

# リウマチ画像解析の学習データと AI ベンチマークを公開

～1,048 手の X 線データセットがリウマチ診断支援の進化を加速～

## ポイント

- ・手首 X 線マルチタスクデータセット「RAM-W600」を公開。
- ・セグメンテーションと標準的骨侵食スコアリングに対応し、高品質なアノテーションを提供。
- ・関節リウマチ自動評価 AI ベンチマークとして活用可能で、診断支援やアルゴリズム比較を促進。

## 概要

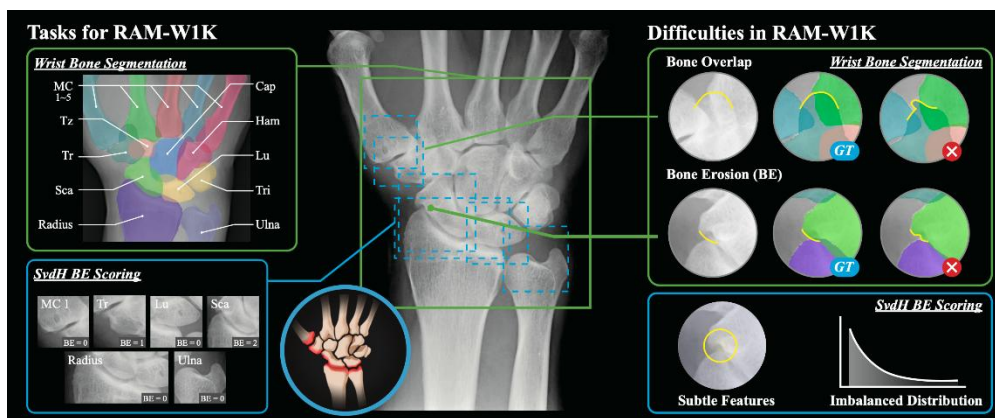
北海道大学大学院保健科学研究所の神島保教授、同大学量子集積エレクトロニクス研究センターの池辺将之教授、同大学大学院保健科学院博士後期課程の王昊霖氏、東京科学大学工学院システム制御系の奥富正敏特任教授、博士後期課程の楊松暁氏、同大学総合研究院の欧亜非研究員らの研究グループは、関節リウマチ（RA）の診断支援に向け、手首 X 線画像に基づく初のマルチタスクデータセットと AI ベンチマーク<sup>\*1</sup>を公開しました。

RA は代表的な自己免疫疾患であり、臨床現場では X 線画像が関節破壊評価に広く用いられています。特に手首は診断上重要な部位ですが、複雑な骨構造や疾患進行による骨変形のため、高精度なアノテーションが難しく、コンピューター支援診断（CAD）研究は限られていました。

研究チームは手首 X 線画像に基づく初の公的マルチタスクデータセット RAM-W600 を公開しました。対象は 388 名の患者の X 線画像 1,048 枚で、618 枚に骨インスタンスセグメンテーション注釈、800 枚に標準的評価法である、Sharp/van der Heijde (SvdH) 法による骨びらんスコアが付与されています。さらに、Unet、TransUNet などの代表的な深層学習アーキテクチャや、SAM といった基盤モデルを用いた多様な AI ベンチマーク実験が実施され、性能比較が行われました。

RAM-W600 は、RA 関連の多様な研究課題に資する可能性を有しています。また、手根骨骨折の局在化など、手首に関連する他の課題にも応用可能です。本データセットが手首領域における CAD 研究の障壁を大幅に低減し、RA 研究及び臨床応用の発展を促進することを期待しています。

本研究成果は、2025 年 12 月 2 日（火）から米国サンディエゴで開催された AI 関連難関国際会議 NeurIPS の Dataset and Benchmark Track で発表されました。また、2026 年 6 月 10 日（水）～12 日（金）に横浜で開催される画像センシングシンポジウムにおいても発表される予定です。



公開されたマルチタスクデータセットと AI ベンチマーク

## 【背景】

関節リウマチ（RA）は早期から手首関節を侵しやすい疾患であり、X線画像は広く用いられているものの、骨構造の複雑性や病的変形により、専門医による評価への依存度が高く、効率性に乏しいことがコンピューター支援診断（CAD）の発展を制限してきました。既存の公開データセットは、ピクセルレベルの分割アノテーションを欠くか、あるいは臨床的スコアが不足しており、RA 特異的な研究ニーズを十分に満たしていませんでした。本研究では、RA に特化した世界初の手首関節 X線画像多タスク大規模公開データセット RAM-W600 を構築しました。本データセットは、手首関節骨のインスタンス分割及び Sharp/van der Heijde (SvdH) 骨侵食スコアという二つの主要タスクを網羅し、高品質なアノテーションとベンチマークを提供します。これにより、関節間隙狭小化の定量化や骨びらん検出といった多様な下流タスクを支援し、RA 関連 CAD 研究のさらなる進展を後押しします。

## 【研究手法・研究成果】

本研究では、RA に特化した大規模公開データセット RAM-W600 を構築しました。本データセットは、四つの医療機関から収集した 1,048 枚の手首関節 X線画像（388 名の患者に由来）で構成され、そのうち 618 枚には骨単位のピクセルレベル分割アノテーションが、800 枚には SvdH 基準に準拠した骨侵食スコアが付与されています（図 1）。研究チームはこれを基盤として多タスクベンチマーク実験を実施し、手首関節骨インスタンス分割及び BE 分類に対する性能評価を行いました。評価には、Unet、Unet++、TransUNet、SwinUMamba などの多様なディープラーニングアーキテクチャや、汎用モデル（SAM、MedSAM）を採用し、DSC、NSD、BACC、F1-score といった指標を用いて体系的に性能比較を行いました（図 2）。

結果として、分割タスクでは教師ありモデルが総じて優れた性能を示し、特に SwinUMamba は DSC 97.75% という高い精度を達成しましたが、骨重なり部位や骨侵食領域では依然として顕著な精度不足が見られました。一方、汎用モデルの SAM は明らかに劣る結果となりました。BE 分類タスクでは、モデル全体の性能は限定的であり、最高でも平衡精度は約 52.6%にとどまり、F1-score も低値で、クラス不均衡に起因する感度不足が課題として確認されました。総じて、RAM-W600 は RA 手首関節画像に対する高品質な分割・スコアリング資源を提供するとともに、既存手法が複雑な臨床病変に直面した際の限界を浮き彫りにし、今後より堅牢な CAD 技術の開発に向けた基盤を提示しています（図 3）。

本研究成果は、2025 年 12 月 2 日（火）から米国サンディエゴで開催された AI 関連難関国際会議 NeurIPS の Dataset and Benchmark Track で発表されました。論文名：RAM-W600: A Multi-Task Wrist Dataset and Benchmark for Rheumatoid Arthritis (RAM-W600：関節リウマチに向けた手首関節マルチタスクデータセットとベンチマーク) <https://arxiv.org/abs/2507.05193>

また、2026 年 6 月 10 日～12 日に横浜で開催される画像センシングシンポジウムにおいても発表される予定です。

## 【今後への期待】

今後の研究は、多尺度コンテキストと境界感受性メカニズムを融合した専用ネットワークの開発に取り組み、骨重なり部位や骨侵食領域における分割精度の向上を目指す必要があります。同時に、微細な病理的特徴を捉えるモデル構造を設計することで、早期病変に対する感度を高め、病期の異なる症例に対する堅牢性を強化することが求められます。さらに、本データセットにおける骨侵食

サンプルの分布不均衡を克服するために、少数クラスの拡張戦略やより効率的な特徴抽出法を探索することが不可欠です。また、多地域・多集団にわたるデータの拡充により、モデルの施設間及び集団間における汎化能力を高め、RA 画像解析の自動化と臨床応用価値を一層推進していくことが期待されます。

## 学会情報

発表名 RAM-W600: A Multi-Task Wrist Dataset and Benchmark for Rheumatoid Arthritis (RAM-W600 : 関節リウマチに向けた手首関節マルチタスクデータセットとベンチマーク)

著者名 楊 松暁<sup>1</sup>、王 昊霖<sup>2</sup>、付 堯<sup>3</sup>、田 野<sup>4</sup>、神島 保<sup>5</sup>、池辺将之<sup>3</sup>、欧 亜非<sup>6</sup>、奥富正敏<sup>1</sup> ( <sup>1</sup>東京科学大学工学院システム制御系、<sup>2</sup>北海道大学大学院保健科学院、<sup>3</sup>北海道大学量子集積エレクトロニクス研究センター、<sup>4</sup>東京大学生産技術研究所、<sup>5</sup>北海道大学大学院保健科学研究所、<sup>6</sup>東京科学大学総合研究院)

学会名 NeurIPS2025 (AI 関連難関国際会議)

開催日 2025 年 12 月 2 日 (火) ~12 月 7 日 (日)

## お問い合わせ先

北海道大学大学院保健科学研究所 教授 神島 保 (かみしまたもつ)

T E L 011-706-2824 F A X 011-706-2824 メール ktamotamo2@hs.hokudai.ac.jp

U R L <https://kamishima.hs-hokudai.jp/>

## 配信元

北海道大学社会共創部広報課 (〒060-0808 札幌市北区北 8 条西 5 丁目)

T E L 011-706-2610 F A X 011-706-2092 メール jp-press@general.hokudai.ac.jp

東京科学大学総務企画部広報課 (〒152-8550 東京都目黒区大岡山 2 丁目 12 番 1 号)

T E L 03-5734-2975 F A X 03-5734-3661 メール media@adm.isct.ac.jp

【参考図】

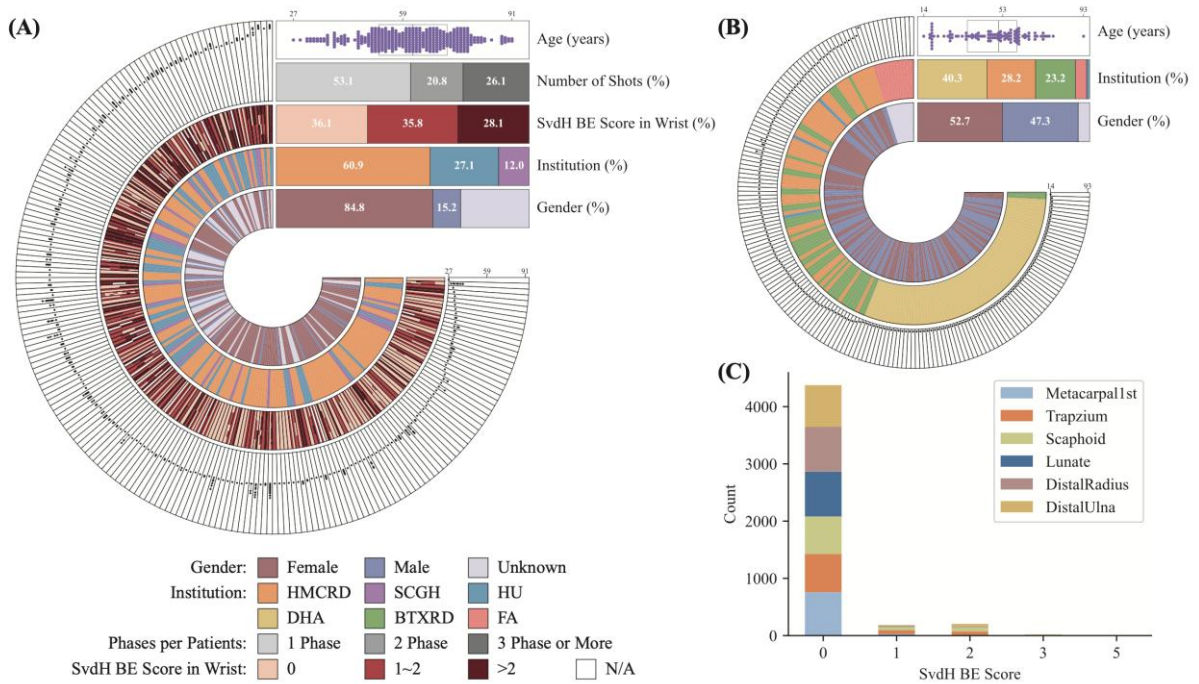


図 1. データベースにおけるデータ分布と統計情報。本研究で用いたデータベースの構成を示しており、  
 (A) 患者単位での分布  
 (B) 施設・性別の集計分布  
 (C) 部位別のスコア分布  
 を通じて、データの全体像を可視化している。

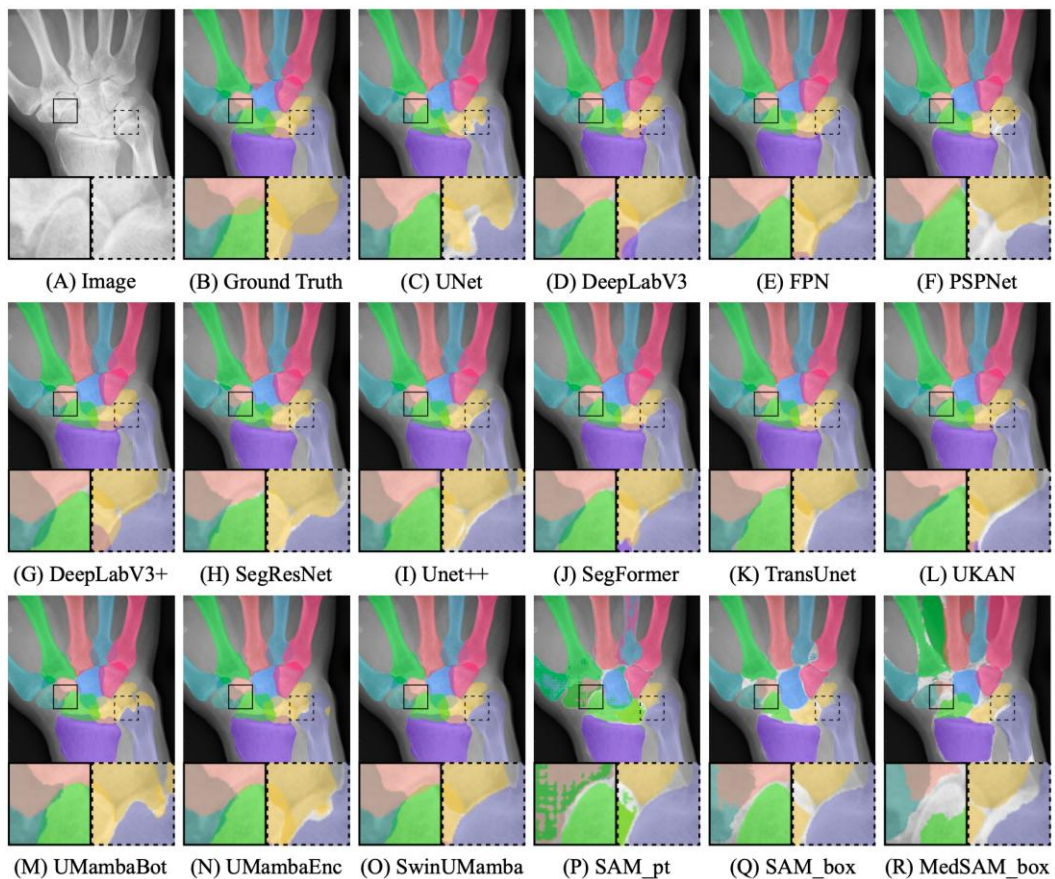


図 2. AI ベンチマークの可視化結果。

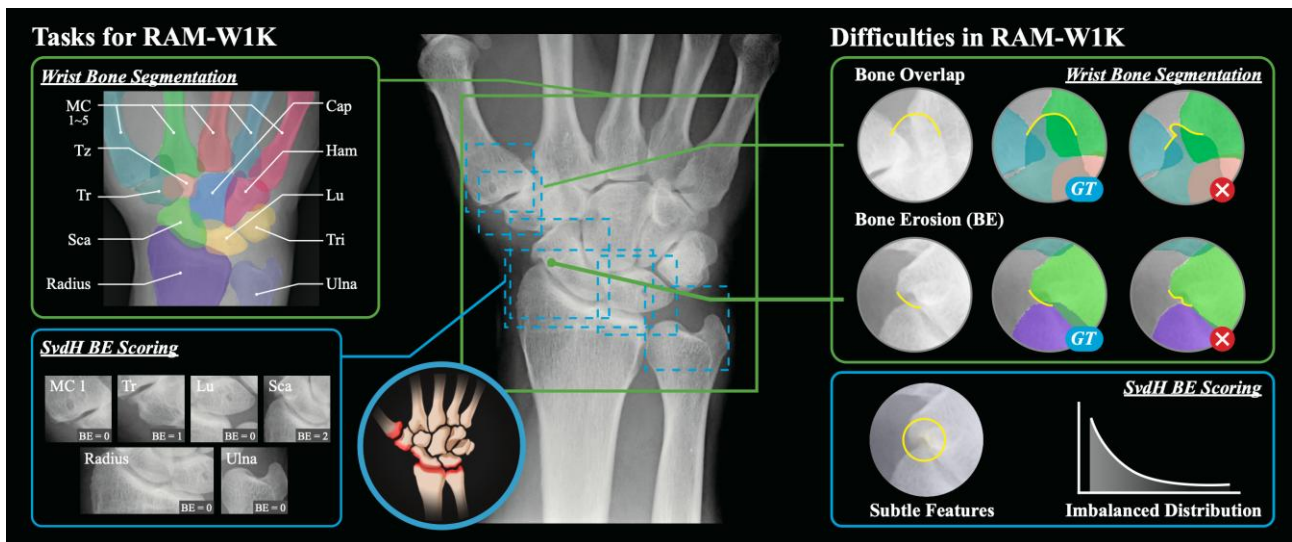


図 3. 公開されたマルチタスクデータセットと AI ベンチマーク。

**【用語解説】**

\*1 AI ベンチマーク … 統一されたデータセット、明確なタスク、標準化された評価指標に基づき、異なるモデルやアルゴリズムを体系的かつ客観的に比較・評価するための枠組みであり、再現性と対照性のある性能基準を提供することで手法の改良や課題の可視化を促し、医学画像解析においては臨床応用に資する CAD 開発の基盤となる。