 85% vs 79%). **B.** Accuracy of the DL model compared to RECIST when trained using only scans used for RECIST assessment vs all scans while on treatment (CR/PR: 84% vs 76%, SD: 82% vs 74%, POD: 91% vs 82%).

Supplementary Figure 4



Supplementary Figure 4: Performance of RECIST vs DL model PFS Kaplan-Meier PFS estimates depict comparison of outcomes defined by DL model-determined PFS or RECIST-defined PFS for the **A.** training **B.** internal test and **C.** external test cohorts.