

# Automatic Cephalometric Landmark Localization on CBCT-Derived Digitally Reconstructed Radiographs for Skeletal Malocclusion Classification

Benjamin Hou<sup>1</sup> · Konstantina Almpanti<sup>2</sup> · Janice S. Lee<sup>2</sup> · Zhiyong Lu<sup>1</sup>  
<sup>1</sup> Division of Intramural Research, National Library of Medicine

<sup>2</sup> Craniofacial Anomalies and Regeneration Section, National Institute of Dental and Craniofacial Research  
National Institutes of Health, Bethesda, MD, USA.

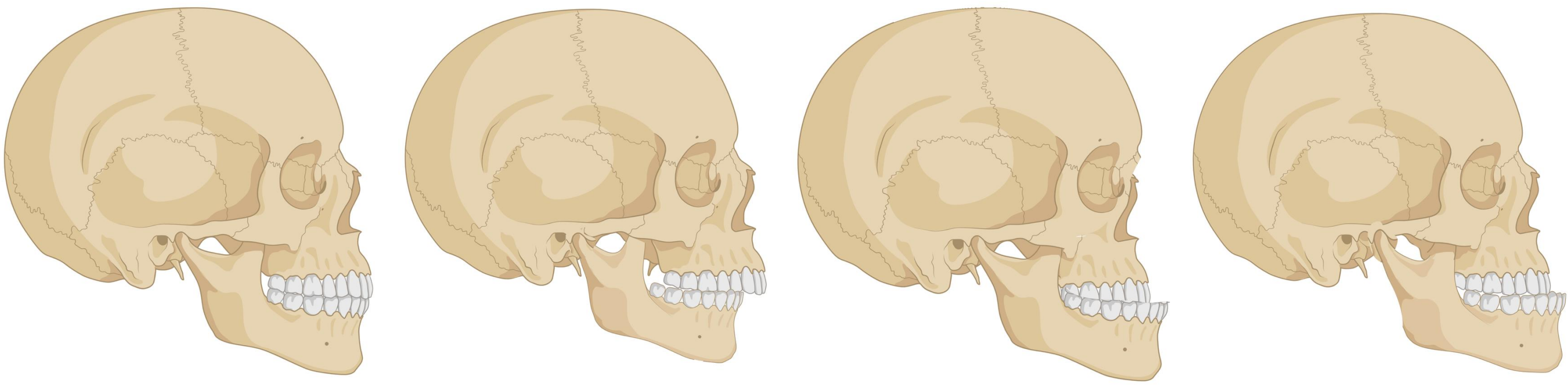


ODIN 2026

Oral & Dental Image Analysis Workshop  
MICCAI 2026 · Strasbourg · Paper #10

## CLINICAL QUESTION / MOTIVATION

Can 2D cephalometric landmark localization be applied to CBCT-derived DRRs for downstream skeletal malocclusion assessment?



CLASS I

CLASS II

CLASS IIIA

CLASS IIIB

Jaw position relative to the anterior cranial base

- Class I: Both jaws within the normative range.
- Class II: Maxilla forward and/or mandible back.
- Class IIIA: Maxillary retrusion predominates.
- Class IIIB: Mandibular protrusion predominates.

## EXPERIMENT 1 — Develop and Validate CephViT

Aariz Dataset  
1,000 cephalograms

CephViT

27 landmarks

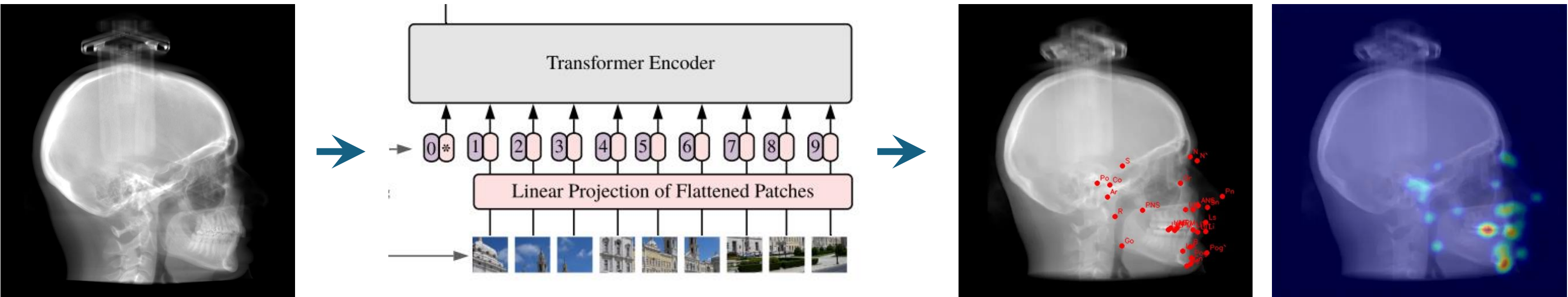
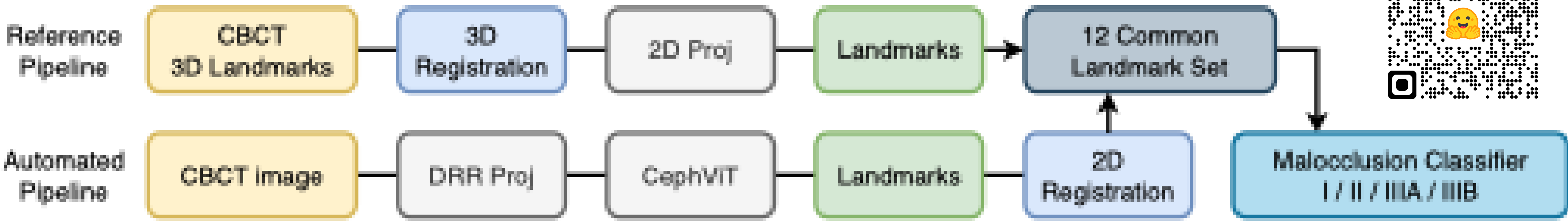


Table 1. Landmark detection performance on the Aariz test set. MRE is mean radial error; SDR is the percentage of landmarks detected within each distance threshold.

| Method         | MRE $\pm$ SD (mm) | SDR (%) |        |        |        |
|----------------|-------------------|---------|--------|--------|--------|
|                |                   | 2.0 mm  | 2.5 mm | 3.0 mm | 4.0 mm |
| Khan et al.    | 1.92 $\pm$ 7.85   | 78.54   | 85.72  | 89.64  | 94.49  |
| Khalid et al.  | 1.87 $\pm$ 4.01   | 75.17   | 82.43  | 88.78  | 93.01  |
| Khan et al.    | 1.69 $\pm$ 3.36   | 81.18   | 87.28  | 90.82  | 94.82  |
| CephViT (ours) | 1.28 $\pm$ 1.42   | 82.10   | 87.90  | 92.00  | 95.70  |

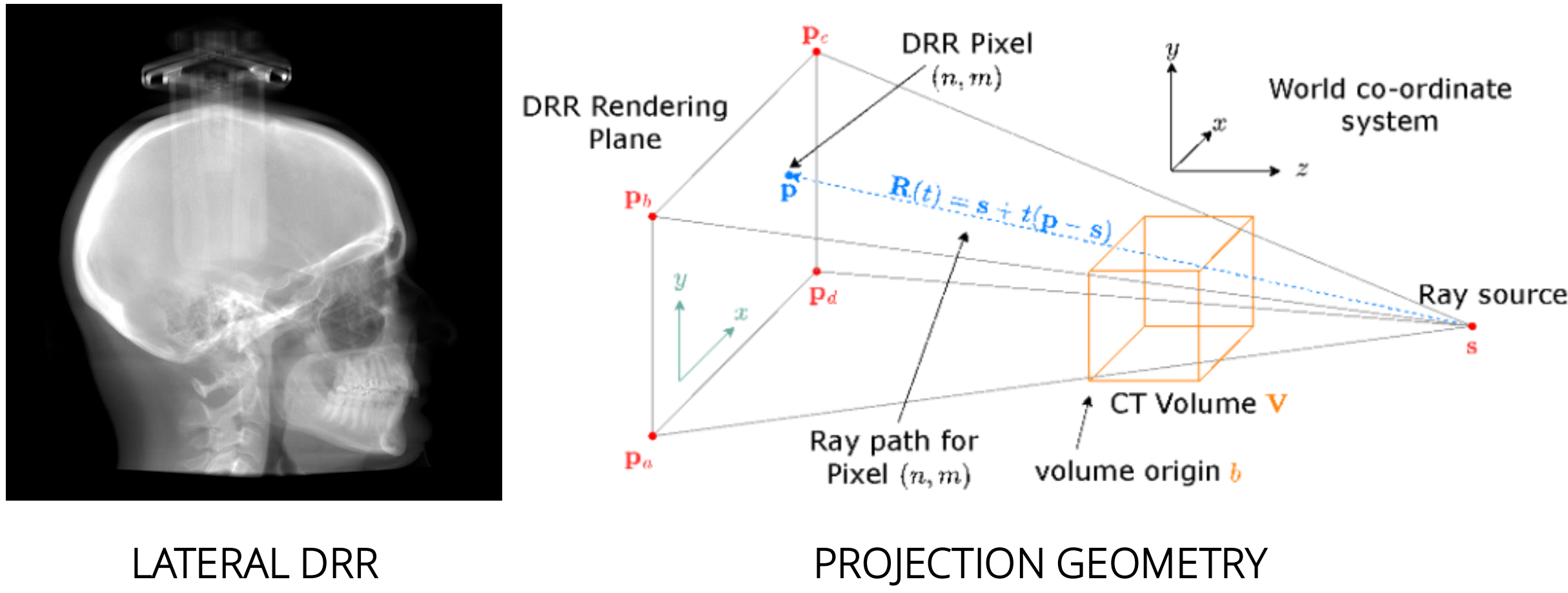
CephViT had the lowest mean error among the methods shown (1.28  $\pm$  1.42 mm), with 92.0% of landmarks detected within 3 mm.

## EXPERIMENT 2 — Apply CephViT to CBCT-Derived DRRs for Malocclusion Prediction



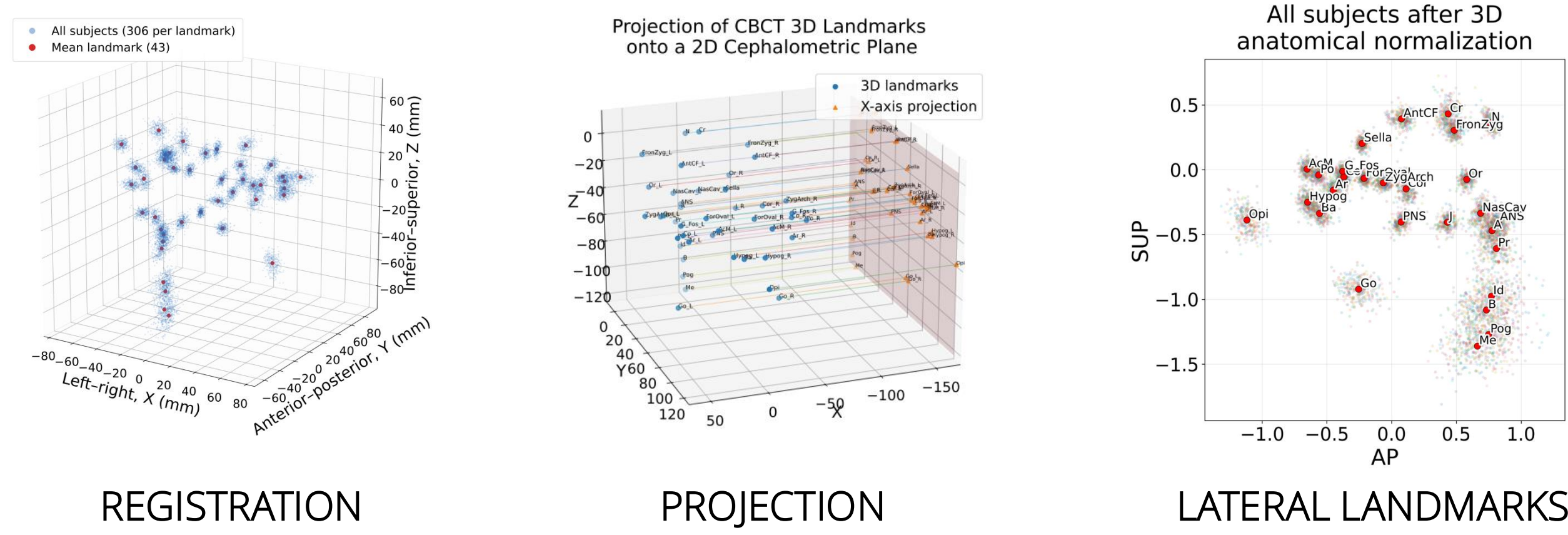
### DRR Creation

Lateral DRRs from CBCT volumes using Siddon-Jacobs ray tracing.



### 3D Landmark Projection

Anatomical normalization and sagittal projection of 3D CBCT landmarks onto a 2D cephalometric plane.



### 2D Landmark Alignment

DRR-localized landmarks before and after similarity alignment to a representative reference shape.

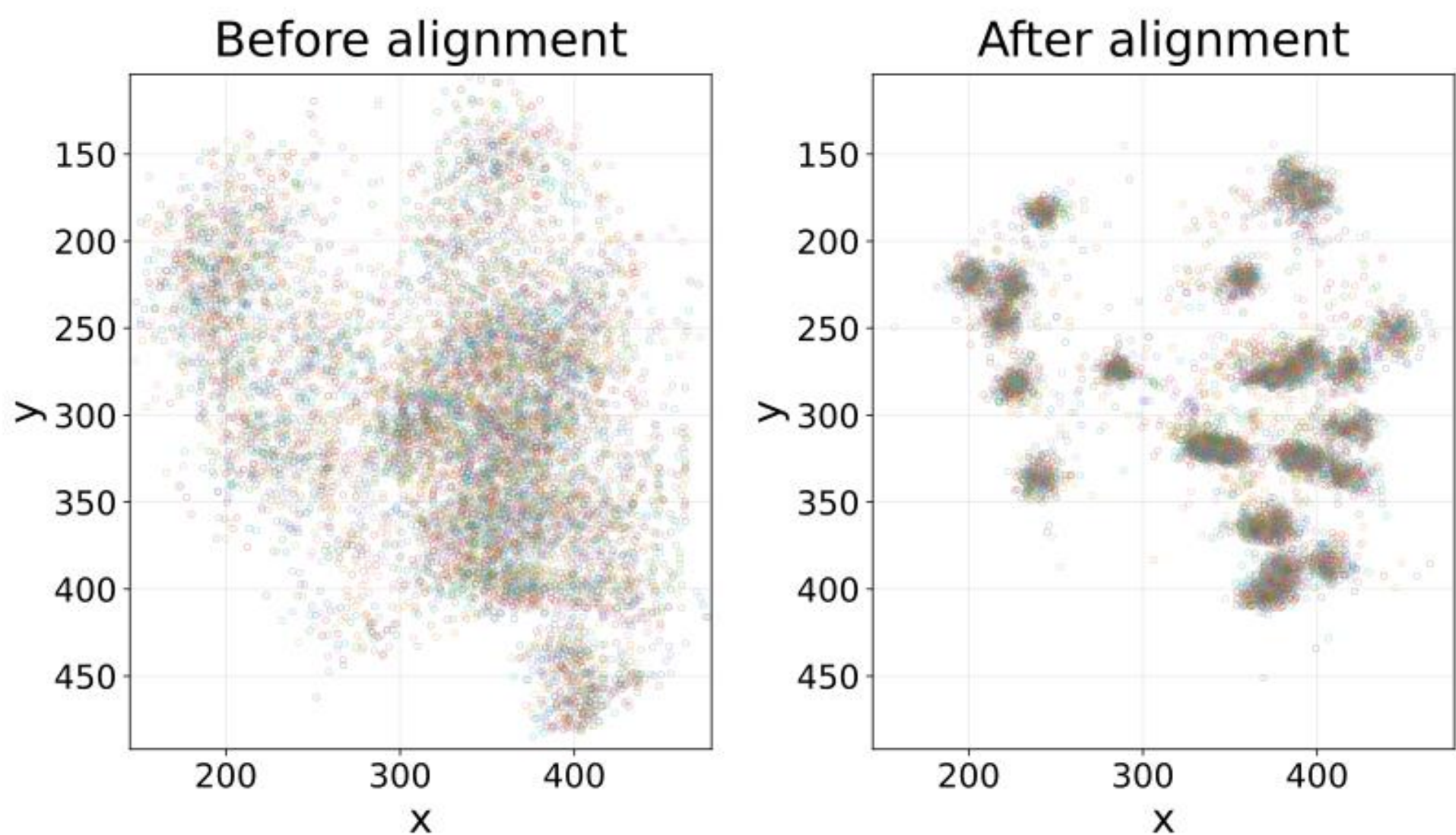


Table 2. Five-fold cross-validated malocclusion classification

| Metric      | Support | Reference landmarks | DRR-localized landmarks | $\Delta$ | 95% CI for $\Delta$ | Paired $p$ |
|-------------|---------|---------------------|-------------------------|----------|---------------------|------------|
| Accuracy    | 300     | 0.683               | 0.700                   | +0.017   | [-0.050, 0.083]     | 0.696      |
| Macro F1    | 300     | 0.642               | 0.647                   | +0.005   | [-0.066, 0.074]     | 0.894      |
| Weighted F1 | 300     | 0.688               | 0.708                   | +0.020   | [-0.045, 0.084]     | 0.555      |
| Class I     | 99      | 0.667               | 0.691                   | +0.024   | [-0.071, 0.119]     | —          |
| Class II    | 99      | 0.768               | 0.794                   | +0.026   | [-0.052, 0.107]     | —          |
| Class IIIA  | 27      | 0.431               | 0.364                   | -0.067   | [-0.223, 0.080]     | —          |
| Class IIIB  | 75      | 0.703               | 0.740                   | +0.036   | [-0.065, 0.139]     | —          |

- Accurate landmarking: 1.28  $\pm$  1.42 mm MRE; 92.0% within 3 mm.
- 2D CephViT successfully applied to CBCT-derived DRRs.
- Similar classification accuracy: 70.0% DRR versus 68.3% reference.
- No significant accuracy difference: paired  $p$  = 0.696.

### Acknowledgements

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH authors are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. This work utilized the computational resources of the NIH High-Performance Computing Biowulf cluster (<https://hpc.nih.gov>).

