

1244P AI-based accurate PD-L1 IHC assessment in non-small cell lung cancer across multiple sites and scanners.

Ramona Erber¹, R. Banisch², C. Schaaf⁴, N. Abele¹, F. Keil³, J. James⁵, P. Maxwell⁵, M. Päpper², P. Frey², K. Daifalla², S. Günther², A. Hartmann¹, Tobias Lang²

¹ Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Comprehensive Cancer Center Erlangen-EMN (CCC ER-EMN), Erlangen, Germany; ² Mindpeak GmbH, Hamburg, Germany; ³ Institute of Pathology, University of Regensburg, Regensburg, Germany; ⁴ Department of Internal Medicine II, Klinikum rechts der Isar of the TU Munich, Munich, Germany; ⁵ Queen's University Belfast, North Ireland, UK

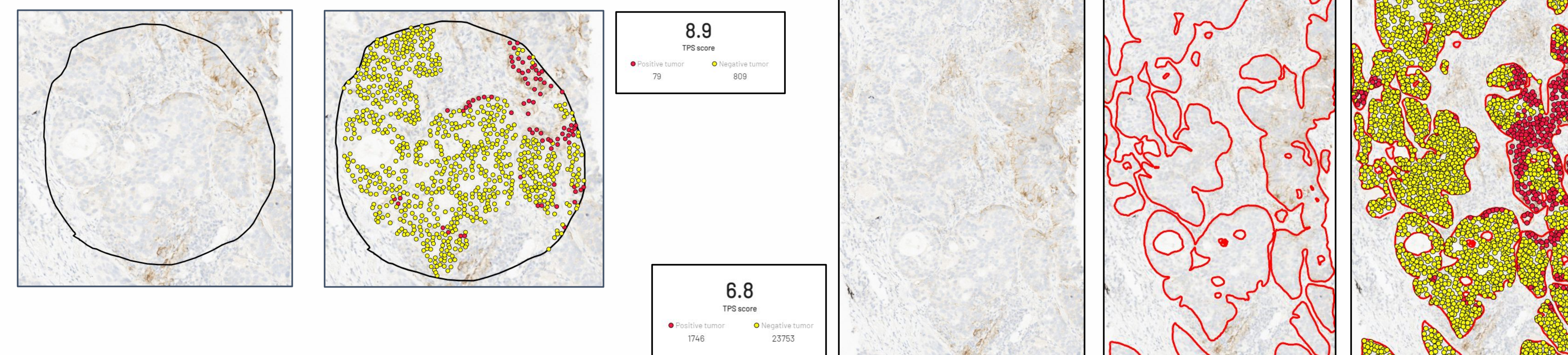
Background: Immunohistochemical PD-L1 expression, a tissue-based biomarker that predicts response to immune checkpoint therapy, is used to guide treatment decisions in advanced non-small cell lung cancer (NSCLC). Artificial intelligence (AI) technology may optimize PD-L1 scoring in regard to standardization, accuracy and efficiency. However, previous such approaches failed to show consistency across samples derived from different sites and scanning devices.

Methods:

We developed an AI software for automated PD-L1 tumor proportion scoring (TPS) in NSCLC according to international guidelines [Mindpeak PD-L1 NSCLC RoI^{CE} and PD-L1 NSCLC (WSI)^{RuO}] and evaluated its performance in this study.

Mindpeak PD-L1 NSCLC RoI^{CE}

1. Pathologist draws ROI
2. AI identifies cells and scores PD-L1



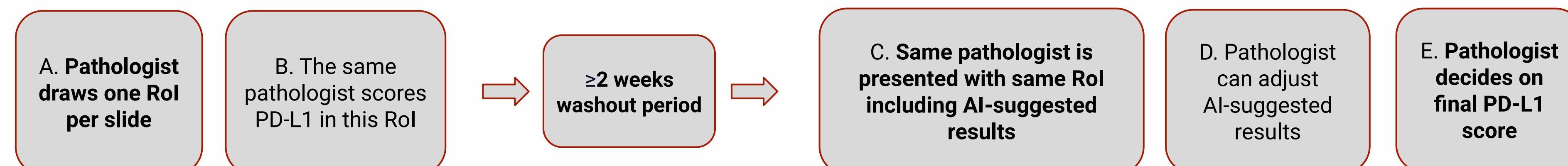
Study cohort:

N=146 WSIs from **3 institutions** (all stained with SP263 [Ventana])
4 different scanners (3D Histech, Leica, Hamamatsu, Philips)

Study setup:

1. RoI scoring

(performed by **4 pathologists** with multi-year experience in PD-L1 scoring)



2. WSI scoring (N =142; 4 cases excluded, which could not be scanned with sufficient quality for the whole-slide images)

AI scores PD-L1 on entire WSIs fully automatically (no human intervention). Comparison to scores from clinical workflow. For 24 cases, clinical scores were classified as potentially imprecise by the authors. These were re-scored by two pathologists. The majority vote of the three assessments was used to determine the final scoring category (3/5 cases changed category for 1%/50% threshold respectively).

As a reference, according to literature^[1]:

Human interobserver variability is 84.2% for the 1% cutoff and 81.9% for the cutoff at 50% for PD-L1 scoring in NSCLC.

References:

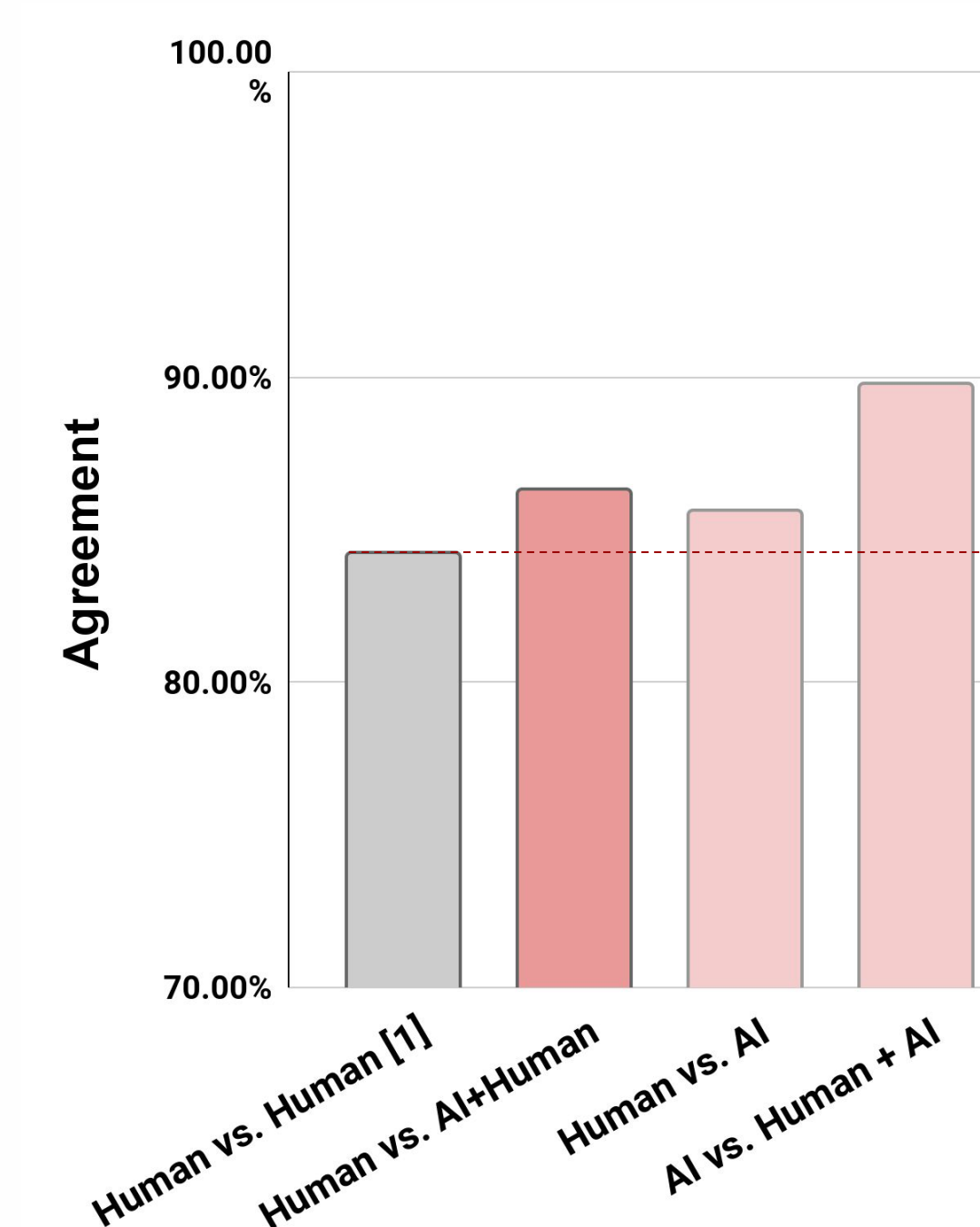
[1]: Cooper WA et al., "Intra- and Interobserver Reproducibility Assessment of PD-L1 Biomarker in Non-Small Cell Lung Cancer, Clin Cancer Res. 2017 Aug 15;23(16):4569-4577. doi: 10.1158/1078-0432.CCR-17-0151. PMID: 28420726.

Correspondence: PD Ramona Erber, MD; ramona.erber@uk-erlangen.de, +49 9131 8543634,

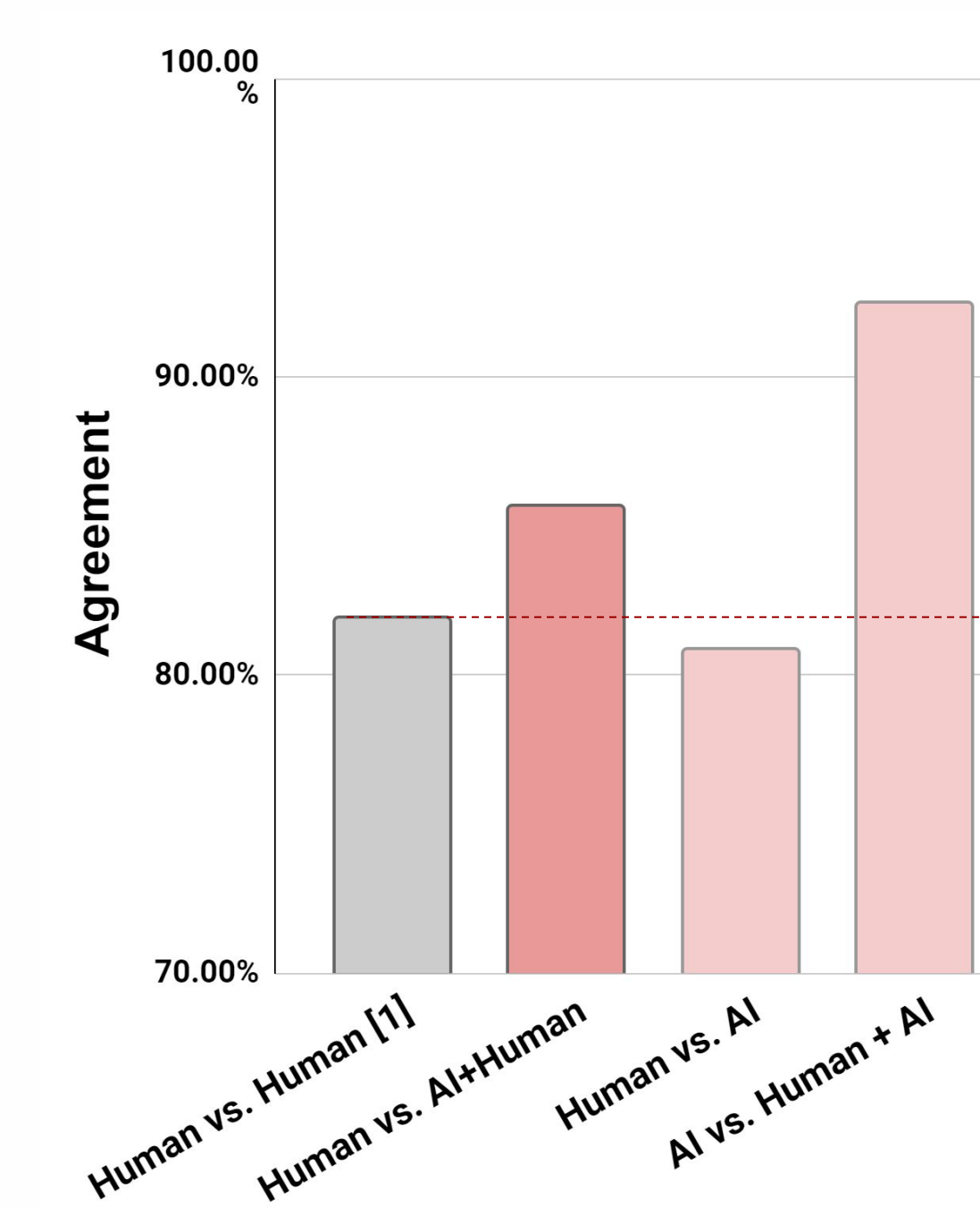
(DOI: Consulting: AstraZeneca/Daiichi Sankyo; Expert: Mindpeak; Honoraria: Novartis, AstraZeneca/Daiichi Sankyo, Diaceutics, Mindpeak, Pfizer; Research Funding: Palleos, Gilead Sciences, Biocartis, Owkin)

Results: RoI scoring

TPS 1% - Threshold

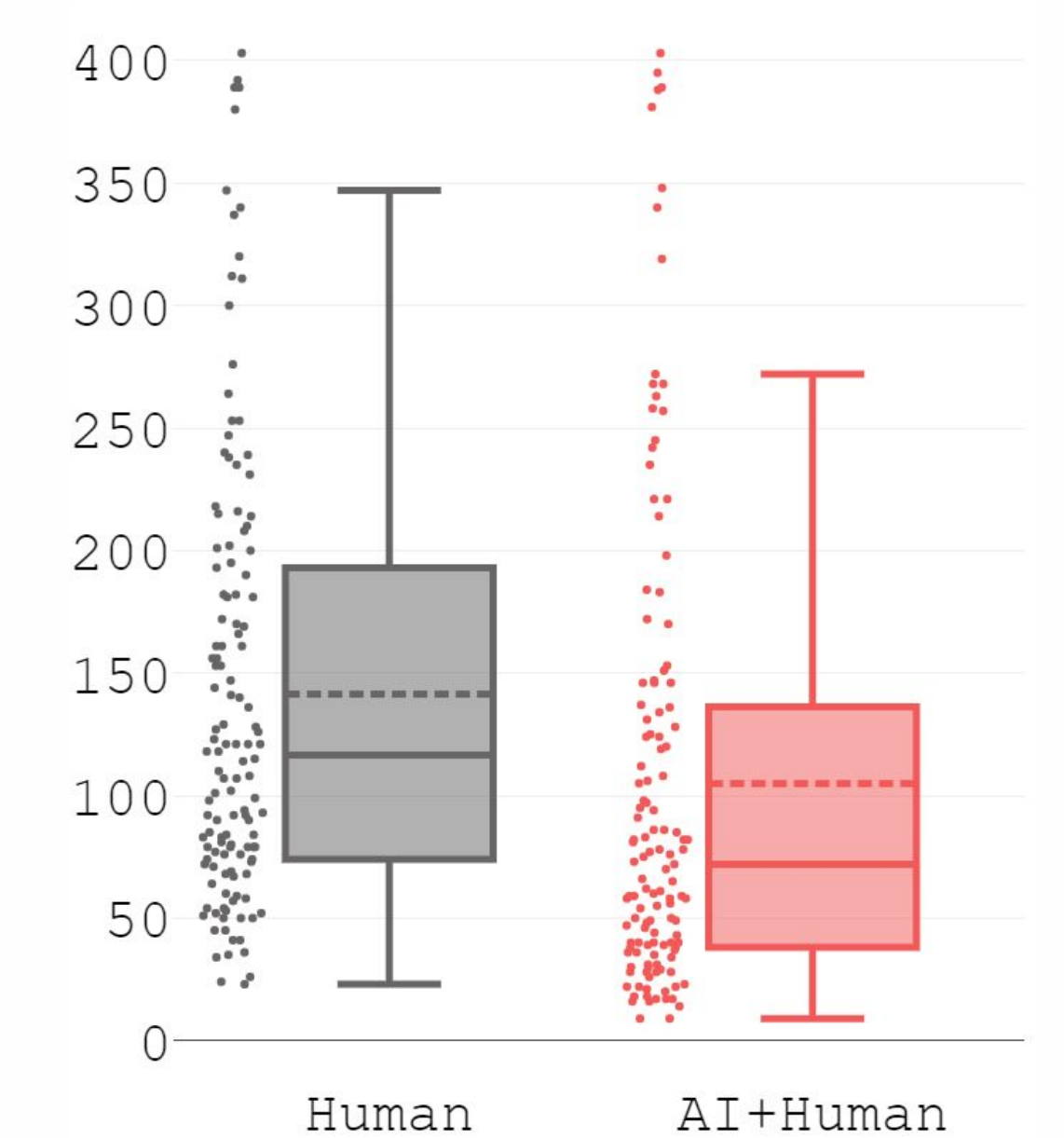


TPS 50% - Threshold



Scoring time [s/RoI]

(Median: 117 sec vs. 72 sec; p<0.01)



Use of AI as assistance tool results in higher agreement rates than those reported without AI use^[1].

AI-Assistance significantly reduces scoring time by around 40%.

Results: Autonomous AI WSI scoring

TPS 1% - Threshold

Agreement (AI vs. Human): 92.3%

AI-only	PD-L1 score	Human	
		<1%	≥1%
	<1%	33	3
	≥1%	8	98

TPS 50% - Threshold

Agreement (AI vs. Human): 93.7%

AI-only	PD-L1 score	Human	
		<50%	≥50%
	<50%	97	9
	≥50%	0	36

Fully autonomous AI PD-L1 scoring on entire WSIs shows >90% agreement with human scores.

This is higher than the human inter-reader agreement for NSCLC PD-L1 scoring reported in literature^[1].

Conclusions:

Using challenging validation data from three institutions and four scanners, ROI scoring with the support of our AI PD-L1 quantifier showed **high agreement with pathologists, while significantly reducing assessment time**. Of note, **fully automatic AI scoring on WSIs resulted in equally high agreement rates**. These results demonstrate suitability and safety of the AI system for application in clinical routine, ultimately improving accuracy and efficiency of PD-L1 scoring.