Sharp.AI: Longitudinal AI-Based Evaluation of Radiographic Progression in PsA

Zijun Gao[1†], Darshana Govind[1†], Dimitri Kessler[1], Chaitanya Parmar[1], Kenneth Broos[1], Nicholas Fountoulakis[1], Xiaoqin Tang[1], Luiza Gabriel[1], Elizabeth Hsia[1], Gabriela Oana Cula[1], Lenore Noonan[1], Philip Murphy[1], Anna Beutler[1], Terence Rooney[1], Michael Deman[1], Peter Ory[2], Kristopher Standish[1*], Robert Janiczek[1*]

[1]Johnson & Johnson, Spring House, PA, USA. [2]Virginia Mason Franciscan Health-St. Anne Hospital, WA, USA

Johnson&Johnson Innovative Medicine

Background

Accurate radiographic assessment is essential for monitoring **disease progression and treatment efficacy** in PsA clinical trials. While PsA-modified van der Heijde Sharp (vdHS) scoring remains the gold standard, its labor-intensive nature and limited sensitivity to subtle longitudinal change motivate the need for **scalable AI-based approaches with ability to capture progression over short trial intervals.**

Objective

To develop and evaluate **Sharp.AI**, an end-to-end deep learning pipeline for cross-sectional and longitudinal radiographic assessment.

Method

• Data & Study Design

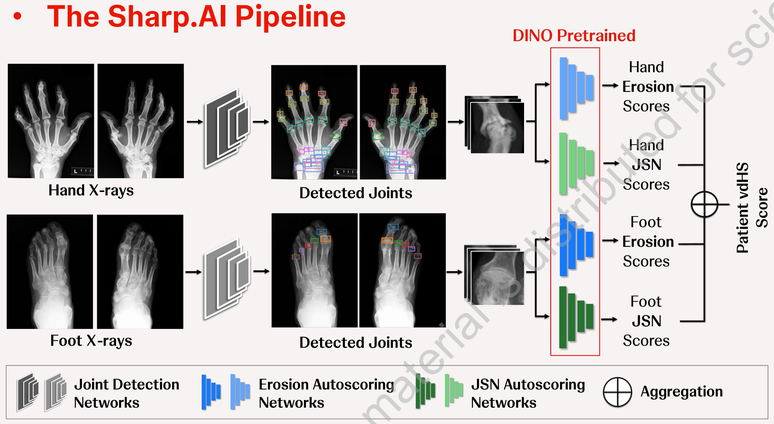
- Longitudinal PsA and RA clinical trial X-rays
- Multiple timepoints (baseline → follow-up)
- Expert reader panel as reference
- Multi-site data spanning hundreds of sites: PsA (19 countries) and RA (31 countries).

Table 1. Demographic of (a).PsA and (b).RA Dataset

a. PsA Data		Missing	Overall	Training	Validation	Holdout Test
N			1401	840	211	350
Age, Median [Q1,Q3]			48.0 [39.0,55.0]	48.0 [38.0,56.0]	46.0 [36.0,53.0]	48.0 [40.2,56.0]
Sex, n (%)	Female		679 (48.5)	380 (45.2)	100 (47.4)	199 (56.9)
	Male		722 (51.5)	460 (54.8)	111 (52.6)	151 (43.1)
BMI, Mean (Std)		2	30.2 (7.0)	30.5 (7.2)	29.3 (7.0)	30.0 (6.6)
PsA-modified vdHS, median [Q1,Q3]		13	12.5 [3.5, 34.2]	12.0 [3.5, 34.1]	13.0 [4.0, 36.2]	12.8 [3.6, 33.5]

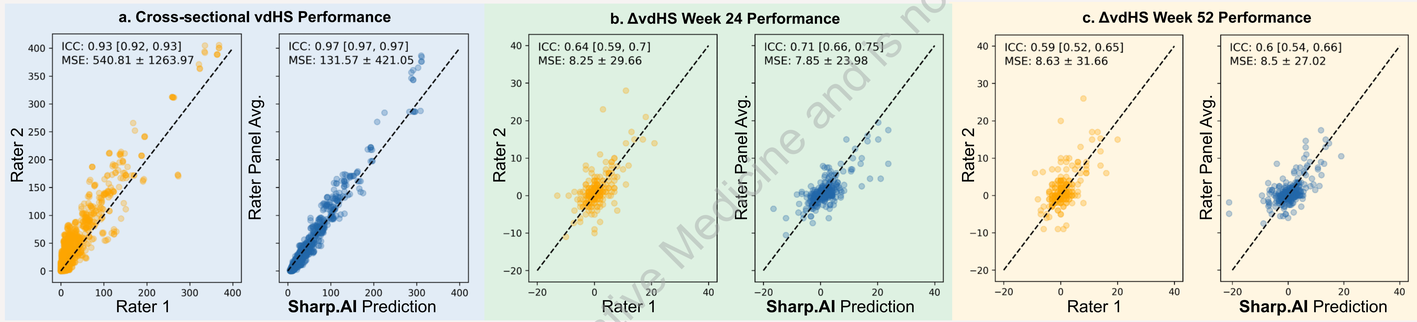
b. RA Data		Missing	Overall	Training	Holdout Test
N			2749	2153	596
Age, Median [Q1,Q3]			53.0 [44.0,60.0]	52.0 [44.0,60.0]	54.0 [45.0,62.0]
Sex, n (%)	Female		2219 (80.7)	1737 (80.7)	482 (80.9)
	Male		530 (19.3)	416 (19.3)	114 (19.1)
BMI, Mean (Std)		4	27.2 (6.1)	27.1 (6.0)	27.6 (6.5)
vdHS, median [Q1,Q3]		27	16.5 [4.5,49.4]	16.5 [4.5,49.0]	16.5 [4.5,51.0]

• The Sharp.AI Pipeline



Results

• Patient-level Performance on PsA Test Set



End-to-end **Patient-level**	Rater 1 vs. Rater 2		**Sharp.AI** vs. Rater Panel Avg.	
Task (Evaluated Samples #)	ICC [95% CI]	MSE (Mean ± Std.)	ICC [95% CI]	MSE (Mean ± Std.)
a. Cross-sectional vdHS (1,575)	0.93 [0.92, 0.93]	540.81 ± 1263.97	**0.97 [0.97, 0.97]**	**131.57 ± 421.05**
b. ΔvdHS Week 24 (420)	0.59 [0.52, 0.65]	8.63 ± 31.66	**0.6 [0.54, 0.66]**	**8.5 ± 27.02**
c. ΔvdHS Week 52 (424)	0.64 [0.59, 0.7]	8.25 ± 29.66	**0.71 [0.66, 0.75]**	**7.85 ± 23.98**
ΔvdHS Week 100 (123)	**0.88 [0.83, 0.91]**	**11.27 ± 38.76**	0.86 [0.80, 0.90]	15.07 ± 41.36

Q1: 25th percentile. Q3: 75th percentile. MSE: mean squared error; Std.: Standard deviation; ICC: intraclass correlation coefficients; CI: Confidence Interval

• Joint-level Performance on PsA Holdout Test Set

End-to-end **Joint-level**	Rater 1 vs. Rater 2		**Sharp.AI** vs. Rater Panel Avg.	
Task (Evaluated Joints #)	ICC [95% CI]	MSE (Mean ± Std.)	ICC [95% CI]	MSE (Mean ± Std.)
Foot BE (13,014)	0.87 [0.85, 0.88]	0.86 ± 5.49	**0.94 [0.94, 0.95]**	**0.39 ± 2.1**
Foot JSN (13,014)	0.91 [0.9, 0.92]	0.23 ± 0.81	**0.97 [0.97, 0.98]**	**0.08 ± 0.24**
Hand BE (43,187)	0.87 [0.86, 0.88]	0.3 ± 1.2	**0.9 [0.89, 0.91]**	**0.15 ± 0.72**
Hand JSN (43,185)	0.91 [0.9, 0.91]	0.23 ± 0.85	**0.97 [0.97, 0.97]**	**0.08 ± 0.31**

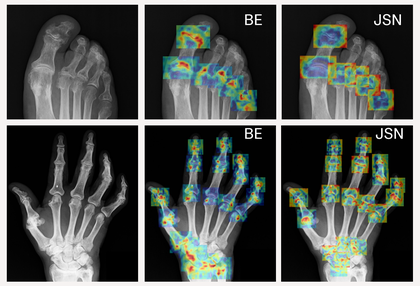
• Patient-level Performance on RA Holdout Test Set

End-to-end **Patient-level**	Rater 1 vs. Rater 2		**Sharp.AI** vs. Rater Panel Avg.	
Task (Evaluated Samples #)	ICC [95% CI]	MSE (Mean ± Std.)	ICC [95% CI]	MSE (Mean ± Std.)
Cross-sectional vdHS (2,883)	**0.96 [0.96, 0.96]**	**164.41 ± 352.43**	0.95 [0.95, 0.96]	179.76 ± 344.71
ΔvdHS ≤ Week 24 (377)	0.64 [0.57, 0.69]	8.21 ± 34.42	**0.65 [0.59, 0.71]**	**10.19 ± 26.45**
ΔvdHS Week 52 (685)	**0.66 [0.61, 0.7]**	**9.1 ± 29.03**	0.64 [0.59, 0.68]	14.63 ± 36.95
ΔvdHS Week 104 (685)	**0.79 [0.75, 0.82]**	**11.93 ± 39.49**	0.73 [0.68, 0.77]	19.71 ± 75.86
ΔvdHS ≥ Week 208 (230)	0.64 [0.56, 0.71]	23.17 ± 77.33	**0.66 [0.58, 0.73]**	**32.02 ± 81.82**

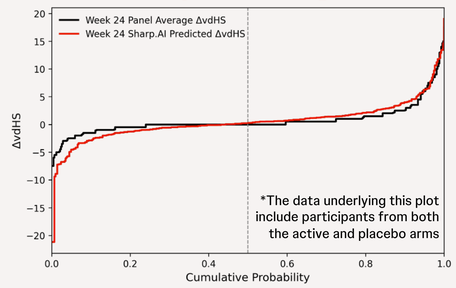
Conclusions

- Sharp.AI provides **accurate cross-sectional patient-level vdHS** estimates.
- The derived **longitudinal measures** align with expert-annotated structural progression.
- Sharp.AI **generalizes well to RA** scoring despite not being trained on RA data.
- The Sharp.AI pipeline provides a **scalable, objective** alternative to manual scoring, with the potential to **accelerate clinical trial workflows** and **improve consistency** in monitoring structural joint damage.
- Future work includes performing **analytical validation** on a **hold out test dataset** to enable regulatory discussion on the potential for Sharp.AI to support and/or **replace a central reader panel** for radiographic analyses endpoints.

• Explainability - Attention Map



• Sensitivity to Change*



References

1. van der Heijde D. *Imaging and scoring methods in psoriatic arthritis*, 2005.
2. Govind D. et al. *ViT for automated radiographic assessment in PsA*, 2024.
3. Wassenberg S. *Radiographic scoring methods in psoriatic arthritis*, 2015.
4. Deimel T. et al. *Deep learning for radiographic progression scoring*, 2020.

Acknowledgements

Acknowledgement to the **AutoPiX** project supported by the Innovative Health Initiative Joint Undertaking and its members, and other contributing partners under grant agreement No.101194766.

Contact Information

Corresponding author emails:
kstandis@its.jnj.com; rjanicze@its.jnj.com

Conflict of Interest

Authors who are employees of Johnson & Johnson may own stock or stock options in the company. Anna Beutler has previously received consulting fees from Kiniksa Pharmaceuticals and Boston Pharmaceuticals.

Mobile Version QR Code

The QR code is intended to provide scientific information for individual reference, and the information should not be altered or reproduced in any way.