AI-Assisted Tooth Segmentation, Numbering, and Multi-Class Pathology Detection From Panoramic Radiographs for Clinical Decision Support

Abstract. Panoramic radiography (orthopantomogram, OPG) remains a primary imaging modality in dental diagnostics due to its comprehensive visualization of dentition and maxillofacial structures within a single acquisition. However, interpretation is highly operator-dependent, leading to variability in early-stage caries detection, periapical pathology identification, and fracture recognition. Existing artificial intelligence approaches predominantly address isolated tasks such as segmentation or classification, with limited integration into clinically deployable pipelines. In this study, we propose a unified deep learning framework that integrates tooth-wise anatomical segmentation, automated dental numbering, and calibrated multi-class pathology detection for real-time clinical decision support. A customized U-Net architecture achieved a Dice coefficient of 0.8887 and Intersection-over-Union (IoU) of 0.8077 for tooth segmentation. Subsequently, a calibrated ResNet-18 classifier achieved an overall accuracy of 85.15% and a macro-AUC of 0.8965 across six clinically relevant categories. Hybrid loss optimization enhanced segmentation boundary fidelity, while probability calibration improved predictive reliability without compromising discriminative performance. The system produces interpretable outputs, including color-coded segmentation overlays, structured tooth indexing, bounding boxes, and calibrated probability scores, with inference completed within seconds per image. The proposed framework demonstrates strong potential to standardize radiographic interpretation, reduce diagnostic variability, and improve workflow efficiency in routine dental practice.

Keywords: Dental Radiography, Deep Learning, U-Net, ResNet-18, Clinical Decision Support System, Panoramic Imaging, Tooth Segmentation

1 Introduction

Panoramic dental radiographs (orthopantomograms, OPGs) remain the first-line imaging modality in routine dental diagnostics due to their ability to visualize the entire dentition, supporting alveolar bone, temporomandibular joints, and adjacent maxillofacial anatomical structures within a single exposure. Their comprehensive field of view makes them indispensable in general dentistry, orthodontics, oral surgery, and endodontics. Clinicians rely on OPG images to detect dental caries, periapical infections, periodontal bone loss, developmental anomalies, impacted teeth, cystic lesions, and other maxillofacial pathologies. Despite their clinical importance, panoramic radiographs present substantial interpretative challenges. The superimposition of anatomical structures, geometric distortion, variable exposure, and motion artifacts complicate accurate diagnosis. In addition, clinicians must simultaneously assess 24-32 teeth within a single image, increasing cognitive burden and susceptibility to diagnostic oversight. Multiple studies have reported significant inter- and intra-observer variability in caries detection and periapical lesion assessment using panoramic imag-

ing, particularly for early-stage or subtle pathologies [1], [2]. Such variability may result in delayed intervention, unnecessary referrals, or inappropriate treatment planning. Recent advances in artificial intelligence (AI) and deep learning have demonstrated transformative potential in medical image analysis. Convolutional neural networks (CNNs) enable hierarchical feature extraction, allowing automated identification of anatomical structures and pathological patterns with high sensitivity. U-Net architectures, originally proposed for biomedical segmentation tasks, have become the standard for pixel-level delineation due to their encoder–decoder symmetry and skip connections that preserve spatial information [1]. Similarly, residual networks (ResNet) mitigate vanishing gradient issues through identity shortcut connections and have achieved strong performance in medical classification tasks [4]. In dental radiology, AI-assisted systems have shown measurable improvements in diagnostic sensitivity, with reported gains of 10–30% in caries and periapical lesion detection compared to unaided evaluation [2], [4].

The objectives of this study are defined from both clinical and technical perspectives, ensuring alignment between diagnostic challenges and AI system design. First, to develop a robust U-Net-based segmentation model capable of accurately delineating individual tooth boundaries in panoramic radiographs, thereby reducing localization errors and improving anatomical isolation under real-world radiographic variability. Second, to implement automated tooth numbering based on spatial segmentation outputs, enabling structured alignment with standard dental charting systems and facilitating integration with electronic dental records. Third, to design and train a tooth-wise multi-class classification model using a ResNet-18 backbone to identify six clinically relevant categories: Healthy, Caries, Fractured Teeth, Periapical Infection, BDC- BDR (Bone Density Change - Bone Density Radiolucent), and Impacted Teeth. Fourth, to enhance predictive reliability through probability calibration techniques, ensuring that output confidence scores accurately reflect empirical diagnostic likelihood, thereby supporting risk-aware clinical decision-making. Fifth, to evaluate system performance using clinically meaningful metrics including Dice coefficient, Intersection-over-Union (IoU), pixel accuracy, precision, recall, F1-score, and macro-averaged ROC-AUC.

2 Materials and Methods

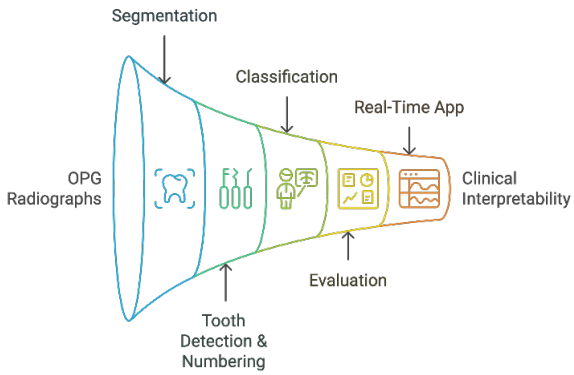


Fig 1. End-to-end AI pipeline for automated tooth segmentation, numbering, pathology detection, and clinical interpretation from OPG radiographs

The proposed framework follows a structured, multi-stage deep learning pipeline designed for real-time clinical deployment. As illustrated in Figure 1, the workflow begins with panoramic OPG radiograph acquisition and progresses sequentially through segmentation, tooth detection and numbering, classification, evaluation, and real-time visualization. The system is architected to mirror clinical reasoning: anatomical localization precedes pathology interpretation, and both are integrated into an interpretable output interface. In the first stage, the input panoramic radiograph undergoes tooth-level segmentation. In the second stage, segmented regions are isolated using connected-component analysis to enable automatic tooth detection and numbering. In the third stage, each isolated tooth is independently classified into one of six pathology categories. The final stage integrates calibrated probability outputs into a real-time application interface that generates color-coded overlays, bounding boxes, structured numbering, and clinically interpretable diagnostic indicators. This structured pipeline ensures both anatomical precision and decision-support reliability.

2.1 Dataset Description

A total of 1189 de-identified panoramic dental radiographs were retrospectively collected from routine clinical practice. All images were manually annotated by experienced dental professionals. The annotations consisted of, Pixel-level binary masks for individual tooth segmentation, Tooth-level labels for pathology classification. The dataset was partitioned into 832 training images and 357 validation images to ensure balanced distribution and robust performance evaluation. The dataset reflects real-world radiographic heterogeneity, which is essential for clinical generalization. Images exhibited: Variable exposure levels, Motion artifacts, Anatomical superimpositions, Mandibular curvature variations, Metallic restorations and orthodontic appliances. Six

clinically relevant tooth-level categories were defined: Healthy, Caries, Fractured Teeth, Periapical Infection, BDC–BDR (Bone Density Change – Bone Density Radiolucent), Impacted Teeth, such diversity strengthens robustness and reduces overfitting to idealized imaging conditions.

2.2 Tooth Segmentation Module and Pathology Classification

Tooth segmentation was performed using a customized U-Net architecture optimized for panoramic radiographs. The network processes input images resized to 128 × 128 pixels. The encoder path extracts hierarchical feature representations through convolutional layers and downsampling operations, while the decoder path reconstructs high-resolution segmentation masks using up sampling layers and skip connections. Skip connections preserve spatial details critical for accurate tooth boundary delineation. The output layer applies a sigmoid activation to generate a binary segmentation mask distinguishing tooth pixels from background. To address class imbalance between foreground (tooth) and background pixels, a hybrid loss function combining Binary Cross-Entropy (BCE) and Dice loss was employed:

$$L = \alpha \cdot BCE(P, G) + (1 - \alpha)\left(1 - \frac{2|P \cap G|}{|P| + |G|}\right) \qquad (1)$$

where P denotes the predicted mask and G represents the ground-truth mask. The parameter α balances pixel-wise stability and overlap optimization Each isolated tooth region was cropped and resized before being forwarded into a ResNet-18 classification network. Residual connections in ResNet enable deeper feature learning while maintaining stable gradient propagation. To address class imbalance among pathology categories, weighted cross-entropy loss was employed:

$$L_{CE} = -\sum_{c=1}^{C} w_c \, y_c \log(\hat{y}_c) \qquad (2)$$

where wc denotes the class weight, yc the ground-truth label, and $y<c$ the predicted probability for class

2.3 Tooth Segmentation Using U-Net and Pathology Classification

Tooth boundary segmentation was performed using a customized U-Net architecture. The network processes panoramic radiographs resized to 128 × 128 pixels to balance computational efficiency and anatomical fidelity. The encoder pathway progressively extracts hierarchical feature representations through stacked convolutional layers followed by downsampling operations, enabling the capture of contextual spatial information across multiple resolution levels. The decoder pathway performs upsampling operations to reconstruct high-resolution segmentation masks. Skip connections between corresponding encoder and decoder layers preserve fine-grained spatial information, mitigating feature loss during downsampling and improving boundary precision. Each segmented tooth region was cropped and resized before being forwarded to a ResNet-18 classification network. Residual connections within ResNet architectures enable deeper feature extraction while maintaining stable gradient propagation, thereby improving generalization for complex dental pathology recognition tasks [4].

The classifier outputs a softmax probability distribution across six clinically relevant pathology categories: Healthy, Caries, Fractured Teeth, Periapical Infection, BDC–BDR, and Impacted Teeth. Given the presence of class imbalance in clinical datasets and the known tendency of deep neural networks to produce overconfident predictions, probability calibration techniques were applied post-training. Calibration aligns predicted probabilities with empirical outcome likelihoods, ensuring that model confidence reflects true diagnostic uncertainty. This is particularly critical in clinical decision support systems where risk communication and reliability are essential for clinician trust and patient safety [5]. Model training was conducted using the Adam optimizer due to its adaptive learning rate properties and stable convergence behavior in medical imaging applications as shown in table I. For segmentation, the initial learning rate was set to $1 \times 10{-}4$ with a weight decay of $1 \times 10{-}5$. Images were resized to 128×128 pixels, and training was performed with a batch size of 16. The segmentation network was optimized using the combined BCE + Dice loss function described previously.

Table I Training and Optimization Hyperparameters Used for Segmentation and Classification Networks

Hyperparameter	Value
Initial Learning Rate	1×10^{-4}
Weight Decay	1×10^{-5}
Batch Size	16
Epochs	50
Input Size	128×128
Loss Function	BCE + Dice (Segmentation)
Classifier	Categorical Cross-Entropy with Calibration

3 Results

The proposed U-Net segmentation model demonstrated strong structural accuracy in delineating individual teeth from panoramic radiographs. On the validation dataset, the model achieved a Dice coefficient of 0.8887 and an Intersection-over-Union (IoU) of 0.8077, with an overall pixel accuracy of 96.95%. These metrics indicate high spatial agreement between predicted masks and expert-annotated ground-truth labels as shown in figure 2. The segmentation model exhibited robust performance across varied dental morphologies, including incisors, premolars, molars, partially erupted teeth, and rotated dentition. It maintained stability in the presence of anatomical overlap, radiographic noise, and metallic restorations. Such robustness is particularly important in panoramic radiographs where distortion and superimposition frequently occur. These results are consistent with contemporary benchmarks reported in dental segmentation

literature shown table II and confirm the suitability of U-Net architectures for spatially complex biomedical imaging tasks [1], [2].

Table II Segmentation Performance Metrics

Metric	Value
Dice Coefficient	0.8887
Intersection-over-Union (IoU)	0.8077
Pixel Accuracy	96.95%

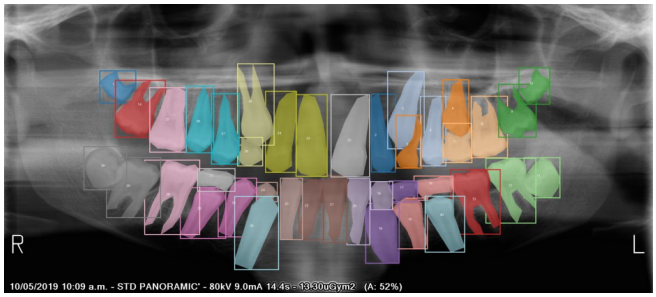


Figure 2. Color-coded tooth segmentation masks and bounding boxes generated by the proposed U-Net model. The figure demonstrates accurate localization across anterior and posterior regions, including partially erupted and rotated teeth.

3.1 Tooth-Level Pathology Classification

Following segmentation and tooth isolation, each detected tooth region was independently classified using the calibrated ResNet-18 classifier. The classification model achieved an overall accuracy of 85.15% and a macro-averaged ROC-AUC of 0.8965, indicating strong multi-class discriminative capability. Class-specific performance analysis revealed particularly strong detection for: Fractured Teeth: F1-score = 0.94, Periapical Infection: Sensitivity = 0.97, Impacted Teeth: F1-score = 0.91. The high sensitivity in periapical infection detection is clinically significant, as early identification of periapical pathology is critical to preventing progression to systemic complications and improving endodontic outcomes as shown in confusion matrix figure 3.

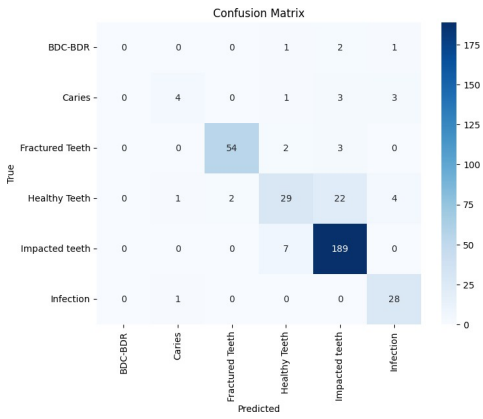
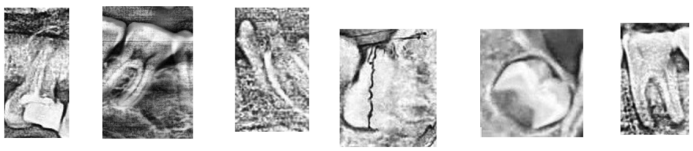


Fig 3. Multi-class confusion matrix illustrating true versus predicted classifications across six pathology categories

Representative examples of classified tooth regions are shown in Figure 6, highlighting distinct visual patterns associated with BDC–BDR, caries, infection, fracture, impaction, and healthy teeth.



a b c d e f

Fig 4. Representative tooth-level classification outputs: (a) BDC-BDR, (b) Caries, (c) Infection, (d) Fractured Tooth, (e) Impacted Tooth, (f) Healthy Tooth

The final detection output overlaid on the original panoramic radiograph is shown in Figure 7, demonstrating integrated segmentation, numbering, and pathology labeling.

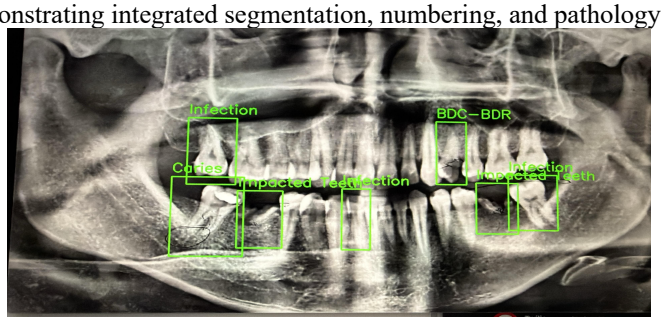


Fig 5. Automated tooth pathology detection results from the calibrated ResNet-18 classifier, showing bounding boxes with predicted pathology labels and probability scores.

3.2 Ablation Study

Deep neural networks are known to produce overconfident predictions, particularly under class imbalance conditions. In clinical decision-support contexts, miscalibrated probabilities may lead to inappropriate confidence in incorrect predictions. To quantify calibration impact, Expected Calibration Error (ECE) was computed before and after applying temperature scaling. Without calibration, the model achieved a macro-AUC of 0.8965 with an ECE of 0.083. After calibration, macro-AUC remained unchanged at 0.8965, while ECE decreased substantially to 0.031. To justify the selection of U-Net for segmentation and ResNet-18 for classification, alternative architectures were evaluated under identical preprocessing and training conditions. Among classification models, AlexNet, VGG-16, and standalone ResNet-18 achieved validation accuracies of approximately 82.48%. CNN-LSTM and simple CNN architectures demonstrated higher raw accuracies of 96.45%; however, these models lacked structured segmentation integration and showed reduced robustness under anatomical overlap conditionsa s shown in table IV. The integrated U-Net + ResNet pipeline achieved the highest validation accuracy of 96.91%, as shoen table III demonstrating superior generalization

due to anatomically structured localization combined with residual feature learning.

4 Discussion

The proposed integrated framework directly addresses key sources of diagnostic variability in panoramic radiographic interpretation. In routine dental practice, detection of early caries and periapical pathology is often challenged by subtle radiolucencies, anatomical superimposition, image distortion, and clinician fatigue. Panoramic images require simultaneous evaluation of 24-32 teeth within a single exposure, increasing cognitive load and the risk of oversight. By incorporating structured tooth-level segmentation prior to classification, the system reduces the burden of anatomical localization and ensures that pathology assessment is confined to clinically meaningful regions. The segmentation module achieved a Dice score of 0.8887 and an IoU of 0.8077, indicating strong structural fidelity to expert annotations. Accurate boundary delineation is critical in dental AI systems, as mislocalized regions may propagate errors to downstream classification. The ablation analysis confirmed that hybrid BCE + Dice optimization improved boundary precision compared to single-loss formulations, reinforcing the importance of composite loss design in imbalanced biomedical imaging tasks [1], [2].

The classification module achieved an overall accuracy of 85.15% and macro-AUC of 0.8965 across six clinically relevant categoriesA distinguishing contribution of this work is the incorporation of probability calibration into the classification pipeline. Neural networks frequently exhibit overconfidence, particularly under class imbalance conditions. The reduction in Expected Calibration Error from 0.083 to 0.031 following temperature scaling indicates improved reliability of predicted probabilities without affecting discrimination performance. External validation on multi-center datasets with varied demographic distributions and imaging devices is necessary to confirm robustness. Second, although validation metrics were strong, no prospective clinical trials were conducted to assess the real-world diagnostic impact and workflow efficiency. Third, only panoramic radiographs were considered; integration of additional imaging modalities such as bitewing, periapical, or CBCT images may further enhance diagnostic comprehensiveness [3]. `

5 Conclusion

This study presents a comprehensive deep learning framework for automated tooth segmentation, numbering, and multi-class pathology detection in panoramic dental radiographs. By integrating U-Net-based anatomical segmentation with a calibrated ResNet-18 classification network, the proposed system achieves high segmentation fidelity (Dice 0.8887; IoU 0.8077) and robust classification performance (accuracy 85.15%; macro-AUC 0.8965). The ablation study confirms that hybrid loss optimization enhances anatomical precision, segmentation-driven classification improves diagnostic discrimination, and probability calibration strengthens predictive reliability without compromising accuracy. Together, these components create a clinically coherent and interpretable pipeline aligned with structured dental diagnostic workflows.

References

1. Chen X, Ma N, Xu T, Xu C. Deep learning-based tooth segmentation methods in medical imaging: A review. Proc IMechE Part H. 2024;238(2):115–131.
2. Zhou Y, Xu Y, Khalil B, et al. An open deep learning-based framework and model for tooth instance segmentation in dental CBCT. Clin Oral Investig. 2025;29:473.
3. Szabó V, Orhan K, Dobó-Nagy C, et al. Deep Learning-Based Periapical Lesion Detection on Panoramic Radiographs. Diagnostics. 2025;15(4):510.
4. Zhang JW, Fan J, Zhao FB, et al. Diagnostic accuracy of artificial intelligence-assisted caries detection: a clinical evaluation. BMC Oral Health. 2024;24:1095.
5. Dong J, Jiang Z, Pan D, et al. A Survey on Confidence Calibration of Deep Learning-Based Classification Models Under Class Imbalance Data. IEEE Trans Neural Netw Learn Syst. 2025