SEGMENTATION-GUIDED 3D DENTAL ROI TRANSFORMER FOR STRUCTURED CLINICAL REPORT GENERATION FROM CBCT

Changkai Ji[1], Yusheng Liu[1], Yuxian Jiang[1], and Lisheng Wang[1]
[1] School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China
changkaiji@sjtu.edu.cn
Code: github.com/duola-wa/MICCAI-2026-STS-Task-3

INTRODUCTION

- Task 3: derive clinically relevant information from 3D dental CBCT volumes and generate seven structured report fields.
- Clinical need: connect volumetric evidence with tooth-specific diagnoses, treatment actions, medications, and follow-up information.
- Our solution: a segmentation-guided, FDI-aligned 3D framework that unifies whole-dentition context and local tooth evidence.

CHALLENGES

- **Large 3D volumes** make full-resolution modeling expensive; sparse sampling can lose crown-to-root and periapical continuity.
- **Tooth-specific findings** must be aligned to anatomical positions while preserving relationships across adjacent teeth and dental arches.
- **Noisy segmentation** may remove diagnostic context, while only 50 labeled training cases increase the risk of overfitting.

METHOD

- **Anatomical prompting:** nnU-Net maps teeth to 32 FDI slots and defines an expanded dentition ROI without hard masking.
- **Hierarchical 3D representation:** a residual 3D CNN encodes global context, while ROI pooling creates FDI-aligned tooth tokens.
- **Structured prediction:** a Tooth Transformer fuses local and global evidence for multi-task report generation.

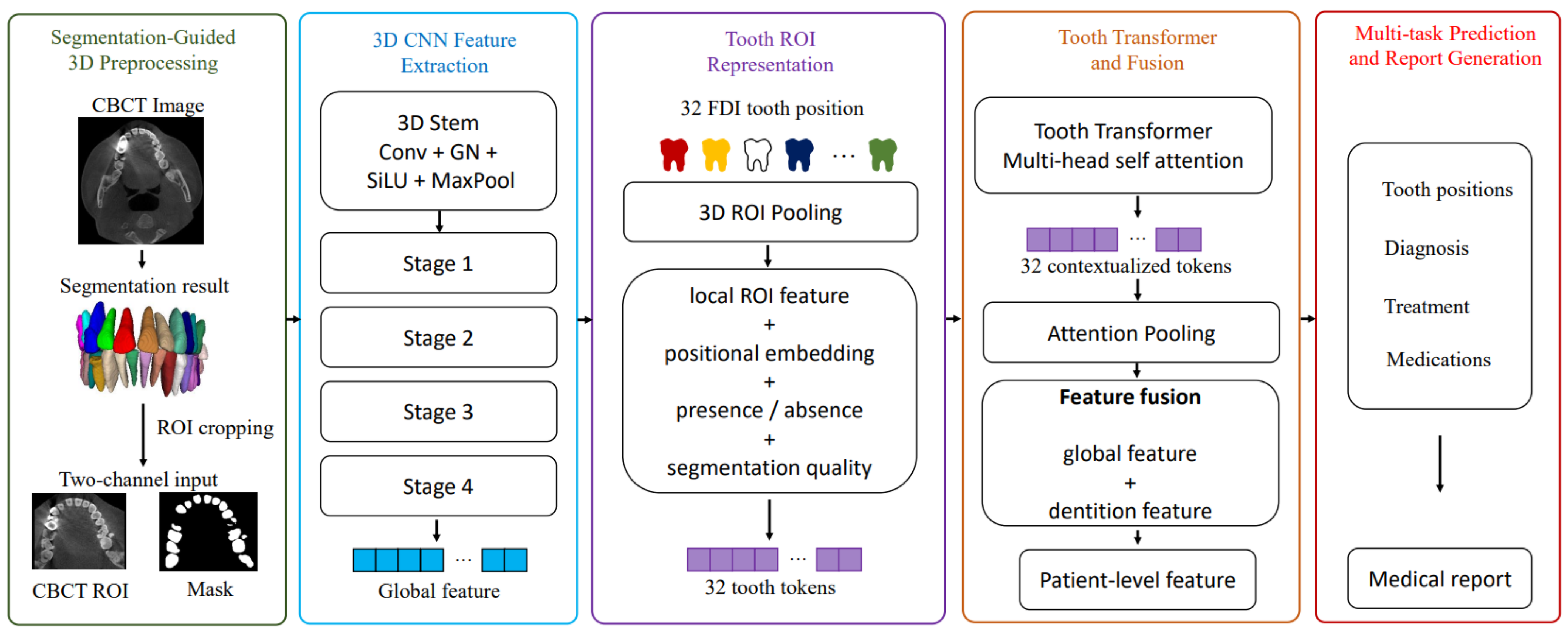


Fig. 1. Overview of the proposed framework. Global CBCT features and FDI-aligned tooth-level features are jointly modeled for structured clinical prediction and report generation.

(a) Context-preserving 3D preprocessing

- Expanded ROIs: 10 mm local tooth context and 15 mm dentition context retain roots, periapical tissues, periodontal structures, neighboring teeth, and alveolar bone.

(b) FDI-aligned tooth representation

- Each token: local 3D feature, FDI position, tooth presence, segmentation-quality descriptors.
- Reliability: quality gating and 0.35 segmentation dropout reduce dependence on imperfect localization.

(c) Multi-task report generation

- Outputs: tooth positions, diagnoses, tooth-diagnosis associations, treatment, medications, and auxiliary information.
- Deployment: deterministic templates render seven report fields; GPT is used for training-label extraction.

EXPERIMENTS

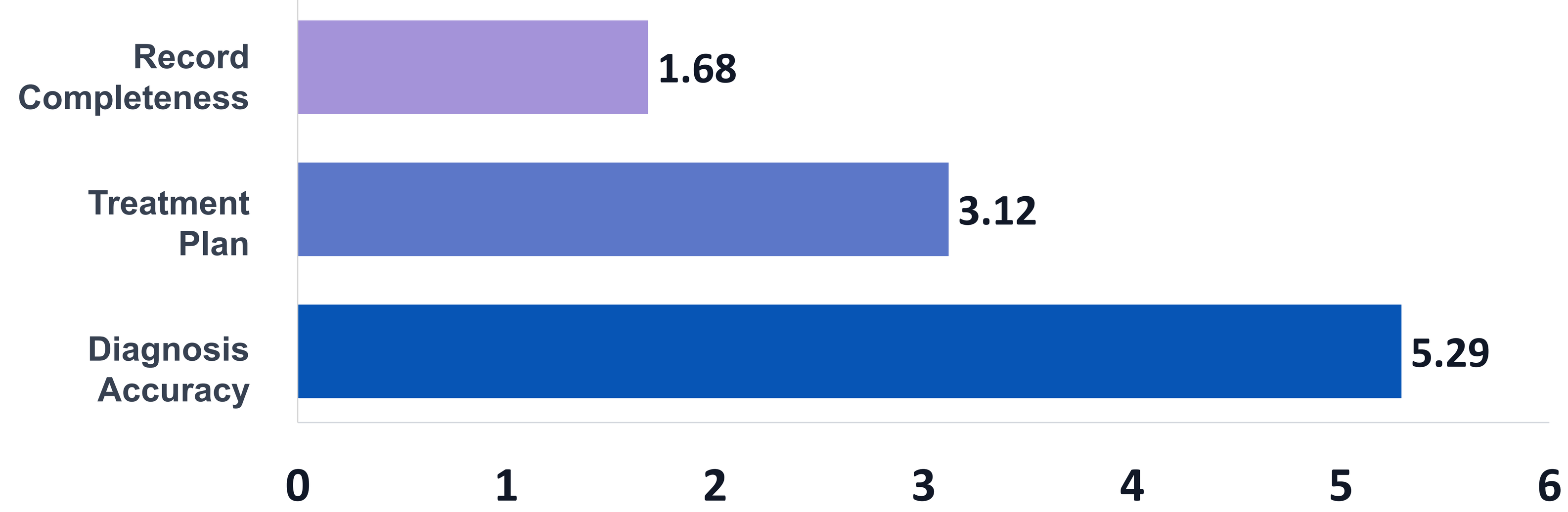
Dataset and training

- **MMDental** pairs 3D CBCT volumes with expert medical records covering tooth notation, diagnosis, treatment, medication, and follow-up.
- **Supervision**: the 3D model is trained on the 50 labeled cases; 200 additional unlabeled volumes are provided by the challenge.
- **Protocol**: five-fold patient-level cross-validation (40 train / 10 validation per fold) and model ensembling.

50	200	50
LABELED CASES	UNLABELED VOLUMES	VALIDATION CASES
supervised training	provided by the challenge	official evaluation

Quantitative results

Total Score 10.08 | 2nd place in MICCAI STSR 2026 Task 3



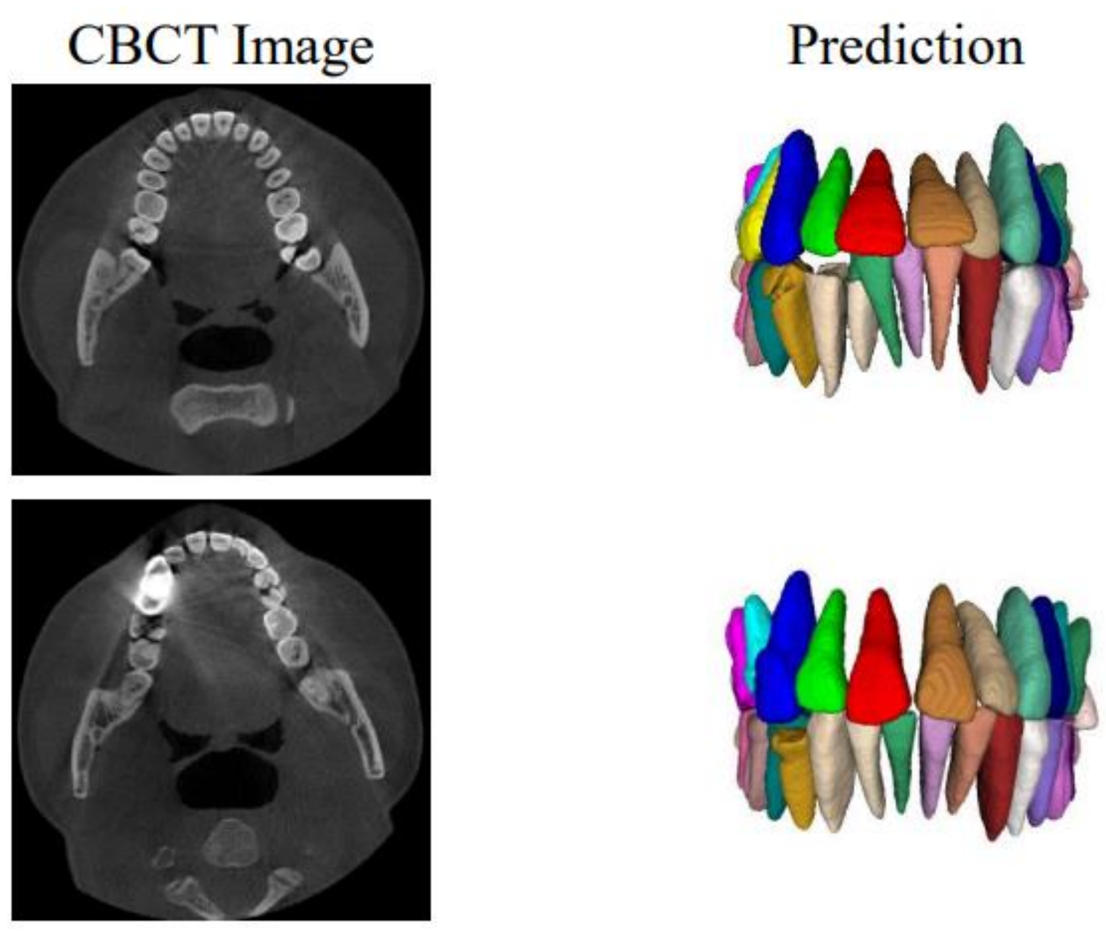


Fig. 2. CBCT slices and nnU-Net tooth-instance predictions.

CONCLUSION

- **Segmentation acts as a localization prompt**, preserving extra-tooth clinical context while aligning 32 FDI tooth positions.
- **Deterministic reporting**: multi-task predictions are converted into seven structured report fields through template-based rendering, without online GPT inference.
- **Global and tooth-level 3D evidence are complementary**: their fusion achieved a 10.08 total score and 2nd place in Task 3.

ACKNOWLEDGEMENTS

We thank the organizers of the MICCAI 2026 STSR Challenge for providing the MMDental benchmark dataset and evaluation platform. We also appreciate their prompt and helpful responses to our questions throughout the challenge.