Automated Early Detection of Rheumatoid Arthritis in Radiographs Using Neural Networks

Jerome TEYI[1], Brice Kevin Kollo Nzima[2], Prof. Sophie Dabo[3]

[1] African Institute for Mathematical Sciences (AIMS), Senegal
[2] Faculty of Medicine and Biomedical Sciences, University of Yaoundé I, Cameroon
[3] University of Lille, France

jerometeyi2@gmail.com brice.kollo@univ-yaounde1.cm sophie.dabo@univ-lille.fr

Abstract

Rheumatoid arthritis (RA) poses diagnostic challenges in resource-limited African settings because early erosive changes on radiographs are subtle and advanced imaging is often unavailable. We propose a joint-centric framework for automated erosion detection on RAM-W600 wrist radiographs, formulated as binary joint-level classification (non-erosive vs. erosive) under severe imbalance (91% non-erosive). An ImageNet-pretrained ResNet-18 trained with focal loss achieves AUC = 0.654, sensitivity = 0.53, and specificity = 0.71 on the test set at threshold 0.314. In addition, joint space width (JSW) analysis validates sub-pixel measurements through controlled synthetic joint space narrowing, revealing strong anatomical variability across joints. Overall, this work provides a reproducible baseline and highlights data and measurement constraints that must be addressed for scalable, trustworthy AI-assisted screening in African healthcare.

Keywords: Rheumatoid Arthritis; Wrist Radiographs; Joint-Level Analysis; Erosion Detection; Class Imbalance; Generative Augmentation

1 Introduction

Rheumatoid arthritis (RA) is a chronic systemic autoimmune disease in which early structural damage may be subtle and spatially localized, yet delayed recognition can lead to irreversible disability (1). In many resource-limited clinical contexts, including several African settings, limited access to advanced imaging and specialist expertise increases reliance on conventional radiography for screening and follow-up. However, radiographs are less sensitive in early disease and their interpretation is time-consuming and subject to inter-observer variability, particularly when damage is mild (1).

In this paper, we study automated *joint-centric* assessment of early RA damage on wrist radiographs from the RAM-W600 dataset (**?**). We focus on joint-level erosion detection under severe class imbalance and report a reproducible baseline based on transfer learning trained with focal loss, together with clinically interpretable operating points. Beyond classification, we analyze joint space width (JSW) and validate sub-pixel measurement sensitivity through *controlled synthetic* joint space narrowing, highlighting strong joint-dependent anatomical variability. Our main contributions are:

- a joint-level baseline for early erosion detection on RAM-W600 under extreme class imbalance;
- an evaluation protocol reporting AUC and sensitivity–specificity trade-offs at an optimized decision threshold;
- a complementary JSW analysis with controlled synthetic narrowing to validate measurement sensitivity and characterize joint-dependent variability.

2 Related Work

Recent progress in medical imaging AI has been driven by the shift from rule-based systems to data-driven learning, where models learn predictive mappings directly from labeled images. In particular, deep learning enables

end-to-end hierarchical representation learning that reduces dependence on handcrafted features and improves the extraction of subtle patterns from noisy, high-dimensional radiographs (2; 3). In musculoskeletal radiography, large-scale datasets such as MURA have facilitated systematic benchmarking and transfer learning, demonstrating strong performance for abnormality detection across upper-extremity studies (4).

For rheumatoid arthritis (RA), the literature has increasingly moved from global image-level decisions toward clinically aligned pipelines that localize joints and predict damage grades to emulate expert scoring systems, thereby reducing subjectivity and reader variability (5). Beyond categorical scoring, progression-sensitive methods target the quantification of joint space narrowing (JSN), where changes may be sub-pixel and difficult to detect visually; sub-pixel estimation and registration-based strategies have been proposed to improve sensitivity to early structural progression (6). Other work addresses the challenge of anatomical overlap in projection radiography through bone layer separation models that aim to disentangle superimposed structures and enhance downstream measurement reliability (7).

Despite these advances, early RA remains challenging because erosions are subtle, spatially localized, and often underrepresented in labeled datasets, while generalization across acquisition settings and the need for joint-specific explanations continue to limit clinical deployment (1; 3). In this context, our work positions itself as a *joint-centric* approach focused on early erosion detection under severe imbalance, complemented by a quantitative joint space width (JSW) analysis to validate measurement sensitivity and characterize joint-dependent variability.

3 Methods

Dataset. Experiments are conducted on the RAM-W600 dataset (8), comprising 618 postero–anterior wrist radiographs from 335 patients with joint-level Sharp/van der Heijde (SvdH) erosion annotations and joint masks. Data are split at the *patient level* into 70% training, 10% validation, and 20% test sets to prevent information leakage. After ROI extraction, this yields 3,354 training, 486 validation, and 960 test joint samples. Erosion detection is formulated as a binary task (SvdH = 0 vs. ≥ 1), with severe imbalance (91% non-erosive joints).

Preprocessing. Radiographs are converted to grayscale, enhanced using CLAHE, and resized to 512×512. Joint ROIs are extracted using provided masks, cropped with a margin, and resized to 224×224. ROI-level augmentations (horizontal flips, small rotations, scaling, brightness/contrast perturbations) are applied during training.

Model and imbalance handling. A ResNet-18 backbone (9) pretrained on ImageNet is used with a linear classification head; the backbone is frozen and only the head is optimized. Let z denote the output logit and $p = \sigma(z)$ the predicted erosion probability. To address class imbalance, training minimizes focal loss (10):

$$\mathcal{L}(p, y) = -\alpha y(1-p)^{\gamma} \log(p) - (1-\alpha)(1-y)p^{\gamma} \log(1-p),$$

with $\alpha = 0.85$ and $\gamma = 1.5$, which down-weights easy negatives and emphasizes minority erosive samples. Optimization uses AdamW (11) with early stopping on validation loss.

Evaluation. Metrics are computed at the joint level on the held-out test set using the model selected on validation. Performance is measured using ROC-AUC (12) and PR-AUC (13). The decision threshold is determined on the validation set (max F1) and fixed for test evaluation; sensitivity and specificity are reported at this operating point.

4 Results

Quantitative performance. Table 1 reports joint-level erosion detection performance on the validation and held-out test sets. The model achieves stable discrimination with a ROC-AUC of 0.666 on validation and 0.654 on test. Using the decision threshold selected on the validation set by maximizing the F1-score

($\tau = 0.314$), sensitivity reaches 0.477 (validation) and 0.530 (test), while specificity remains above 0.70 on both splits, reflecting a clinically meaningful screening-oriented trade-off under strong class imbalance.

Table 1: Joint-level erosion detection performance.

Split	Loss	ROC-AUC	Sensitivity	Specificity
Validation	0.0572	0.6663	0.4773	0.7443
Test	0.0575	0.6536	0.5301	0.7058

Operating-point behavior. Figure 1 shows normalized confusion matrices computed at $\tau = 0.314$. Most non-erosive joints are correctly rejected (specificity > 0.70), while approximately half of erosive joints are detected (sensitivity ≈ 0.5), highlighting the difficulty of early-stage erosion detection.

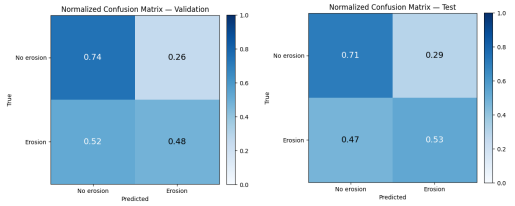

Figure 1: Normalized confusion matrices for joint-level erosion prediction on the validation (left) and test (right) sets, computed at $\tau = 0.314$.

Qualitative analysis. Figure 2 presents representative joint-level predictions on the test set. Correct detections often correspond to localized cortical irregularities, whereas most errors arise in visually ambiguous cases, with predicted probabilities close to the decision threshold, indicating uncertainty-aware behavior rather than overconfident misclassification.

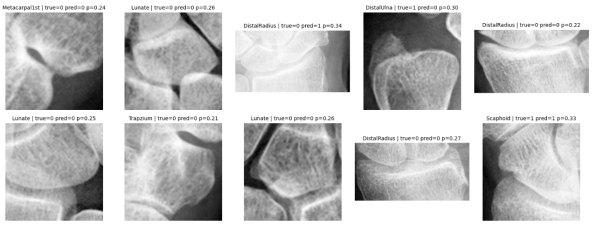

Figure 2: Representative joint-level erosion predictions on the test set. For each ROI, the true label, predicted label, and predicted probability are shown using $\tau = 0.314$.

5 Discussion

Main findings. We evaluated a joint-centric pipeline for early RA erosion detection on wrist radiographs from RAM-W600 (8). The resulting discrimination is *moderate but stable* across validation and test splits, consistent with the intrinsic difficulty of detecting subtle, localized erosions in projection radiography under severe class imbalance. By selecting an operating threshold on the validation set (F1-optimized), we intentionally favor sensitivity to reduce missed erosions, at the cost of additional false positives, a trade-off that is reasonable for a screening/triage assistant rather than a standalone diagnostic tool.

Relation to prior work. Compared with prior RA radiograph pipelines that target global or aggregated scoring (5; 14), our setting is more constrained (joint-level, early-dominant cohort) and therefore harder. More broadly, our results align with the observation that deep learning can extract clinically relevant patterns from radiographs, but performance is limited by label scarcity, subtle signals, and dataset shift (3; 15).

Limitations. First, erosion-positive samples are rare and severe grades are sparse, which constrains supervision and naturally limits achievable sensitivity. Second, radiography is less sensitive than advanced imaging modalities for very early inflammatory changes, making borderline erosions visually ambiguous even for experts (1). Third, RAM-W600 is not African data; deployment in African settings may face domain shift due to differences in devices, protocols, and population characteristics, requiring local validation and potentially adaptation before clinical use (16; 17).

Practical implications for African deployment. RA care in many Sub-Saharan contexts faces constraints in specialist access, treatment pathways, and follow-up capacity (16; 18; 19; 17). In this context, a lightweight joint-level tool could serve as a *decision-support and triage* layer on top of standard X-ray workflows (20; 21): (i) flag joints with elevated erosion probability for closer review, (ii) standard-

ize reporting across readers, and (iii) support task-shifting where rheumatology expertise is limited. For feasibility, the approach should be designed for low-cost operation and for multilingual reporting to maximize adoption in heterogeneous clinical environments.

6 Conclusion

This work investigated an end-to-end, joint-centric framework for automated assessment of early-stage rheumatoid arthritis (RA) from wrist radiographs using the RAM-W600 dataset (8). Motivated by the subtle and localized nature of early erosions, the proposed pipeline combines standardized preprocessing, mask-guided joint ROI extraction, exploratory joint space analysis, and deep learning for joint-level erosion prediction.

The results support three main conclusions. First, RAM-W600 is dominated by absent-to-mild radiographic damage at both global and joint levels, with the majority of joint surfaces labeled as non-erosive. While clinically realistic, this distribution makes early RA detection intrinsically challenging due to subtle visual cues and severe class imbalance. Second, within this setting, the proposed erosion classifier generalizes consistently and achieves moderate discrimination on the held-out test set (ROC-AUC = 0.654), with sensitivity = 0.53 and specificity = 0.71 at the F1-optimized operating threshold ($\tau = 0.314$). Importantly, the joint-centric formulation yields localized predictions that remain directly linked to anatomically meaningful regions of interest. Third, the quantitative joint space width (JSW) analysis reveals substantial anatomical variability across wrist joints, indicating that JSW is not joint-invariant; consequently, JSW is better used as a complementary structural signal for validation and insight rather than as a direct prediction target in the current formulation.

Future work will focus on improving both performance and practical readiness for African deployment. Methodologically, we will explore joint-level multi-task learning for erosion and joint space narrowing, integrate longitudinal information when repeated studies are available, and refine JSW validation using pixel spacing metadata. For deployment, we will prioritize lightweight and compute-efficient models, robust performance under dataset shift, and offline-first inference to handle limited connectivity. We also plan to support clinically usable outputs through joint-level explanations and simple reporting interfaces, and to evaluate the pipeline on external radiographs acquired under different protocols and geographic contexts to assess generalization in real-world African healthcare settings.

Acknowledgments

I am deeply grateful to my supervisor, Professor Sophie Dabo Niang, whose kindness, attentive guidance, and pursuit of excellence shaped this work in every way. I could not have asked for a better supervisor.

My sincere thanks go to my tutor, Dr. El Hadji Lamine SOKHNA, for his care, encouragement, and unwavering support throughout my journey.

I would like to express my sincere appreciation to Professor Brice Kevin Kollo Nzima, rheumatologist and faculty member at the Faculty of Medicine and Biomedical Sciences, University of Yaoundé I, for his collaboration and valuable clinical insight.

I also warmly thank Dr. DJONGANG FOSSUA Arlette Stella for her availability, guidance, and valuable support during this work.

References

[1] G. S. Firestein *et al.*, *Kelley's Textbook of Rheumatology 2-volume Set*. Kelleys Textbbok of Rheumatology, Saunders, 9th ed., 2012.

[2] I. Goodfellow *et al.*, *Deep Learning: Adaptive Computation and Machine Learning*. The MIT Press, 2016.

[3] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[4] P. Rajpurkar *et al.*, "Mura: Large dataset for abnormality detection in musculoskele-

tal radiographs," *arXiv*, vol. 1712.06957, dec 2017.

[5] J. Tan *et al.*, "Automated scoring of radiographic joint damage in rheumatoid arthritis using deep learning," *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2313–2324, 2019.

[6] Y. Ou *et al.*, "A sub-pixel accurate quantification of joint space narrowing progression in rheumatoid arthritis," *Medical Image Analysis*, vol. 86, p. 102760, 2023.

[7] H. Wang *et al.*, "Bls-gan: A deep layer separation framework for eliminating bone overlap in radiographs," *IEEE Transactions on Medical Imaging*, vol. 43, no. 4, pp. 1450–1462, 2024.

[8] S. Yang *et al.*, "Ram-w600: A multi-task wrist dataset and benchmark for rheumatoid arthritis," *arXiv preprint arXiv:2507.05193*, 2025.

[9] K. He *et al.*, "Deep residual learning for image recognition," *CVPR*, 2016.

[10] T.-Y. Lin *et al.*, "Focal loss for dense object detection," *arXiv preprint arXiv:1708.02002*, 2017.

[11] I. Loshchilov *et al.*, "Decoupled weight decay regularization," in *International Conference on Learning Representations (ICLR)*, 2019.

[12] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[13] T. Saito *et al.*, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PLOS ONE*, vol. 10, no. 3, p. e0118432, 2015.

[14] E. Ureten *et al.*, "Automated scoring of erosion and joint space narrowing in ra radiographs using yolov3 and resnet," *Journal of Medical Imaging*, 2021.

[15] D. Shen *et al.*, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.

[16] M. Kolou, "Challenges of rheumatoid arthritis management in sub-saharan africa in the 21st century," *Open Journal of Rheumatology and Autoimmune Diseases*, vol. 13, no. 1, pp. 17–40, 2023.

[17] E. N. Nwamaka, "Management of rheumatoid arthritis in africa," *research invention journal of research in medical sciences*, 2024.

[18] S. Ndongo *et al.*, "Disease-modifying treatment for inflammatory rheumatism in sub-saharan africa: Outcome at 6 months of 205 senegalese patients with rheumatoid arthritis," *Médecine et Santé Tropicales*, vol. 22, pp. 385–389, 2012.

[19] O. Adelowo *et al.*, "Rheumatoid arthritis among nigerians: The first 200 patients from a rheumatology clinic," *Clinical Rheumatology*, vol. 29, pp. 593–597, 2010.

[20] World Health Organization, "Rheumatoid arthritis," *World Health Organization*, 2023.

[21] Mayo Clinic Staff, "Rheumatoid arthritis," *Mayo Clinic*, 2023.