Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence for Automated Detection and Segmentation of Pulp Stones in Cone Beam Computed Tomography Scans.

Sally Mansour

sally.mansour@dentistry.cu.edu.eg

Faculty of dentistry Cairo university https://orcid.org/0009-0008-6257-1753

Hanaa Mansy

Faculty of dentistry Cairo university

Sabah Sobhy*

Al-Azhar University,

William Scarfe

University of Louisville School of Dentistry

Mohamed Elsharkawy

J.B. Speed School of Engineering, University of Louisville, Louisville, KY 40292, USA

Yara Helaly

Faculty of Dentistry, Cairo University

Article

Keywords: pulp stones, cone-beam computed tomography, semantic segmentation, deep learning, DeepLabV3+, endodontics

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11554539/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is no conflict of interest

Manuscript

1. Introduction

The dental pulp is a vascularized and innervated connective tissue occupying the pulp chamber and root canal complex of human teeth. In response to physiological stimuli such as natural aging, occlusal trauma, carious lesions, and chronic low-grade inflammation, the pulp may undergo isolated dystrophic mineralization, resulting in calcified nodules, termed pulp stones.[1,2] Pulp stones can occur as free, adherent, or embedded nodules within the coronal or radicular pulp, with a reported prevalence of 8% to 95%, depending on population, imaging modality used, and diagnostic criteria.[3,4,5]

Clinically, pulp stones are significant because they can obstruct endodontic access, impede access to the root canal orifice, and increase the risk of instrument separation or perforation during root canal treatment.[6,7] Radiography is the main, non-invasive means of identifying these calcifications prior to treatment. Conventional two-dimensional radiography including bitewing and periapical techniques is limited due to superimposition of overlying anatomic structures, projection geometry distortion, and inability to detect calcifications smaller than approximately 200 μm.[8,9] Cone-beam computed tomography (CBCT) overcomes these limitations by providing distortion-free, high-resolution, three-dimensional volumetric data and enabling multiplanar reconstruction, substantially improving the sensitivity and specificity of pulp stone detection compared with two-dimensional techniques. [10,11]

However, interpretation of CBCT datasets specifically for pulp stone detection is time-consuming, requires training, and is subject to both intra- and inter-observer variability. With increasing availability and use of CBCT in dental practice, automated and reproducible analytical tools are needed to support the clinician. Artificial intelligence (AI), particularly convolutional neural network (CNN)-based semantic segmentation, has demonstrated strong performance for dental image analysis tasks, including detection of caries, periapical pathosis, periodontal bone loss, and various anatomical structures.[12,13,14,15]

Semantic segmentation assigns a class label to every pixel in an image, providing a spatially complete representation of the target anatomy which makes it superior for segmentation of small objects, like pulp stones. Encoder–decoder architectures are the dominant framework for this task in medical imaging. The original U-Net introduced skip connections enabling recovery of fine spatial detail during

upsampling. [16] Subsequent architectures (Attention U-Net [17], U-Net++ [18], and DeepLabV3+ [19]) have extended this paradigm through attention gating, nested dense skip pathways, and atrous spatial pyramid pooling, respectively.

Previous studies have applied a range of artificial intelligence (AI) approaches for pulp stone detection in two-dimensional dental radiographs. Pulp stone identification was investigated as a classification task using periapical radiographs [20],while in different studies, the problem was formulated as object detection on bitewing and panoramic radiographs, respectively. [21,22] Gurhan et al. [23] adopted a two-stage framework consisting of tooth localization followed by classification to identify pulp stones in panoramic images. Only a limited number of studies have approached the task through image segmentation, employing Mask R-CNN and U²-Net, respectively, to delineate pulp stone boundaries on panoramic radiographs. [24,25]

Despite the growing body of research on AI-based pulp stone detection in two-dimensional dental imaging modalities (Table 1) [5,20,21,22,23], the application of AI to three-dimensional imaging remains largely unexplored. To date, only Mahabob et al. [11] have utilized cone-beam computed tomography (CBCT) for automated pulp stone detection, highlighting a significant gap in the current literature. They did so through demographic clustering rather than image-based analysis, or semantic segmentation. Furthermore, no author has systematically compared multiple deep learning architectures under standardized experimental conditions for this specific task. The present study addresses this gap by (i) evaluating four established segmentation architectures on annotated CBCT datasets and (ii) introduces a new architecture, DACA-DeepLab.

2. Materials and Methods

2.1 Ethical Considerations

This retrospective study was conducted in accordance with the Declaration of Helsinki (revised 2013) and received approval from the Research Ethics Committee of Faculty of Dentistry, Cairo University on 25/11/2025. All CBCT examinations had been acquired for routine diagnostic or treatment-planning purposes prior to this study. Informed consent was obtained from all patients for being included in the study.

2.2 Study Sample

Power analysis was designed to have adequate power to apply a two-sided statistical test of the null hypothesis that results of deep learning model are as accurate as a radiologist's opinion. Sample size calculation was performed using Statskingdom sample size calculator. By adopting a (95%) confidence interval and by using a sensitivity value of 80% of the DL group [21] and 100 % for the ground truth, the sample size calculated based on sensitivity was 385 image.

2.3 CBCT Data Acquisition

CBCT examinations were retrieved from the CBCT database at the Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Cairo University in addition to a private dentomaxillofacial radiology in Cairo, Egypt.

A systematic retrospective search of the CBCT database identified 112 eligible examinations (50 male, 62 female; age range, 16 to 65 years). All the scans demonstrated fully erupted permanent teeth with complete root formation, and complete volumetric data for the tooth of interest. Exclusion criteria included deciduous teeth, permanent teeth exhibiting developmental anomalies, external or internal root resorption, prior root canal treatment, post-restoration, or root caries. Additionally, CBCT images of suboptimal quality or with significant beam hardening artifacts were excluded.

All scans were acquired using a Planmeca Promax 3D MID CBCT machine, operating at 90 kVp, 6.3 to 8 mA, with an exposure time 12.5s. Scans had a voxel size of 200 μm and variable field-of-view (FOV) of 5 × 5-5 × 8 cm. 523 images with (n = 305) and without (n = 218) pulp stones were included.

2.4 Image Annotation

CBCT scans were imported into Romexis 6.0 (Planmeca, Finland) software for data review. The assessment was performed on para-coronal and para-sagittal planes oriented along the long axis of the tooth by two oral and maxillofacial radiologists (OMRs), each with more than 10 years of experience in CBCT interpretation. By consensus (YH, SM), the presence or absence of pulp stone(s) was noted by navigating the software. Finally, the most representative consecutive coronal and sagittal slices

through the annotated images were exported as PNG files at the original CBCT resolution for use as model input.

All the saved images uploaded to Roboflow (Version 1.0, USA) [26], a Cloud based platform designed to manage the entire computer vision workflow, including image annotation, dataset organization, preprocessing, and model training and deployment. On the platform, manual annotation was performed using the polygonal bounding tool to precisely outline each pulp space and pulp stone. Each exported slice was labelled with one of three classes (Class 0, background; Class 1, pulp cavity and; Class 2, pulp stone). When present, pulp stones (Class 2), are categorized within the pulp cavity group (Class 1) according to their anatomical location (Figure 1). The remaining cases were classified as pulp cavity only, without the presence of pulp stones. This decision preserved the clinical semantic structure of the segmentation task, consistent with the primary aim of delineating pulp-containing spaces.

Annotation was performed through a consensus-based protocol: both radiologists and the endodontist reviewed each cross-sectional slice jointly and assigned the final class boundaries by mutual agreement. In cases of initial disagreement regarding the precise extent of the pulp cavity and pulpstone boundary, the slice was re-examined in conjunction with adjacent axial, coronal, and sagittal reconstructions until a consensus annotation was reached. This approach was adopted to ensure that the ground truth labels used for model training and evaluation reflected single and standard interpretations,

Finally, all the annotated images and original images were saved in PNG format, in two separate folders. Subsequently, all files were shared with a computer science engineer for further analysis.

2.5 Deep Learning Protocols

This study pursues two objectives. First, four established encoder–decoder architectures — U-Net, Attention U-Net, U-Net++, and DeepLabV3+ — are trained and evaluated under a common protocol to provide a reproducible benchmark for CBCT-based segmentation of the dental pulp space. Second, we propose DACA-DeepLab (Dual-Axis Channel-Aware Attention DeepLabV3+), a network built specifically to recover small, low-contrast calcified structures such as pulp stones. DACA-DeepLab retains the macro-architecture of DeepLabV3+ — a feature-extracting encoder, an Atrous Spatial Pyramid Pooling (ASPP) context module, and a decoder that fuses semantic context with low-level spatial detail — and introduces three components: a Pyramid Channel-Spatial Attention (PCSA)

module in the encoder, a Dual-Axis Cross-Attention (DACA) decoder, and a Boundary-Focus Refinement (BFR) head supervised by a Sobel-based boundary loss. Computational details of all used models, along with their supporting equation, are described in Appendix A.

2.6 Dataset and Preprocessing

Each example consists of a CBCT cross-sectional image paired with an expert color-coded annotation. Ground-truth masks were derived by converting each annotation to HSV color space and thresholding the red and blue regions; morphological closing followed by contour filling produced solid regions, assigned to the pulp cavity (class 1, red) and the pulp stone (class 2, blue), with all remaining pixels as background (class 0). Images and masks were resized to 256 × 256 pixels — nearest-neighbor interpolation for masks to preserve discrete labels — and intensities scaled to [0, 1].

Data augmentation was applied on the fly to the training subset only and comprised random horizontal and vertical flips, random 90°/180°/270° rotations, brightness and contrast jitter, and additive Gaussian noise. The validation and test subsets were never augmented, so that model selection and final evaluation reflect performance on unmodified images.

2.7 Data Splitting

The labelled dataset was partitioned into three mutually exclusive subsets using a fixed random seed so the partition is identical across every model and can be reproduced exactly. The split was implemented in two stages: the full set of image–annotation pairs were divided into a held-out test set and a development set; the development set was then divided again into training and validation subsets. Applying the same seed at both stages guarantees that all architectures are trained, tuned, and evaluated on the same images, a prerequisite for a fair comparison. The three subsets play deliberately separate roles, as summarized in Table 1.

Table 1. Distribution and Role of each Data Subset.

Subset	Proportion	Role in the experiment

Training	70%	Parameter updates; augmented on the fly at each epoch
Validation	15%	Checkpoint selection and early stopping; never augmented, never used for updates
Test	15%	Held out entirely until training is complete; source of all reported metrics

2.8 Training Configuration

The proposed model was implemented in PyTorch and trained on a single NVIDIA GPU, with the full training configuration summarized in Table 2. The four comparison models were trained on the same data split, input resolution, batch size, and epoch budget using the Adam optimiser and a combined unweighted cross-entropy–Dice loss, without class weighting or boundary supervision. All DACA-DeepLab variants, including the DeepLabV3+ baseline within the second study, were trained under the full DACA-DeepLab configuration: AdamW, inverse-frequency class weighting, and the combined boundary loss.

Table 2. Training Configuration for DACA-DeepLab.

Setting	Value
Framework	PyTorch, single NVIDIA GPU
Mixed precision	Automatic mixed precision (autocast + gradient scaling) on GPU
Reproducibility	Fixed random seed (42); deterministic cuDNN
Input resolution	256 × 256 pixels, RGB, scaled to [0, 1]
Classes	3 (background, pulp cavity, pulpstone)
Batch size	8
Optimiser	AdamW (weight decay = 1×10^{-4})
Initial learning rate	1×10^{-3}
LR schedule	Cosine annealing with warm restarts ($T_0 = 10$, T_mult = 2)
Maximum epochs	50
Early stopping	Patience 10 epochs on validation mean Dice (min. $\Delta = 1 \times 10^{-4}$)

Class weighting	Inverse-frequency weights on the cross-entropy term
Loss (DACA-DeepLab)	Weighted CE + soft Dice + $\lambda \cdot$ boundary, $\lambda = 0.3$
Model selection	Checkpoint at highest validation mean Dice (pulp + canal)

Aligning every remaining hyperparameter across the comparison models and the proposed model — optimiser family, learning-rate schedule, class weighting, and early stopping — is recommended so that reported differences can be attributed unambiguously to architecture.

2.9 Evaluation Metrics

Segmentation performance was quantified per class using true positives (TP), false positives (FP), and false negatives (FN) of the predicted mask against the ground truth, from which the Dice Similarity Coefficient (Eq. A13), Intersection-over-Union (Eq. A14), precision (Eq.A15), and recall (Eq. A16) were calculated, following standard definitions used in medical image segmentation evaluation . [30] Mean Dice (mDice) summarized performance across the tissue classes, excluding background (Eq.A17), and was supplemented by class-wise ROC curves with AUC, normalized confusion matrices, and per-sample Dice distributions, all computed on the held-out test set using the best-checkpoint weights. Full equations for all reported metrics are provided in Appendix A.

3. Results

This section reports segmentation performance across two separate experiments. Experiment 1 evaluates four established encoder–decoder architectures under a common training protocol using the Adam optimiser and an unweighted combined cross-entropy–Dice loss. Experiment 2 presents an incremental additive study of the proposed DACA-DeepLab architecture, trained with AdamW, inverse-frequency class weighting, and an additional Sobel-guided boundary loss. Because the two experiments differ in training configuration, their results are reported and interpreted separately; direct cross-experiment numerical comparisons are not made.

3.1 Benchmark Comparison of Established Architectures

Table 3 reports the per-class and overall segmentation performance of U-Net, Attention U-Net, U-Net++, and DeepLabV3+ on the held-out test set. All metrics were computed using the best-checkpoint

weights selected based on highest validation mean Dice. Final performance per class and training convergence for the four baseline architectures are shown in Figures 2. Representative qualitative outputs for the four baseline architectures are shown in Figure 3.

Table 3. Segmentation performance per-class of the four baseline architectures on the held-out test set.

Model	Class*	Dice	IoU	Precision	Sensitivity†	Specificity	F1	Accuracy‡	AUC	mDice§
U-Net	0	0.993	0.986	0.992	0.993	0.855	0.993	0.987	0.994	0.762
	1	0.822	0.698	0.830	0.815	0.992	0.822	0.984	0.991	
	2	0.702	0.541	0.713	0.692	0.999	0.702	0.997	0.993	
Attention U-Net	0	0.992	0.985	0.992	0.992	0.853	0.992	0.985	0.992	0.742
	1	0.808	0.678	0.807	0.808	0.991	0.808	0.983	0.988	
	2	0.677	0.511	0.680	0.673	0.999	0.677	0.997	0.991	
U-Net++	0	0.993	0.986	0.993	0.993	0.872	0.993	0.987	0.994	0.759
	1	0.827	0.705	0.825	0.829	0.992	0.827	0.984	0.990	
	2	0.692	0.529	0.686	0.698	0.999	0.692	0.997	0.985	
DeepLab V3+	0	0.994	0.987	0.994	0.994	0.878	0.994	0.988	0.996	0.775
	1	0.838	0.721	0.835	0.841	0.992	0.838	0.985	0.994	

2	0.711	0.552	0.723	0.700	0.999	0.711	0.997	0.997

*0, background; 1, pulp cavity; 2, pulp stone: Sensitivity = Recall: ‡Accuracy = one-vs-all pixel accuracy per class. §mDice = mean Dice over the pulp cavity and pulp stones classes with background excluded.

3.2 Study of DACA-DeepLab

Table 4 presents the incremental results on the held-out test set, comparing the performance of DeepLabV3+ baseline trained under the same AdamW and class-weighted loss configuration with PCSA, DACA, and BFR. ROC performance for pulp cavity and pulp stone(s) between DeepLabV3+ (0.985 / 0.996), + PCSA (0.988 / 0.998), + PCSA + DACA (0.990 / 0.997) and, Full DACA-DeepLab (0.987 / 0.995) is not statistically significant. Dice distribution per-sample for the DACA-DeepLab variants is shown in Figure 4. Figure 5 demonstrates the convergence in accuracy as a function of passes through the training data (epoch). Qualitative predictions across the second study variants and the corresponding encoder attention patterns are shown in Figures 6 and 7.

Table 4. Second study results for DACA-DeepLab on the held-out test set. Metrics are reported as foreground-averaged values across pulp cavity and pulp stone(s).

Model Variant	Dice	IoU	Precision	Sensitivity	Specificity	F1	Accuracy	AUC	mDice
DeepLabV3+	0.593	0.429	0.485	0.783	0.985	0.593	0.967	0.990	0.606
+ PCSA	0.663	0.500	0.551	0.849	0.987	0.663	0.972	0.993	0.685
+ PCSA + DACA	0.687	0.527	0.623	0.766	0.989	0.687	0.975	0.993	0.733
Full DACA-DeepLab	0.628	0.463	0.529	0.772	0.985	0.628	0.968	0.991	0.647

Foreground metrics are averaged over Class 1 (pulp cavity) and Class 2 (pulp stone), excluding background. mDice = mean Dice over the two tissue classes.

4. Discussion

This study represents the first systematic comparison of four encoder–decoder deep learning architectures (U-Net, Attention U-Net, U-Net++, and DeepLabV3+) for semantic segmentation of dental pulp cavity and pulp stone(s) in CBCT cross-sectional images, alongside a new architecture, DACA-DeepLab, designed to improve boundary-level delineation of calcified pulp structures. This approach was intentional. The initial aspect of the investigation benchmarks established architectures under a standard protocol, while the second study isolates the contribution of each proposed component under a unified imbalance-aware regime. Results from the two experiments were therefore interpreted independently. Aligning the hyperparameters (optimiser family, learning-rate schedule, class weighting, and early stopping) across the comparison models and the proposed model was necessary so that reported differences can be related to specific architectures.

4.1 Clinical Implications and Methodology

Automated segmentation of root canal anatomy on CBCT data has clinical significance for pre-operative treatment planning. Identifying the location and extent of pulp stones prior to root canal treatment allows the clinician to anticipate access difficulty, select appropriate ultrasonic or modified file techniques for stone removal, and consider specialist referral.[7] Integration of a validated segmentation model into the CBCT reporting workflow would also support quantitative longitudinal tracking of calcification progression, which is currently infeasible with manual methods given the time required.

In the present study, CBCT volumes were converted into standardized coronal and sagittal slices—each tooth individually reoriented perpendicular to its long axis because these planes best visualize the pulp chamber and root canal system, closely resemble conventional periapical radiographs, and allow multiple images to be generated per scan, substantially expanding the limited annotated CBCT training data available. [27,28,29,30] This slice-based approach preserves sufficient anatomical detail for accurate deep learning while avoiding the greater computational demands of full 3D volumetric analysis. [27,31,32]

Detection through semantic segmentation was chosen, rather than bounding boxes, because pulp stones are small, morphologically variable calcified structures whose clinical relevance depends not only on their presence but on their precise boundaries and spatial relationship to the surrounding pulp space;

information a bounding box cannot capture. Segmentation performs pixel-level classification, enabling this anatomical detail to inform future quantitative analysis and clinical interpretation, consistent with its established role in computer-aided diagnosis. [33] This approach also provides a finer level of spatial information than region-based or bounding-box annotation, allowing the extent and boundaries of lesions to be characterized more precisely. [34]

The quality of ground-truth annotations is a critical determinant of supervised deep learning performance, particularly in semantic segmentation where models learn pixel-level representations of anatomical structures and their boundaries.[35,36] Roboflow was selected in the present study in accordance with previous studies , [37,38] as it provides an integrated platform for image annotation, dataset management, preprocessing, and export within a standardized workflow. This minimized data-handling errors while ensuring consistency throughout dataset preparation. In addition, it supports polygon-based manual annotation, enabling precise outlining of both the pulp cavity and pulp stones. This capability was particularly valuable because for the present study because pulp stones are small, irregularly shaped calcifications whose boundaries directly influence segmentation accuracy. Furthermore, the platform facilitated consensus-based annotation by multiple experts, allowing refinement of annotations before establishing the final ground truth. [37,38]

Following recommendation of the contributing computer scientist, the final sample size should increase to more than 500 annotated images. This upward adjustment was made because deep convolutional neural networks require enough training examples to learn discriminative feature representations without overfitting, regarding data splitting, we considered that the partition was performed at the level of individual image slices rather than at the patient level; where multiple slices originate from the same tooth or patient, a patient-level (grouped) split is preferable to remove any residual risk of information leakage.

4.2 Artificial intelligence performance

Across the four established architectures, DeepLabV3+ achieved the strongest overall segmentation performance (mDice = 0.775), followed closely by U-Net (0.762) and U-Net++ (0.759), with Attention U-Net weakest (0.742). All four models identified background and pulp cavity tissue reliably, but performance dropped noticeably for the pulp stone class across the board (Dice range 0.677–0.711).

This pattern is clinically expected: pulp stones are small, often isolated calcified foci that vary considerably in size, shape, and position within the pulp cavity, and they can blend into the surrounding dentin walls on CBCT with only subtle contrast differences. Because they occupy a small minority of pixels relative to the surrounding pulp cavity, the models also had comparatively less data from which to learn their appearance, compounding the difficulty. The consistency of this drop across all four architectures suggests the challenge lies primarily in the nature of the target structure itself rather than in any one model's design.

The second experiment, which built the proposed DACA-DeepLab model step by step, showed a related pattern. Adding the pyramid attention module alone improved performance over the baseline DeepLabV3+, and combining it with the cross-attention decoder produced the best result in this experiment (mDice = 0.733), reflecting a genuine benefit from letting the model weigh fine local detail against the broader shape of the pulp cavity when isolating small calcifications. Performance declined, however, once the boundary-refinement component was added to complete the full DACA-DeepLab model (mDice = 0.647). The most likely explanation is that this component, designed to sharpen edge detection, is especially sensitive to the low-contrast, often indistinct margins that separate a pulp stone from adjacent pulp tissue; rather than clarifying the true boundary, it may have reinforced spurious edges arising from image noise or CBCT artifact, particularly given the still-limited training set for a model of this added complexity. This points to a need for further evaluation of this component on a larger, more varied dataset before it can be considered an improvement over the simpler PCSA + DACA configuration.

From a clinical standpoint, these findings support DeepLabV3+, and its PCSA + DACA-enhanced variant, as the more dependable options for automated identification and outlining of pulp stones on CBCT. Because segmentation delineates the stone's boundaries and its relationship to the surrounding pulp cavity rather than simply flagging its presence, it offers information relevant to pre-endodontic treatment planning, where unrecognized pulp stones can obstruct canal negotiation and complicate access cavity preparation. That said, pulp stone segmentation accuracy across all models remains moderate rather than high, so these tools are best positioned as an assistive second check for the radiologist rather than a substitute for direct visual interpretation. Given how easily small, low-contrast pulp stones can be overlooked on routine review, the clinical value of automated segmentation lies less

in achieving perfect boundaries and more in consistently drawing attention to calcifications that might otherwise be missed, supporting rather than replacing the radiologist's judgment.

4.3 Comparison with Published Literature

While most of the previous studies dealt with 2d images, the U2-net segmentation study of Firincioglulari et al. [25] is the most methodologically similar to the present work, reporting a pulp stone segmentation Dice of 0.759 and IoU of 0.686 on 713 panoramic images. Our DeepLabV3+ model achieved a comparable pulp cavity Dice (0.838) and a slightly lower pulpstone Dice (0.711), which is expected given the considerably more detailed anatomical information required at the voxel level in CBCT cross-sections compared with projected panoramic images.

Compared with Mahabob et al. [11], whose aim was to uncover population-level pulp stone prevalence phenotypes using demographic and categorical variables, the present study aimed instead to directly detect and delineate pulp stones from CBCT image data itself. Methodologically, Mahabob et al.[11] applied AI (K-Means clustering, PCA) only after manual visual identification of pulp stones by a single radiologist across 300 scans with simple presence/absence labeling, whereas the present study used supervised deep-learning segmentation trained on densely, pixel-level annotated CBCT slices (523 slices from 112 scans, three-expert consensus), so AI was applied directly to the images rather than to pre-extracted labels. Their main results identified molars and older age (41–50 years) as significant predictors of pulp stone prevalence (69.0% of affected teeth were molars, $P = 0.036$), while our main result was that DeepLabV3+ achieved the best segmentation performance (mean Dice = 0.775) among the tested architectures—so although the outcome types differ (prevalence/clustering versus segmentation accuracy) and are not directly comparable, both studies converge on tooth type and age as clinically important factors worth exploring together in future work.

4.4 Limitations and Future Directions

Several limitations should be acknowledged. Although the dataset was acquired from two independent imaging locations, they both used the same CBCT device, which may limit generalizability to images from different manufacturers. Our study used a three-class annotation scheme; future work with a sufficiently large and densely annotated dataset should investigate four-class segmentation (background, pulp cavity, pulp stone) to enable stone-specific volumetric quantification. Our evaluation

used two-dimensional cross-sectional screen shots. Future studies should be extended to fully volumetric three-dimensional architectures (e.g., 3D U-Net, V-Net). Prospective observer studies comparing AI-assisted and unassisted radiologist performance are required before clinical deployment, and multi-centre validation across diverse demographics and imaging platforms is essential to establish the generalizability of the present findings.

5. Conclusion

This study concludes that deep learning-based semantic segmentation can accurately delineate dental pulp, pulp stone and root canal structures in CBCT cross-sectional images. Among the four architectures evaluated, DeepLabV3+ achieved the highest overall performance (mDice = 0.775), with Dice coefficients of 0.838 for the pulp cavity and 0.711 for the pulp stone, outperforming U-Net, Attention U-Net, and U-Net++. The proposed DACA-DeepLab architecture extends this backbone with PCSA encoder attention modules, dual-axis cross-attention decoding, and Sobel-guided boundary refinement, providing a principled framework for improving the delineation of small, calcified structures at clinically relevant boundaries. These findings contribute a systematic comparative benchmark for CBCT-based dental pulp segmentation and establish a foundation for AI-assisted endodontic diagnosis and pre-treatment planning tools. However, pulp stone segmentation accuracy remained moderate rather than high across all models, these tools are best regarded as an assistive aid to support radiologist interpretation rather than a replacement for direct visual assessment.

Statements and Declarations

Author Contributions

[Conceptualization: SM., YH, WS.; Methodology: SM, YH, SS, ; Formal Analysis: HM,; AI Methodology: ME.; Writing—original draft: SM.; Writing—review and editing: all authors; Supervision: WS.]

Human Ethics and Consent to Participate declarations

This study was conducted in accordance with the Declaration of Helsinki (revised 2013) and was

approved by the Research Ethics Committee of Faculty of Dentistry, Cairo University (approval date: 1/3/2026). Informed consent was obtained from all patients for being included in the study. The Research Ethics Committee of Faculty of Dentistry, Cairo University reviewed and approved the retrospective use of this deidentified imaging data.

Funding

This research received no external funding.

Data Availability Statement

The annotated dataset is available on Roboflow platform through this link: https://universe.roboflow.com/sally-mansour/pulpstone-aernc. The trained model weights are available upon request.

Decalaration of competing Interest

The authors declare no conflicts of interest.

References

1. Goga R, Chandler NP, Oginni AO. Pulp stones: a review. Int Endod J. 2008;41:457-468. https://doi.org/10.1111/j.1365-2591.2008.01374.x

2. Moss-Salentijn L, Hendricks-Klyvert M. Calcified structures in human dental pulps. J Endod. 1988;14:184-189. https://doi.org/10.1016/S0099-2399(88)80262-0

3. Sener S, Cobankara FK, Akgunlu F. Calcifications of the pulp chamber: prevalence and implicated factors. Clin Oral Investig. 2009;13:209-215. https://doi.org/10.1007/s00784-008-0212-x

4. Arys A, Philippart C, Dourov N. Microradiography and light microscopy of mineralization in the pulp of undemineralized human primary molars. J Oral Pathol Med. 1993;22:49-53. https://doi.org/10.1111/j.1600-0714.1993.tb00041.x

5. Yuce F, Öziç MÜ, Tassoker M. Detection of pulpal calcifications on bite-wing radiographs using deep learning. Clin Oral Investig. 2023;27:2679-2689. https://doi.org/10.1007/s00784-022-04839-6

6. Pietrzycka K, Pawlicka H. Clinical aspects of pulp stones: a case report series. Dent Med Probl. 2020;57:213-220. https://doi.org/10.17219/dmp/117944

7. McCabe P, Dummer PMH. Pulp canal obliteration: an endodontic diagnosis and treatment challenge. Int Endod J. 2012;45:177-197. https://doi.org/10.1111/j.1365-2591.2011.01963.x

8. Nayak M, Kumar J, Prasad LK. A radiographic correlation between systemic disorders and pulp stones. Indian J Dent Res. 2010;21:369-373. https://doi.org/10.4103/0970-9290.70803

9. Moss-Salentijn L, Klyvert MH. Epithelially induced denticles in the pulps of recently erupted, noncarious human premolars. J Endod. 1983;9:554-560. https://doi.org/10.1016/S0099-2399(83)80060-0

10. Tassoker M, Magat G, Sener S. A comparative study of cone-beam computed tomography and digital panoramic radiography for detecting pulp stones. Imaging Sci Dent. 2018;48:201-212. https://doi.org/10.5624/isd.2018.48.3.201

11. Mahabob N, Madiraju GS, Bello SM, Abraham S. Artificial intelligence-assisted analysis of prevalence patterns and phenotypic distribution of dental pulp stones using cone-beam computed tomography imaging. Niger J Clin Pract. 2026;29:280-287. https://doi.org/10.4103/njcp.njcp_823_25

12. Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 2018;77:106-111. https://doi.org/10.1016/j.jdent.2018.07.015

13. Krois J, Ekert T, Meinhold L, Golla T, Kharbot B, Wittemeier A, et al. Deep learning for the radiographic detection of periodontal bone loss. Sci Rep. 2019;9:8495. https://doi.org/10.1038/s41598-019-44839-3

14. Orhan K, Bayrakdar IS, Ezhov M, Kravtsov A, Özyürek T. Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. Int Endod J. 2020;53:680-689. https://doi.org/10.1111/iej.13265

15. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: a systematic review. Dentomaxillofac Radiol. 2020;49:20190107. https://doi.org/10.1259/dmfr.20190107

16. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical image computing and computer-assisted intervention – MICCAI 2015. pp 234-241. Cham: Springer, 2015. https://doi.org/10.1007/978-3-319-24574-4_28

17. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa K, et al. Attention U-Net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. 2018. https://doi.org/10.48550/arXiv.1804.03999

18. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNet++: a nested U-Net architecture for medical image segmentation. In Stoyanov D, Taylor Z, Carneiro G, et al. (eds) Deep learning in medical image analysis and multimodal learning for clinical decision support. pp 3-11. Cham: Springer, 2018. https://doi.org/10.1007/978-3-030-00889-5_1

19. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) Computer vision – ECCV 2018. pp 833-851. Cham: Springer, 2018. https://doi.org/10.1007/978-3-030-01234-2_49

20. Selmi A, Syed L, Abdulkareem B. Pulp stone detection using deep learning techniques. Lect Notes Inst Comput Sci Soc Inform Telecommun Eng. 2022;432:113-124. https://doi.org/10.1007/978-3-030-99197-5_10

21. Altindağ A, Uzun S, Bayrakdar İŞ, Çelik Ö. Detecting pulp stones with automatic deep learning in bitewing radiographs: a pilot study of artificial intelligence. Eur Ann Dent Sci. 2023;50:12-16. https://doi.org/10.52037/eads.2023.0004

22. Altındağ A, Bahrilli S, Çelik Ö, Bayrakdar İŞ, Orhan K. The detection of pulp stones with automatic deep learning in panoramic radiographies: an AI pilot study. Diagnostics (Basel). 2024;14:890. https://doi.org/10.3390/diagnostics14090890

23. Gürhan C, Yiğit H, Yılmaz S, Çetinkaya C. A novel deep learning-based pipeline architecture for pulp stone detection on panoramic radiographs. Oral Radiol. 2025;41:285-295. https://doi.org/10.1007/s11282-025-00804-7

24. Manasa GR, Venugopala PS, Santhosh G, Umesh G. Detection of pulp stones in radiographic images: survey using artificial intelligence. In Proceedings of the IEEE COSMIC Conference. pp 226-232. IEEE, 2024. Available at https://ieeexplore.ieee.org/document/10871780

25. Firincioglulari M, Boztuna M, Mirzaei O, Karanfiller T, Akkaya N, Orhan K. Segmentation of pulp and pulp stones with automatic deep learning in panoramic radiographs: an artificial intelligence study. Dent J (Basel). 2025;13:274. https://doi.org/10.3390/dj13060274

26. Dwyer B, Nelson J, Hansen T, et al. Roboflow (Version 1.0) [Software]. 2026. Available at https://roboflow.com (accessed September 2026).

27. Said ZA, Abdulqader AA, Liu F, Jiang FL, Al-Rubaye MH, Alkadasi BA, et al. Automated detection and classification of maxillary sinus variations using slice-based and full-volume CBCT deep learning models. BMC Oral Health. 2026;26:919. https://doi.org/10.1186/s12903-026-08205-w

28. Shetty S, Talaat W, Al-Rawi N, Al Kawas S, Sadek M, Elayyan M, et al. Accuracy of deep learning models in the detection of accessory ostium in coronal cone beam computed tomographic images. Sci Rep. 2025;15:8324. https://doi.org/10.1038/s41598-025-93250-8

29. Bachtiar MM, Fatichah C, Raharjo AB, Astuti ER, Putra RH. Segmentation and visualization of tooth from CBCT images using 2D U-Net++. In Proceedings of the 2025 15th International Conference on Information & Communication Technology and Systems (ICTS). IEEE, 2025. https://doi.org/10.1109/ICTS67612.2025.11369489

30. Feng L, Li Z, Gu Q, Wang Y, Yu X. Adults' dental cone beam computed tomography images dataset for detecting and classifying missing teeth. Sci Data. 2026;13:828. https://doi.org/10.1038/s41597-026-07156-9

31. Yunker A, Luce J, Roeske JC, Kettimuthu R, Kang H, Gros S, et al. Noise2Inverse for 3D low-dose cone-beam computed tomography. In Proceedings of the 2024 IEEE International Conference on Big Data (BigData). pp 2057-2064. IEEE, 2024. https://doi.org/10.1109/BigData62323.2024.10825036

32. Minnema J, van Eijnatten M, der Sarkissian H, Doyle S, Koivisto J, Wolff J, et al. Efficient high cone-angle artifact reduction in circular cone-beam CT using deep learning with geometry-aware dimension reduction. Phys Med Biol. 2021;66:135015. https://doi.org/10.1088/1361-6560/ac09a1

33. Qureshi I, Yan J, Abbas Q, Shaheed K, Riaz AB, Wahid A, et al. Medical image segmentation using deep semantic-based methods: a review of techniques, applications and emerging trends. Inf Fusion. 2023;90:316-352. https://doi.org/10.1016/j.inffus.2022.09.031

34. Hadjiiski L, Cha K, Chan HP, Drukker K, Morra L, Näppi JJ, et al. AAPM task group report 273: recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging. Med Phys. 2023;50:e1-e24. https://doi.org/10.1002/mp.16188

35. Aljabri M, AlAmir M, AlGhamdi M, Abdel-Mottaleb M, Collado-Mesa F. Towards a better understanding of annotation tools for medical imaging: a survey. Multimed Tools Appl. 2022;81:25877-25911. https://doi.org/10.1007/s11042-022-12100-1

36. Athanasiou G, Arcos JL, Cerquides J. Enhancing medical image segmentation: ground truth optimization through evaluating uncertainty in expert annotations. Mathematics. 2023;11:3771. https://doi.org/10.3390/math11173771

37. Bayati M, Alizadeh Savareh B, Ahmadinejad H, Mosavat F. Advanced AI-driven detection of interproximal caries in bitewing radiographs using YOLOv8. Sci Rep. 2025;15:4641. https://doi.org/10.1038/s41598-024-84737-x

38. Salehizeinabadi M, Neghab S, Ameli N, Koucheh Baghi K, Pacheco-Pereira C. Automated classification of dental caries in bitewing radiographs using machine learning and the ICCMS framework. Int J Dent. 2025;2025:6644310. https://doi.org/10.1155/ijod/6644310

Figures and figure legends:

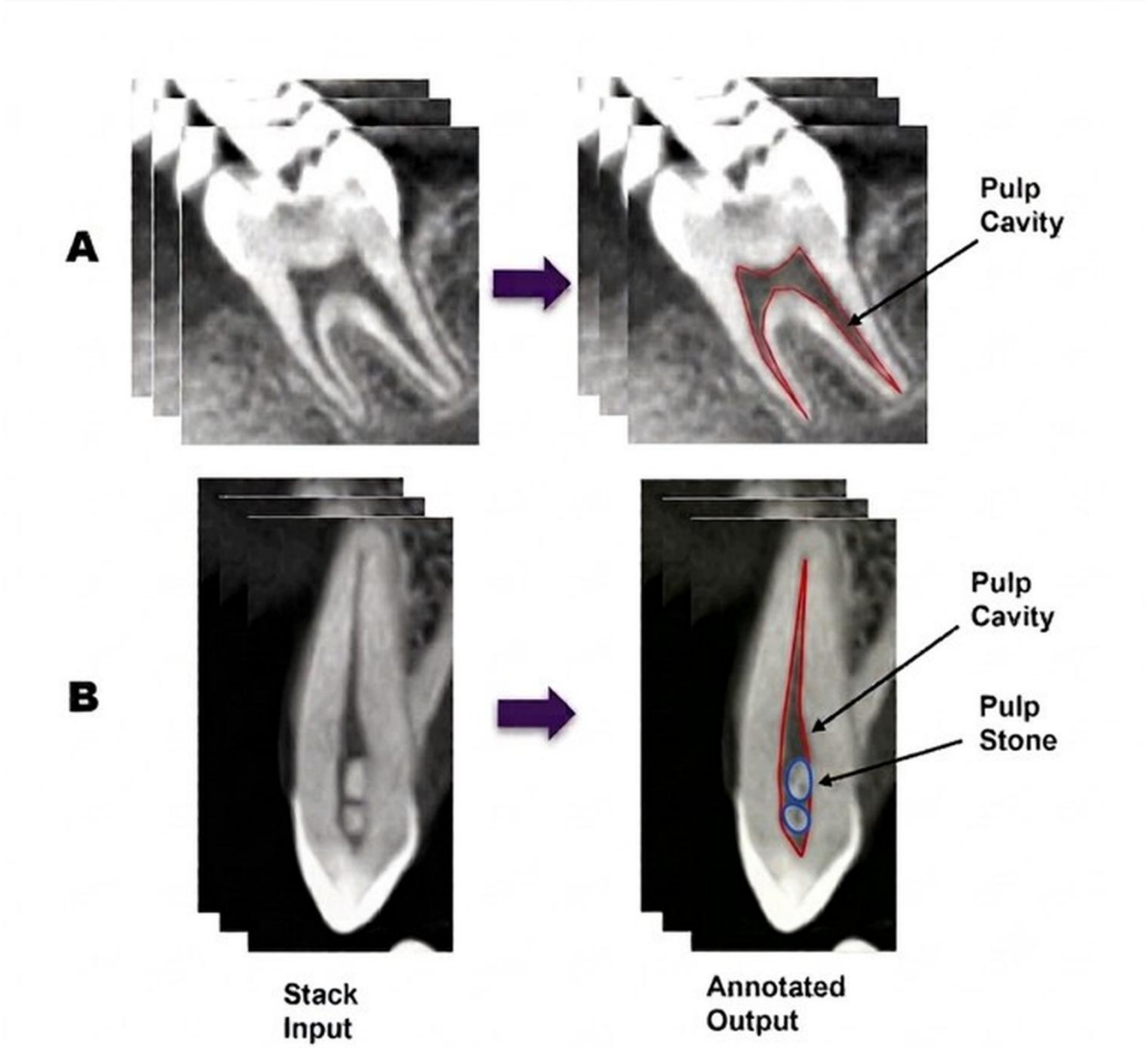


Fig. 1- Annotation workflow for CBCT slices. Raw input slices (left) were manually annotated to delineate the pulp cavity (red outline) and pulp stones (blue outline). (A) Multirooted posterior tooth with the pulp cavity outlined as a single continuous contour and no pulp stones. (B) Single-rooted tooth

showing two discrete pulp stones within the pulp cavity lumen.

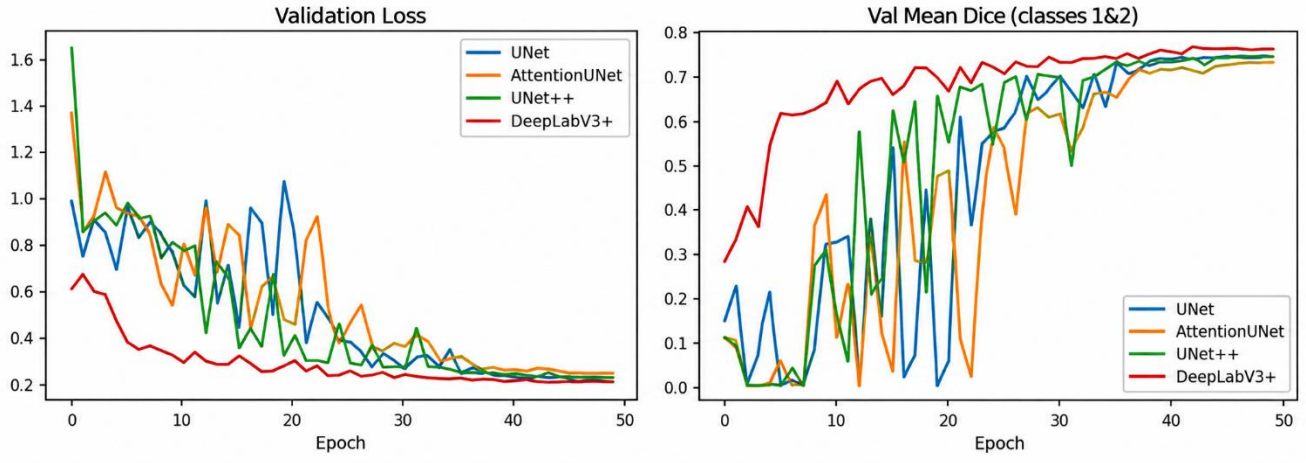


Fig. 2- Validation loss and mean Dice (pulp + pulp stone) across training epochs for U-Net, Attention U-Net, U-Net++, and DeepLabV3+.

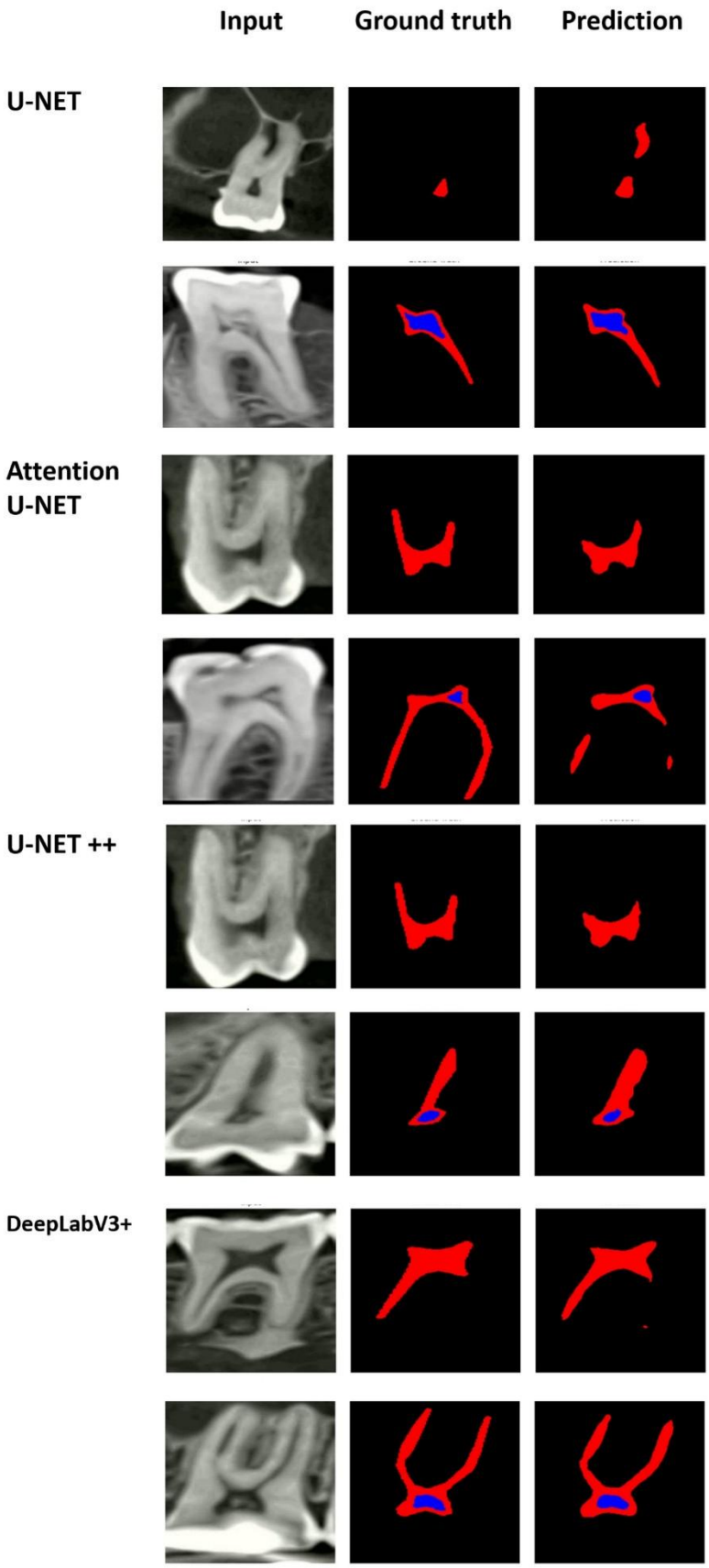

Input
Ground truth
Prediction
U-NET
Attention U-NET
U-NET ++
DeepLabV3+

Fig. 3- Qualitative segmentation results for U-Net, Attention U-Net, U-Net++, and DeepLabV3+ (input, ground truth, prediction; red = pulp/canal, blue = pulp stone). Predictions closely match ground truth, with occasional under-segmentation in thin canal regions.

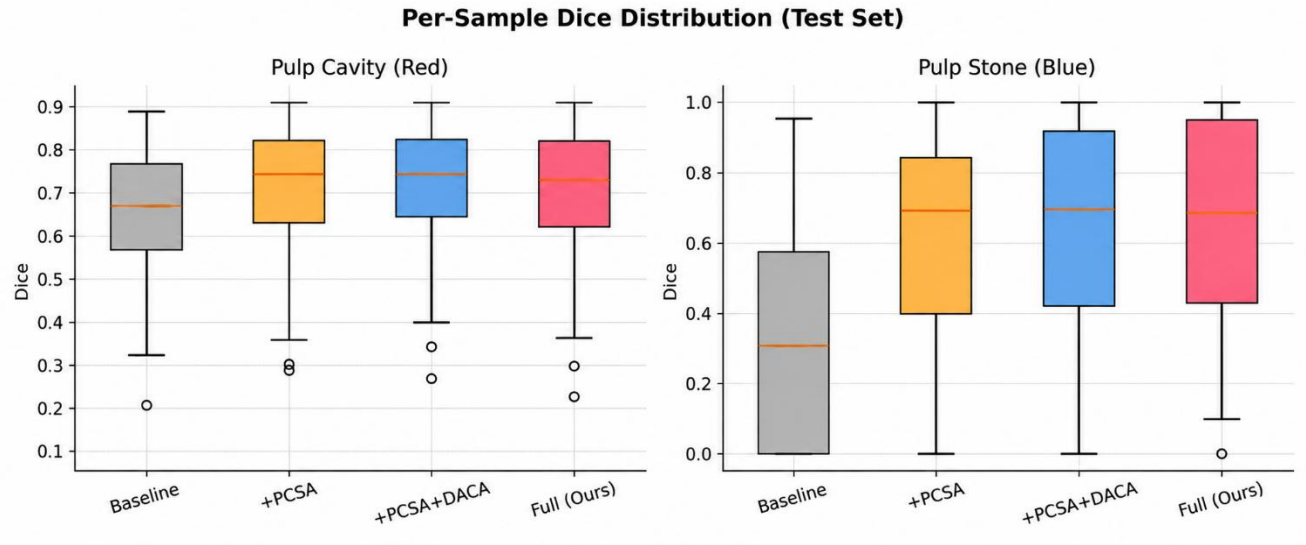


Fig. 4- Per-sample Dice distribution on the test set for the pulp cavity and root canal across the incremental variants.

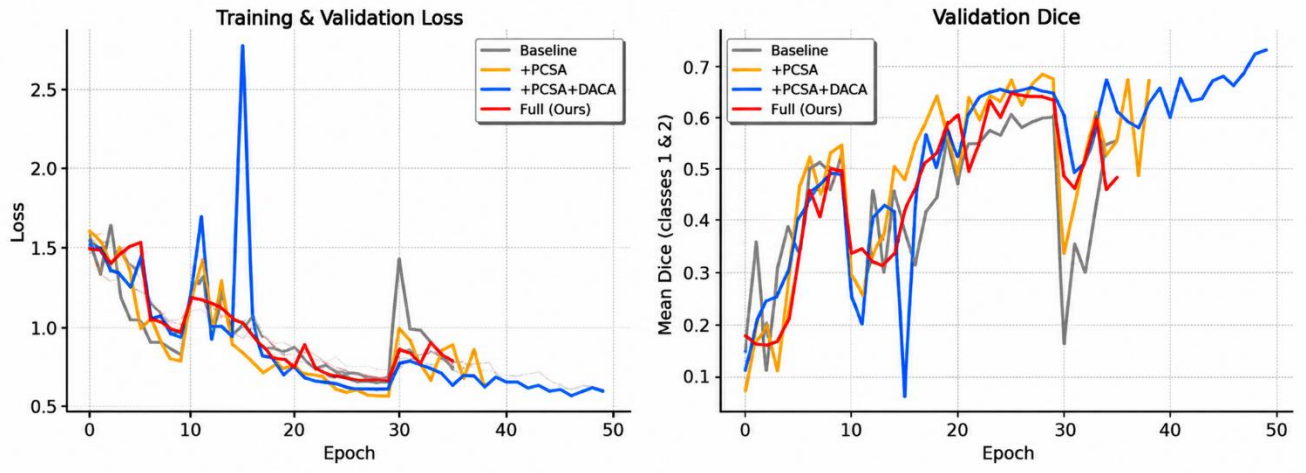


Fig. 5- Training/validation loss and validation mean Dice across epochs for the DACA-DeepLab incremental variants.

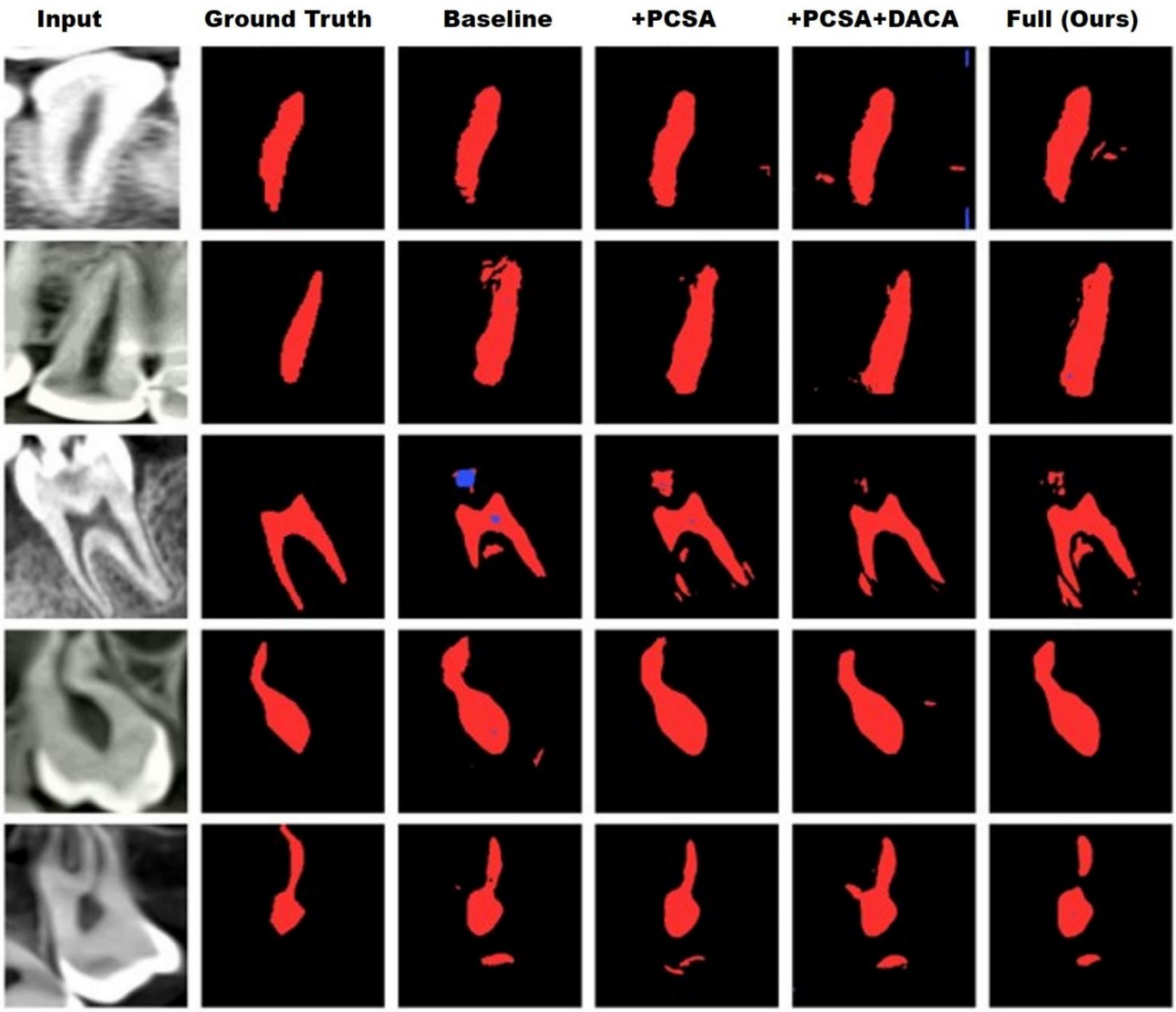


Fig. 6- Five representative CBCT cross-sections demonstrating qualitative predictions of the DACA-DeepLab incremental variants versus ground truth (red, pulp cavity; blue, pulp stone).

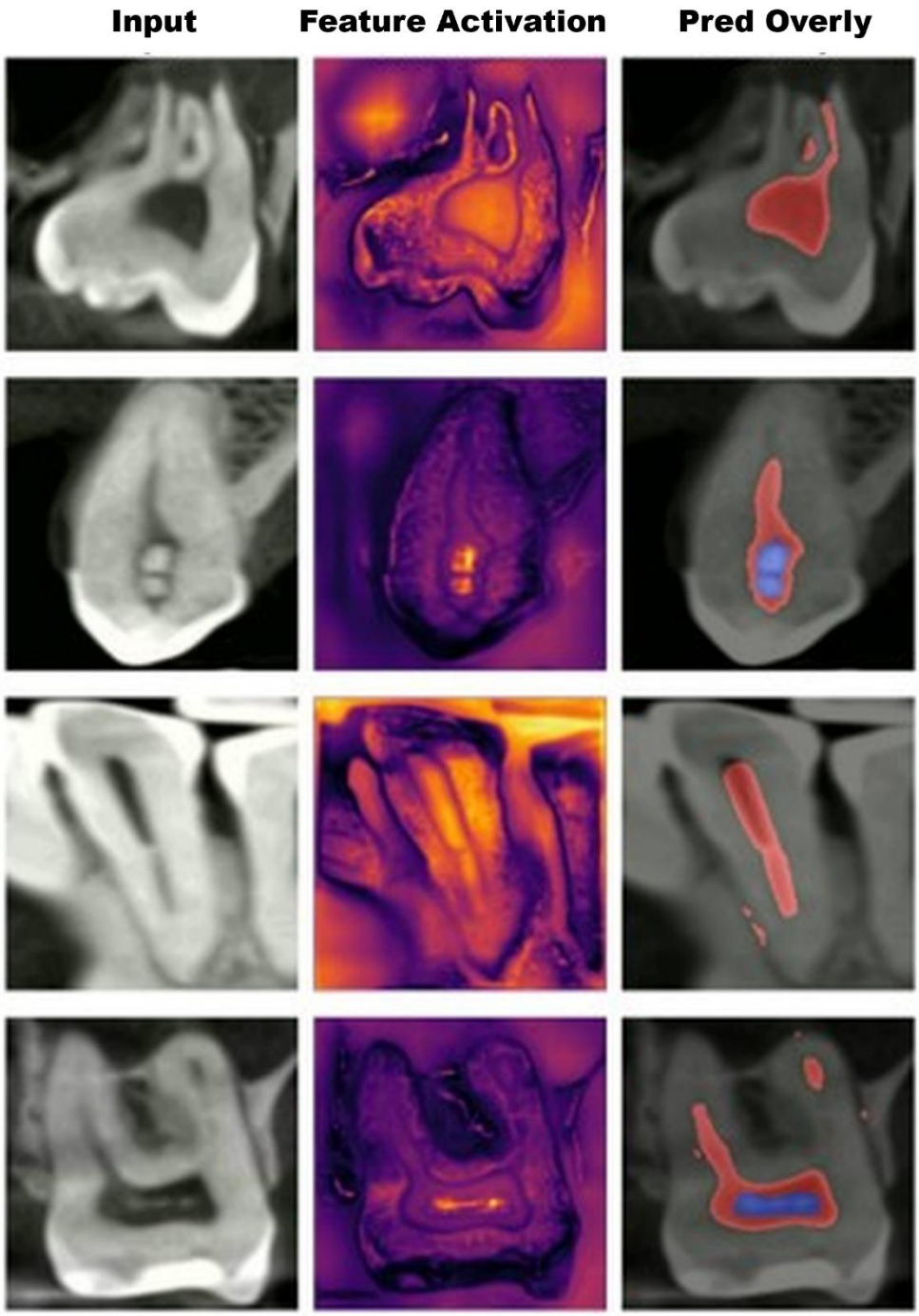


Fig. 7- Four representative CBCT cross-sections demonstrating DACA-DeepLab encoder activation maps and predicted segmentation overlays (red, pulp cavity ; blue, pulpstone). Warmer colors indicate higher network attention.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixA.docx