

From TMJ Localization to Osteoarthritis Classification: An Anatomically Guided 3D Deep Learning Pipeline for Cone-Beam CT

Shradhdha Trivedi¹, Vrundan Sojitra², and Mariela Padilla¹

¹Herman Ostrow School of Dentistry, University of Southern California, Los Angeles, CA, USA

²Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA



1. Introduction and motivation

Clinical relevance: CBCT reveals erosion, osteophytes, sclerosis, subchondral cysts, and flattening associated with temporomandibular joint osteoarthritis (TMJ OA), but subtle or coexisting changes can be difficult to interpret consistently.

Gap: Existing deep-learning approaches often depend on manually defined TMJ regions of interest (ROIs), and it remains unclear how automatically predicted anatomy should be encoded for 3D OA-feature detection.

Goal: Develop an automated pipeline that localizes each TMJ, segments the mandible, and predicts whether radiographic OA features are **present or absent**.

Study question: Does anatomical background suppression or an explicit segmentation-mask channel provide the most useful input to a 3D classifier?

2. Clinical Context

Clinical challenge: OA changes may be subtle or coexist with adaptive remodeling.

Why automation matters: Consistent identification of these features could support imaging assessment and development of future clinical decision-support tools.

3. Dataset and Reference Standard

The study included **216 patients and 432 joint-level TMJ samples (Table 1)**.

(**Internal institutional cohort:** 292 TMJs; **University of Michigan Deep Blue cohort:** 140 TMJs)

Data were divided at the patient level so that the left and right TMJs from the same patient remained in the same partition:

Table 1. Patient-level partitions; counts are joint-level TMJ samples.

Site	Train	Validation	Test	Total
Internal institutional cohort	172	67	53	292
University of Michigan Deep Blue	80	20	40	140
Total	252	87	93	432

Approximately 55% of TMJs had no radiographic osteoarthritis (OA) features, while 45% had OA features. Labels were binary: any radiographic OA grade was classified as OA-features-present, and normal joints as OA-features-absent. Because Deep Blue reported OA at the patient level, an expert independently reviewed each left and right TMJ to assign side-specific labels.

4. AutoTMJ-OA Pipeline

The proposed framework consists of five stages (Fig 1):

TMJ Localization

A Faster R-CNN model with a ResNet-50 backbone and feature pyramid network detected the left and right TMJs on sagittal CBCT slices.

Three-Dimensional ROI Generation

Predictions from anatomically adjacent slices were combined to generate separate left and right 3D TMJ ROIs. Each ROI was expanded by a **5% margin**.

Mandible Segmentation

A 3D SegResNet generated a binary mandible mask containing the condylar process and visible superior ramus.

Anatomy-Guided Input Construction

The CBCT ROI and predicted mandible mask were combined to create four different input representations.

OA-Feature Detection

A 3D ResNet predicted whether radiographic OA features were present or absent.

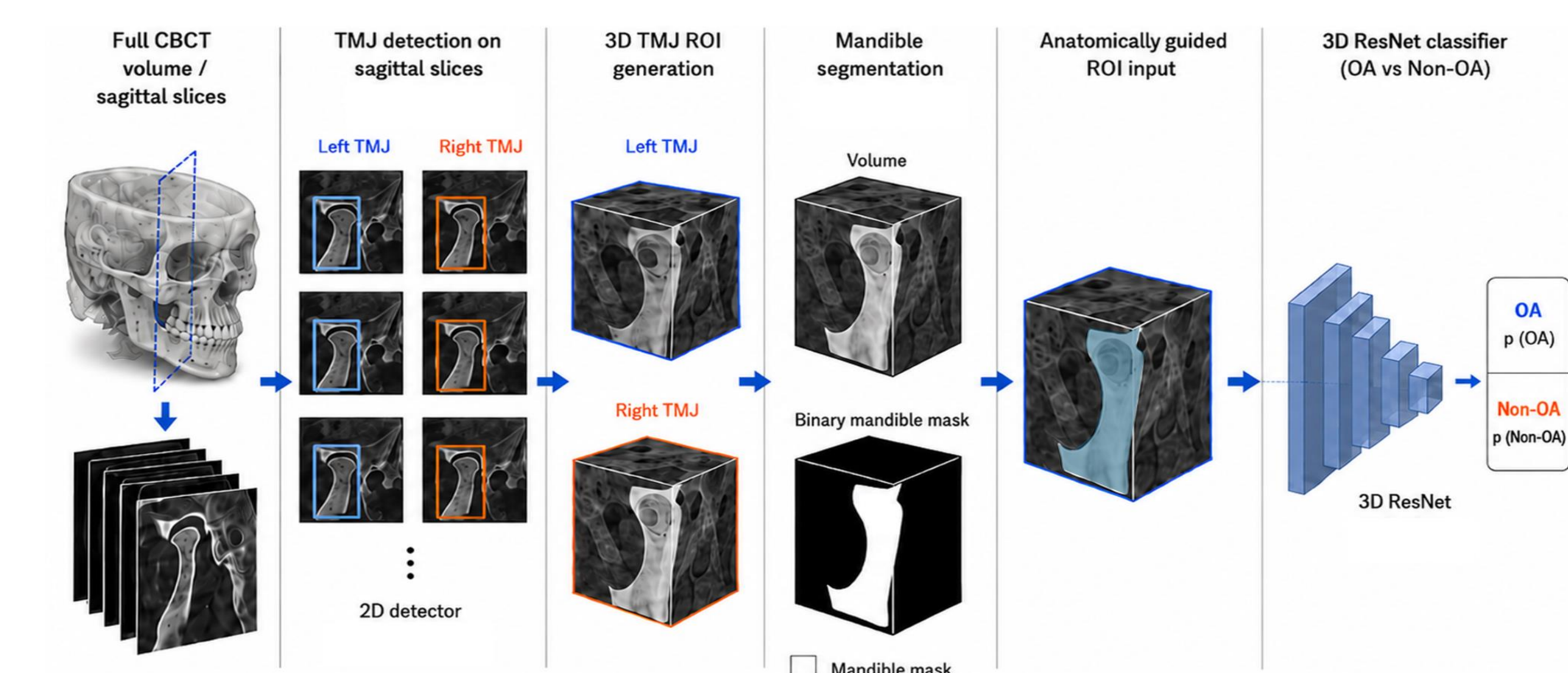


Figure 1. Schematic overview of the proposed pipeline

5. Anatomy-Guided Input Representations

Four inputs were generated from the same normalized CBCT ROI, I , and predicted mandible mask, M . (Fig 2)

A. Volume Only: The complete CBCT ROI was provided to the classifier $[I]$. This representation retained the mandibular condyle, temporal bone, joint space, and surrounding anatomical background.

B. Volume + Mask: The original CBCT ROI and binary mandible mask were provided as two separate channels. $[I, M]$. This representation retained the complete joint context while explicitly providing mandibular geometry.

C. Masked Volume: Background voxels outside the predicted mandible were suppressed $[I \odot M]$. This representation preserved cortical and trabecular information within the mandible while reducing irrelevant anatomical background.

D. Masked Volume + Mask: The masked CBCT volume and binary mandible mask were provided as two channels $[I \odot M, M]$. This representation tested whether an explicit shape channel provides additional information after background suppression.

Each representation was evaluated at five cubic input sizes: 96^3 , 128^3 , 160^3 , 192^3 , 256^3 voxels.

The anatomical field of view remained fixed. Only the sampling resolution and input representation changed.

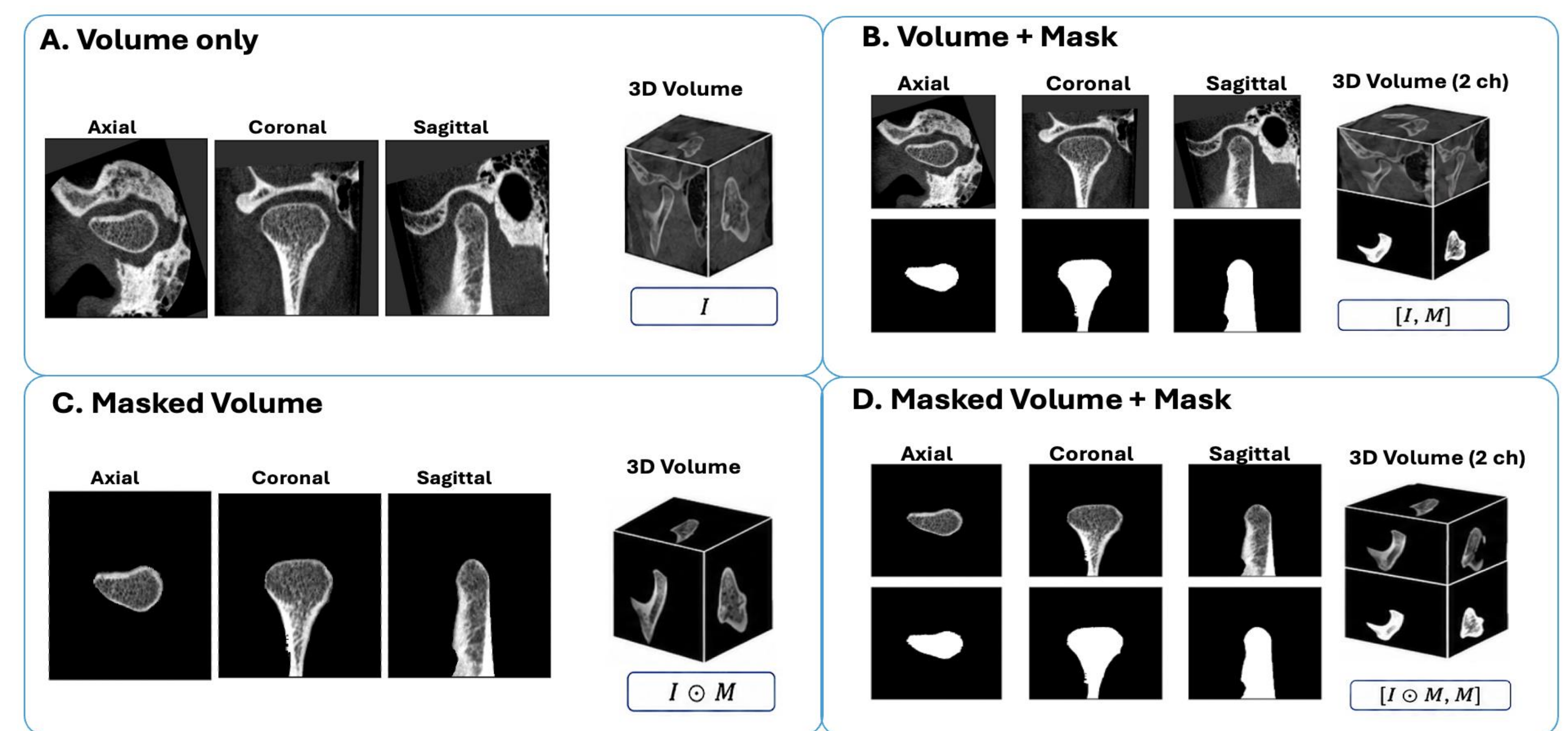


Figure 2. Anatomy-Guided Input Representations

6. Results

Anatomical masking showed the most consistent performance, achieving the highest AUC at four of five input sizes. At 192^3 , masking improved AUC from 0.904 with the complete volume to 0.938, an absolute increase of 0.034. Adding a mask channel improved performance when using the complete CBCT volume, but did not provide further benefit when added to an already masked volume. Performance increased from 96^3 to 192^3 and then plateaued, with minimal additional improvement at 256^3 (Table 2)

Table 2. Test-set AUC/F1 across input representations and cubic input sizes.

Input	96^3	128^3	160^3	192^3	256^3
Masked volume	0.884/0.868	0.908/0.892	0.931/0.914	0.938/0.926	0.935/0.924
Volume + mask	0.848/0.859	0.871/0.865	0.902/0.910	0.930/0.922	0.928/0.925
Masked vol. + mask	0.873/0.864	0.911/0.890	0.919/0.902	0.934/0.920	0.932/0.923
Volume only	0.932/0.923	0.824/0.834	0.856/0.849	0.904/0.900	0.902/0.899

7. Discussion

Background suppression likely helped the model focus on cortical and trabecular changes in the mandibular condyle while reducing irrelevant surrounding information. A separate segmentation channel improved performance with the complete CBCT region of interest but was largely redundant after masking. Performance gains plateaued beyond 192^3 , suggesting this resolution preserves adequate osseous detail with lower computational demand. Future models may benefit from combining focused mandibular anatomy with preserved temporal bone and joint-space context.

8. Conclusion

AutoTMJ-OA provides an automated pipeline for TMJ localization, mandible segmentation, and binary detection of radiographic osteoarthritic features from CBCT. Masked CBCT volume was generally the strongest input representation, achieving the highest observed performance at 192^3 voxels: **AUC = 0.938 and F1 = 0.926**

These findings suggest that **anatomical background suppression contributes more than adding a redundant segmentation-mask channel**. Independent multi-vendor validation is required before clinical application.