applied sciences

MDPI

Article

Context-Aware Caries Segmentation in Periapical Radiographs Using a Hybrid Multi-Task Learning Framework with Partial Annotations

Jesús Antonio Nava-Pintor [1], Héctor A. Guerrero-Osuna [1,*], Fabián García-Vázquez [1], Luis F. Luque-Vega [2], Teodoro Ibarra-Pérez [3], Salvador Ibarra-Delgado [4], Víktor I. Rodríguez-Abdalá [4], Remberto Sandoval-Arechiga [4] and José Ricardo Gómez-Rodríguez [4]

1 Posgrado en Ingeniería y Tecnología Aplicada, Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Zacatecas 98000, Zacatecas, Mexico; jesus.nava@uaz.edu.mx (J.A.N.-P.); fabian.garcia@uaz.edu.mx (F.G.-V.)
2 Department of Technological and Industrial Processes, ITESO, Tlaquepaque 45604, Jalisco, Mexico; luisluque@iteso.mx
3 Unidad Profesional Interdisciplinaria de Ingeniería Campus Zacatecas (UPIIZ), Instituto Politécnico Nacional, Zacatecas 98160, Zacatecas, Mexico; tibarrap@ipn.mx
4 Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Zacatecas 98000, Zacatecas, Mexico; sibarra@uaz.edu.mx (S.I.-D.); abdala@uaz.edu.mx (V.I.R.-A.); rsandoval@uaz.edu.mx (R.S.-A.); jrgrodri@uaz.edu.mx (J.R.G.-R.)
* Correspondence: hectorguerreroo@uaz.edu.mx; Tel.: +52-492-159-5564

Abstract

Dental caries remains a prevalent diagnostic challenge, particularly in periapical radiographs, where geometric distortions and anatomical overlap complicate interpretation. Although deep learning has advanced dental image analysis, most segmentation models depend on fully annotated datasets and rarely exploit anatomical context. This study proposes a hybrid Multi-Task Learning (MTL) framework that jointly performs anatomical segmentation and caries detection in scenarios with partial and asymmetric annotations. The method integrates a U-Net++ dual-head architecture with a shared EfficientNet-B4 encoder, supplemented by pseudo-label generation and selective loss masking to handle incomplete ground truth. We hypothesize that learning healthy dental structures provides a contextual scaffold that enhances the identification of pathology. Experiments on an independent test set validate this hypothesis: the proposed MTL model achieved an DSC of 0.6706 and an IoU of 0.5044, outperforming a specialized single-task baseline. Most notably, sensitivity improved by 7.47%, reducing false-negative pixels by 19.9%. These findings demonstrate that context-aware supervision substantially improves detection robustness on complex periapical radiographs, even when full-pixel-level annotations are unavailable. The proposed framework offers a scalable pathway for developing clinically oriented diagnostic tools in real-world settings where annotation completeness is limited.

Keywords: artificial intelligence; semantic segmentation; deep learning; multi-task learning; periapical radiographs; caries segmentation

Check for updates

Academic Editor: Iole Vozza

Received: 4 December 2025
Revised: 15 December 2025
Accepted: 22 December 2025
Published: 26 December 2025

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Dental caries is one of the most prevalent chronic diseases worldwide, affecting an estimated 2.4 to 3.1 billion people and posing a significant public health challenge in most industrialized countries [1–4].Detecting caries is crucial as it helps prevent serious

complications, especially in the early stages, such as pulpitis, the need for endodontic treatments, and tooth loss [5,6]. Furthermore, addressing oral health concerns is crucial to reducing the overall impact of disease. Research has shown a significant correlation between oral health and various systemic conditions, including cardiovascular, respiratory, and metabolic diseases [7,8].

The diagnosis of dental caries begins with a visual inspection, supported by various imaging techniques such as periapical , bitewing, panoramic, and cone-beam computed tomography (CBCT). Each of these methods offers distinct insights tailored to specific diagnostic needs [9].

Despite advancements in image acquisition and the diverse techniques available, radiographic interpretation remains a challenge. The diagnostic process is inherently subjective and highly dependent on the clinician's experience [10]. This variability leads to significant diagnostic discrepancies, where different professionals may reach divergent conclusions for the same case [11]. While systems like the International Caries Detection and Assessment System (ICDAS) aim to standardize criteria, challenges in radiographic interpretation and inter-observer variability persist [12,13].

In response to these challenges, artificial intelligence (AI), particularly deep learning, has emerged as a promising tool to enhance objectivity and accuracy in dental radiographic diagnosis [14]. While initial research focused on classification (the presence/absence of caries) and object detection (localizing lesions with bounding boxes) [9], these methods provide a limited view that fails to answer key clinical questions regarding the extent, depth, and relationship of caries to adjacent anatomical structures.

To overcome these limitations, the field has evolved towards semantic segmentation, an advanced technique that offers essential pixel-level detail. This method is especially relevant in the context of today's dental workflow, which digital dental imaging sensors have revolutionized. These sensors enable the rapid capture of high-resolution images, creating the ideal conditions for effectively integrating AI-driven computer-aided diagnostic tools [15,16].

This establishes the periapical radiograph as a cornerstone diagnostic instrument in this digital workflow. It provides an in-depth view of a specific tooth, showing everything from the crown to the root apex, which is essential for a thorough evaluation [9]. Despite the critical role of periapical radiographs, there is a noticeable discrepancy in AI-driven radiological research. While Deep Learning has revolutionized medical image analysis, most advanced semantic segmentation models have been predominantly developed for other imaging modalities, leaving the specific challenges of periapical radiography less explored.

In this study, we propose a Multi-Task Learning (MTL) model for caries segmentation in periapical radiographs. Accurate interpretation of these complex images entails more than merely identifying lesions; it requires a thorough understanding of the surrounding anatomical context, a requirement for which the MTL paradigm is particularly well-suited [17]. To address this, our study proceeds with the hypothesis that effective caries detection depends on the adjacent anatomical structural context. By adopting a joint learning strategy, the model can leverage inductive biases to resolve feature ambiguities, using the anatomical representation as a scaffold to support and refine the detection of pathology.

However, achieving this context-aware vision is hindered by a significant challenge: the limited availability of pixel-level annotations. Obtaining complete GT for a wide range of structures (e.g., enamel, dentin, pulp, and lesions) is a labor-intensive process that often yields datasets with partial or asymmetric labels. Therefore, this leads to a severe limitation for standard MTL models, as they often fail to learn robust features when key anatomical targets are missing from the training supervision.

To overcome this barrier and validate our context-driven hypothesis, we introduce a Hybrid Multi-Task Learning methodology designed explicitly for partially annotated periapical radiography datasets. Our framework integrates a semi-supervised strategy with a tailored training process to effectively learn from heterogeneous data sources, ranging from fully supervised images to those with missing labels, thereby unlocking the potential of anatomical context without requiring complete annotations for every sample.

Our main contributions can be summarized as follows:

- A Hybrid MTL Framework, a U-Net++ dual-head architecture with a shared encoder that simultaneously performs multi-class anatomical segmentation and binary caries detection, enabling efficient feature reuse.
- A Proprietary dataset of 686 periapical radiographs, featuring pixel-level expert annotations for anatomical structures (600 images) and pathological lesions (293 images), serving as a specialized resource for the dental AI community.
- The development of a training strategy that handles incomplete Ground Truth through strategic pseudo-labeling and a selective loss mechanism, expanding the usable dataset and ensuring consistent supervision despite asymmetric labels.
- The design of a composite loss function for the caries head, integrating Focal Tversky, Soft Cross-Entropy, and Lovász-Softmax terms, specifically tailored to address severe class imbalance and refine lesion boundaries.
- The empirical validation of the Contextual approach, which prioritizes anatomical learning as a regularizing scaffold, leading to a marked increase in caries detection sensitivity (+7.47%) compared to specialized single-task models, underscoring the importance of context in medical imaging.

The remainder of the article is organized as follows: Section 2 reviews related work in dental radiography analysis and multi-task learning (MTL). Section 3 describes the construction of the dataset and outlines the proposed Hybrid MTL methodology. Section 4 presents both quantitative and qualitative experimental results. Section 5 discusses the findings and limitations of the study, while Section 6 provides the concluding remarks.

2. Related Works

The application of artificial intelligence in dental radiography has evolved rapidly, with research primarily focusing on classification, object detection, and segmentation. While classification models have achieved high accuracy in identifying caries [18–20], and object detection models have improved lesion localization [21–26], both lack the pixel-level precision required for clinical treatment planning. Consequently, recent literature on semantic segmentation of 2D intraoral radiographs can be categorized into three primary domains based on the clinical objective.

The first domain addresses structural segmentation, focusing exclusively on delineating the shape and orientation of teeth or bone without disease diagnosis. Works by Sheng et al. [27] and Deb et al. [28] have leveraged advanced architectures, such as SWin-UNet and attention-gated networks, to achieve high-precision segmentation of individual teeth on panoramic radiographs. Similarly, Yu et al. [29] utilized hybrid Transformer–CNN models (SegFormer) to master contextual understanding of crown and root regions, setting a new benchmark for periodontal assessment. However, these models essentially ignore pathological conditions.

The second domain attempts to establish a comprehensive diagnostic benchmark by jointly segmenting anatomy and pathology. The seminal work of Wang et al. [30] on bitewing radiographs proved the superiority of U-Net over classical methods for multi-class segmentation (enamel, dentin, pulp, caries). This framework was later validated by Lee et al. [31], who demonstrated that simultaneous segmentation significantly enhances

diagnostic sensitivity. However, these successes have been largely confined to bitewing imaging, where anatomical overlap is minimal compared to periapical views.

The third domain specifically targets periapical radiographs, which are critical for detailed diagnostics but present greater geometric challenges. Initial benchmarks by Khan et al. [32] showed modest performance (DSC: 53.2%), constrained by a limited dataset of approximately 200 images, while recent work by Priya and Raja [33] leveraged a significantly larger dataset (1400 images), achieving near-state-of-the-art delineation of caries (DSC: 95.77%), using attention-based networks. Additionally, Ibraheem et al. [34] validated commercial AI software capable of detecting both caries and bone loss with high accuracy. Despite these advances, most periapical studies remain pathology-focused, often excluding the segmentation of healthy internal structures like enamel and dentin which provide essential context.

A summary of these key approaches is provided in Table 1.

Table 1. Summary of Key Approaches in Dental Semantic Segmentation.

Primary Focus	Author (Year)	Modality	Segmentation Targets	Key Contribution
Structural Analysis	Sheng et al. (2023) [27]	Panoramic	Teeth	Proposed SWin-Unet for high-precision structural segmentation.
	Deb et al. (2025) [28]	Panoramic	Teeth (with orientation)	Integrated orientation data for orthodontic planning.
	Yu et al. (2024) [29]	Panoramic	Crown, Root, Bone	Hybrid Transformer–CNN for periodontal assessment.
Multi-Class (Bitewing)	Wang et al. (2016) [30]	Bitewing	Caries, Enamel, Dentin, Pulp	Established the first multi-class deep learning benchmark.
	Lee et al. (2021) [31]	Bitewing	Caries, Anatomy, Restorations	Demonstrated improved clinical sensitivity via multi-class segmentation.
Periapical Analysis	Khan et al. (2020) [32]	Periapical	Caries, Bone recession	First periapical segmentation benchmark (DSC: 53.2%).
	Ari et al. (2022) [35]	Periapical	Teeth, Pulp, Restorations	Applied CLAHE + U-Net for robust anatomical feature segmentation (DSC: 0.82).
	Priya & Raja (2025) [33]	Periapical	Caries (Lesion only)	Achieved high accuracy (DSC: 95.77%) using attention mechanisms.
	Ibraheem et al. (2025) [34]	Periapical	Caries, Bone loss	Validated commercial software performance against clinicians.
Our Approach	This Study	Periapical	MTL (Caries + Full Anatomy (Enamel, Dentin, Pulp))	Methodological Gap: Proposes an MTL framework to handle asymmetric datasets (anatomy vs. pathology) for holistic periapical segmentation.

As outlined in Table 1, existing research predominantly concentrates on panoramic or bitewing radiographs. While periapical radiography is ubiquitous in clinical practice [34,36,37], it remains underexplored in deep learning due to its geometric complexity, including distortion and overlapping structures [38]. This complexity leads to significant performance disparities; for instance, while panoramic segmentation models achieve DSC around 0.91 [29], early multi-class periapical models struggled to exceed DSC 53% [32].

This review highlights a bottleneck in dental AI research. While researchers recognize the theoretical benefits of using anatomical context to aid in pathology detection in other medical domains, there is no empirical evidence that this approach works for periapical caries segmentation. The field lacks a methodological framework for learning from fragmented, asymmetric datasets in the real world, such as separate sources for anatomy and pathology.

As a result, Multi-Task Learning faces limitations in enhancing diagnostic sensitivity for this complex modality due to data scarcity. Our work aims to overcome this limitation. We introduce a Hybrid MTL framework not just as a technical exercise but as a tool for validating whether simultaneous anatomical learning significantly improves caries detection performance.

3. Materials and Methods

The methodology of this study, depicted in Figure 1 , is designed to overcome the challenge of training a dual-task segmentation model (anatomical structures and pathological caries) using datasets with partial and asymmetric annotations. The workflow is organized into four sequential phases: (1) data characterization and strategic splitting, (2) training of a single-task structure model, (3) generation of pseudo-labels to complete the missing Ground Truth (GT), and (4) the training and optimization of the proposed Hybrid MTL framework. To validate this approach, a single-task baseline model for caries segmentation was also developed for comparison.

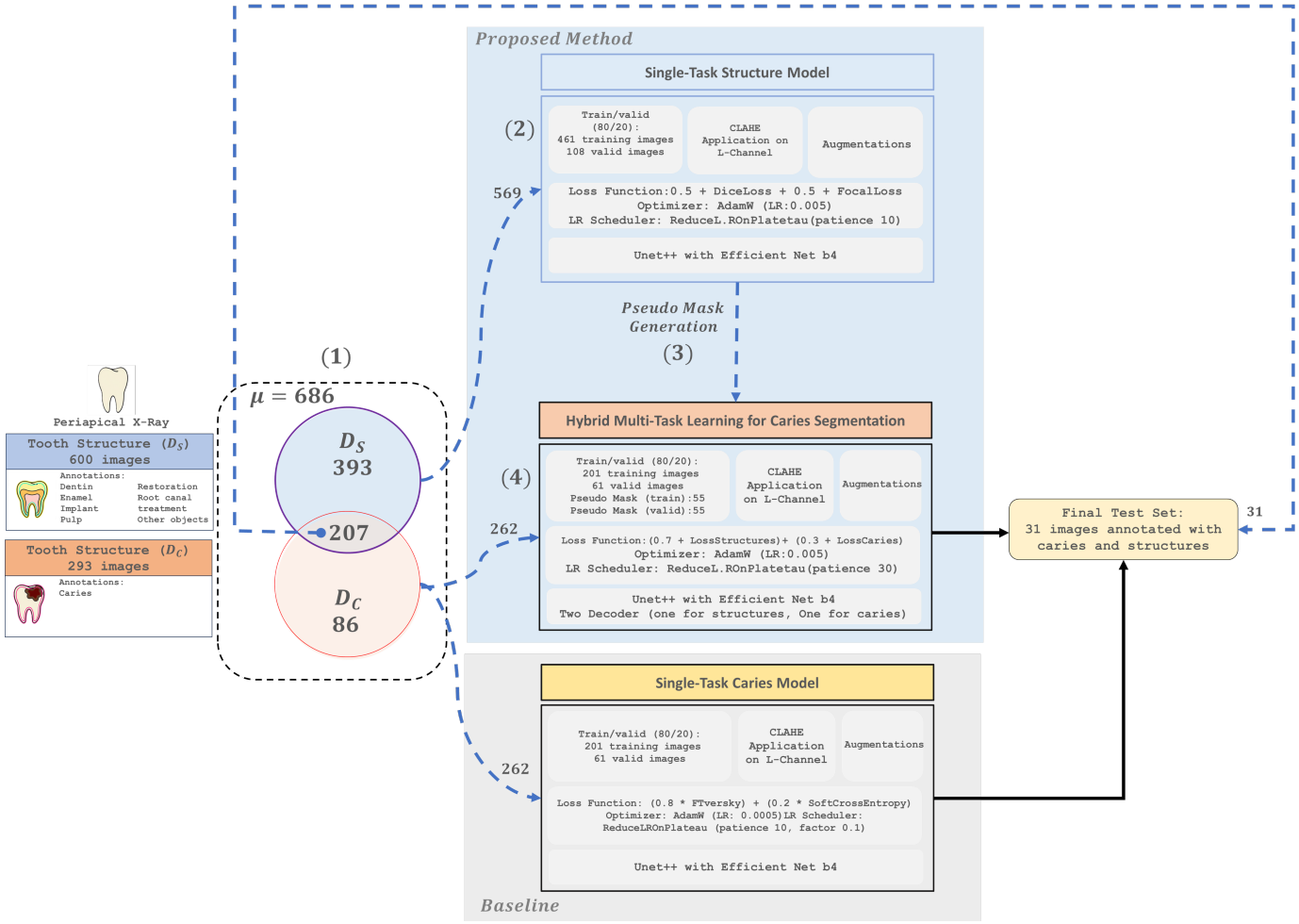


Figure 1. Schematic overview of the proposed Hybrid Multi-Task Learning (MTL) methodology.

3.1. Data Characterization and Strategic Splitting

The dataset was created at the Zentia Dental Clinic in Mexico by dental professionals who acquired periapical radiographs as part of routine clinical consultation procedures. All images were captured using a Hyperlight-G digital radiovisiograph between May 2024 and November 2025. The study population included patients aged 6 to 80 years. We conducted

the study in strict accordance with the ethical principles outlined in the Declaration of Helsinki and obtained informed consent from all participants prior to data collection.

The images were uploaded to the Roboflow platform for a collaborative annotation process managed by a panel of three specialist dentists, each with over 10 years of clinical experience. One specialist performed the initial pixel-level annotations for all classes (structures or caries). These annotations were then subsequently reviewed by the other two specialists. An annotation was considered final and included in the dataset only after consensus among all three experts.

This process resulted in two primary, partially overlapping data collections:

1. Structure Dataset (D_S): Comprising 600 unique images, containing dense, pixel-level annotations for eight anatomical structure classes (enamel, dentin, pulp, restoration, conduct treatment, implant, and other objects).
2. Caries Dataset (D_C): Comprising 293 unique images. This set contained pixel-level annotations for the caries class only.

Figure 2 shows representative examples of annotation classes, provided to illustrate the pixel-level delineation used for both anatomical D_S and D_C.

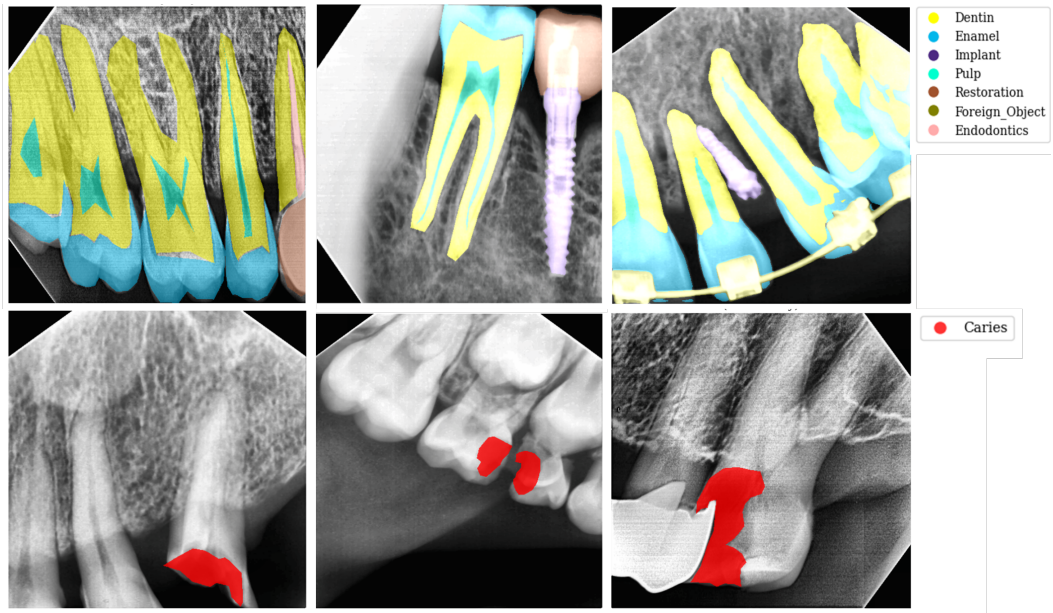


Figure 2. Representative examples of the pixel-level annotations used in Dataset A (D_S) and Dataset B (D_C).

3.2. Hybrid Dataset Splitting Strategy

As noted, both datasets exhibit an inherent imbalance in the number of annotated images. This asymmetry arises from the clinical workflow, in which many dental consultations are unrelated to caries diagnosis and therefore involve radiographs without carious lesions. Consequently, the dataset of anatomical structure annotations is substantially larger than the dataset focused exclusively on caries cases.

To quantify this asymmetry, we analyzed the file base names to characterize the total data universe. A total of 686 unique images were identified ($\mathcal{U}$) and subsequently classified into three disjoint categories:

- Structure-Only ($U_S = D_S \setminus D_C$): 393 images possessing only structure GT.
- Caries-Only ($U_C = D_C \setminus D_S$): 86 images possessing only caries GT.
- Intersection ($U_{Int} = D_S \cap D_C$): 207 images possessing real GT for both tasks.

To manage this data asymmetry, we employed a custom splitting strategy based on the 686 unique base images identified in $\mathcal{U}$. First, a set of 31 images was sampled exclusively from the U_{Int} set, where real GT for both structures and caries was available. This set was held out for the final, blind evaluation of the model. Then, the remaining 655 images formed the pool for model development, containing U_S, U_C, and the remainder of the U_{Int}. This workflow is represented in Figure 1 part (1).

To prevent data leakage, partitioning was performed at the patient level. All radiographs associated with a single patient ID were assigned exclusively to the training, validation, or test set. In this way, the stratification ensures that the model is not evaluated on anatomical structures from individuals present in the training set

3.3. Methodological Framework: A Hybrid MTL Approach

3.3.1. Training the Model for Structure Segmentation

To address the missing structure annotations in the U_C subset, a Single-Task Structure model was first developed. For this study, a Unet++ architecture with an EfficientNet-B4 encoder was trained on 569 images with a validated Ground Truth structure, as illustrated in Figure 1, part (2). This selection is based on recent findings showing enhanced segmentation of delicate anatomical structures in low-contrast regions, such as the boundaries of alveolar bone and enamel, achieved through nested skip connections and deep supervision provided by Unet++ [39].

Methodologically, to prevent data leakage, this training dataset ($\mathcal{T}_{Structure}$) was constructed by combining the U_S and U_{Int} sets, excluding the 31 images reserved for the Final Test Set ($\mathcal{T}_{final}$).

$$\mathcal{T}_{Structure} = (U_S \cup U_{Int}) \setminus \mathcal{T}_{final} \quad (1)$$

$\mathcal{T}_{Structure}$ dataset was further split into training (80%) and validation (20%) subsets.

To standardize the radiographic inputs, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to all images, followed by resizing to 512 × 512. To address potential overfitting caused by limited image data, we incorporated a data augmentation pipeline into the training process, including geometric and photometric transformations such as rotation, flipping, brightness shifts, and gamma contrast adjustments. These techniques have been shown to enhance contrast in dental radiographs without introducing artifacts, improving subsequent segmentation accuracy [40].

Model training was optimized using the AdamW algorithm coupled with a dynamic learning rate scheduler. A critical challenge in this task is the volumetric imbalance between distinct anatomical structures and the background. To address this, we employed a compound loss function defined as the weighted sum of Dice Loss and Focal Loss. The Dice component promotes the spatial overlap of segmented regions, directly optimizing the evaluation metric, while the Focal term addresses class imbalance by down-weighting easy background examples and focusing training on difficult, misclassified regions [41].

A comprehensive summary of the model architecture, training hyperparameters, data-augmentation pipeline, and image pre-processing strategies used for structure segmentation is provided in Table 2.

Table 2. Complete configuration of the UNet++ architecture, training hyper-parameters, data-augmentation pipeline and preprocessing steps.

Parameter	Configuration/Value
	Model Architecture
Architecture	Unet++
Encoder	EfficientNet-B4 (ImageNet weights)
Input Size	512×512 pixels
Classes	8 (Multi-class)
	Optimization & Loss
Optimizer	AdamW
Initial Learning Rate	5×10^{-4}
Scheduler	ReduceLROnPlateau (Patience = 10, Factor = 0.1)
Batch Size	8
Loss Function	$0.5 \cdot \text{DiceLoss} + 0.5 \cdot \text{FocalLoss}$
Early Stopping	30 epochs
	Data Augmentation
Geometric	Horizontal Flip ($p = 0.5$), ShiftScaleRotate ($p = 0.7$)
Noise & Blur	GaussNoise ($p = 0.3$), RandomBlur ($p = 0.3$)
Intensity	RandomBrightnessContrast ($p = 0.5$), RandomGamma ($p = 0.5$)
Preprocessing	CLAHE (ClipLimit = 2.0), Normalization (ImageNet stats)

3.3.2. Generation of Pseudo-Structure Masks

After training the Structure Model, we used it to generate anatomical segmentation for 86 images in the U_C subset that lacked corresponding GT data. These generated predictions are referred to as Pseudo-Structure Masks, which supplement missing labels and serve as target masks for the subsequent phase of Multi-Task Learning. This process is represented in Figure 1 part (3).

Crucially, the custom data loader for the Hybrid MTL model was designed to prioritize real GT over pseudo-labels. Therefore, pseudo-masks were used only for the 86 images from U_C for which the real structure GT was genuinely absent. For images belonging to the U_{Int} set, the manually annotated real GT for structures was always loaded and used.

3.3.3. Training the Hybrid Multi-Task Learning Model

To address the dual objectives effectively, we adapted the U-Net++ architecture by incorporating a Dual-Head Decoder: one head dedicated to the multi-class segmentation of anatomical structures (8 classes) and a second head for binary caries detection. Both decoders share a pre-trained EfficientNet-B4 backbone, promoting feature reuse.

A contribution of this work is the customized loss strategy developed to supervise learning across these asymmetric tasks. For the anatomical structures, we used a combined loss approach ($\mathcal{L}_{Struct}$) that incorporates Dice Loss and Focal Loss with equal weights to balance overlap optimization with hard-sample mining, as described in Equation (2).

$$\mathcal{L}_{Struct} = 0.5 \cdot \mathcal{L}_{Dice} + 0.5 \cdot \mathcal{L}_{Focal} \tag{2}$$

For the caries detection head, we formulated a composite loss ($\mathcal{L}_{Caries}$) tailored to address the severe class imbalance inherent to dental radiography, where carious lesions typically occupy a disproportionately small fraction of the image area compared to dominant structures. This loss integrates Focal Tversky Loss ($\mathcal{L}_{FT}$) to prioritize precision by penalizing false positives, Soft Cross-Entropy ($\mathcal{L}_{SCE}$) for probabilistic stability, and Lovász-Softmax ($\mathcal{L}_{Lovasz}$) to directly optimize the Jaccard index, as described in Equation (3).

$$\mathcal{L}_{Caries} = 0.5 \cdot \mathcal{L}_{FT} + 0.2 \cdot \mathcal{L}_{SCE} + 0.3 \cdot \mathcal{L}_{Lovasz} \tag{3}$$

Finally, the total training objective ($\mathcal{L}_{Total}$) was calculated as a weighted sum of these specific losses, as described in Equation (4):

$$\mathcal{L}_{Total} = W_{Struct} \cdot \mathcal{L}_{Struct} + W_{Caries} \cdot \mathcal{L}_{Caries} \quad (4)$$

Task weights were set to $W_{Struct} = 0.8$ and $W_{Caries} = 0.2$ following preliminary model calibration. Assigning equal or greater weight to the caries task ($W_{Caries} \geq 0.5$) resulted in rapid overfitting of the shared encoder, as indicated by diverging validation loss. Sparse pathological annotations (D_C) did not sufficiently stabilize feature learning; therefore, a dominant weight was assigned to the dense structural task (D_S). This approach leverages anatomical consistency as a regularizer [17], promoting the learning of stable geometric features prior to optimizing for pathological deviations.

Given the hybrid nature of our dataset, the caries loss term ($\mathcal{L}_{Caries}$) is computed selectively: it is active only for training examples with verified caries annotations. Otherwise, it is masked to zero for structure-only images to prevent erroneous supervision.

As for the Single-Task Structure segmentation model, all input radiographs underwent the same standardization and augmentation pipeline. This included applying Contrast Limited Adaptive Histogram Equalization (CLAHE) for image enhancement, resizing to 512 × 512 pixels, and incorporating robust geometric and photometric data augmentation techniques (rotation, flipping, brightness shifts, gamma contrast). A comprehensive summary of the hyperparameters is provided in Table 3.

Table 3. Training Hyperparameters and Configuration for the Hybrid MTL Model.

Parameter	Configuration/Value
	Model Architecture
Architecture	Unet++ (Dual-Head Decoder)
Encoder	EfficientNet-B4 (ImageNet weights)
Input Size	512 × 512 pixels
Outputs	Head 1: Structures (8), Head 2: Caries (2 classes)
	Optimization & Loss
Optimizer	AdamW
Initial Learning Rate	5×10^{-4}
Scheduler	ReduceLROnPlateau (Patience = 10, Factor = 0.1)
Batch Size	4
Total Epochs	300
Loss Strategy	Struct: Dice+Focal | Caries: FT+SCE+Lovasz
Task Weights	$W_{Struct} = 0.8$, $W_{Caries} = 0.2$
Early Stopping	40 epochs (Monitor: DSC-Caries)
	Data Augmentation
Geometric	Horizontal Flip ($p = 0.5$), ShiftScaleRotate ($p = 0.7$)
Noise & Blur	GaussNoise ($p = 0.3$), RandomBlur ($p = 0.3$)
Intensity	RandomBrightnessContrast ($p = 0.5$), RandomGamma ($p = 0.5$)
Preprocessing	CLAHE (ClipLimit = 2.0), Normalization (ImageNet stats)

3.4. Single-Task Baseline Comparison

To evaluate the contribution of the proposed Hybrid MTL framework, we established a baseline model dedicated exclusively to the primary task of caries segmentation. This control experiment serves as a benchmark representing a conventional, specialized single-task approach. By comparing against this baseline, we aim to isolate and quantify the specific impact of the multi-task architecture and the integration of auxiliary anatomical data, independent of backbone capacity or loss-function specifics.

The baseline model is built upon the Unet++ architecture employing an EfficientNet-B4 encoder pre-trained on ImageNet. Structurally, the encoder backbone is identical to the one utilized in our proposed MTL network; however, this control model features only a single decoder head responsible for the binary caries segmentation task. To reflect a standard specialized training scenario and define the data scope clearly, this baseline was trained exclusively on the subset of images containing verified carious lesions ($N = 292$), with no exposure to anatomical structure annotations or pseudo-labels.

To provide a comparison, critical training configurations were maintained consistent with the caries-detection head of the proposed MTL model. Most notably, the baseline employed the same advanced composite loss function, integrating Focal Tversky, Soft Cross-Entropy, and Lovász-Softmax terms with the exact same weights to address severe class imbalance and boundary refinement. Furthermore, the same preprocessing (CLAHE) and data augmentation pipeline were applied. A comprehensive summary of the hyperparameters and configuration for this control experiment is provided in Table 4.

Table 4. Training Hyperparameters and Configuration for the Single-Task Control Model (Caries Only).

Parameter	Configuration/Value
Model Architecture	
Architecture	Unet++ (Single-Head Decoder)
Encoder	EfficientNet-B4 (ImageNet weights)
Backbone Status	Identical to MTL model (Shared Encoder)
Input Size	512×512 pixels
Outputs	Single Head: Caries (2 classes)
Data Scope	
Training Data	Only Caries-Positive images ($N = 292$)
Optimization & Loss	
Optimizer	AdamW
Initial Learning Rate	5×10^{-4}
Scheduler	ReduceLROnPlateau (Patience = 10, Factor = 0.1)
Batch Size	4
Total Epochs	300
Loss Strategy	Composite V5 (FT + SCE + Lovasz)
Loss Weights	$0.5 \cdot \mathrm{FT} + 0.2 \cdot \mathrm{SCE} + 0.3 \cdot \mathit{Lovasz}$
Early Stopping	40 epochs (Monitor: DSC)
Data Augmentation & Preprocessing	
Pipeline Status	Identical to MTL model pipeline
Preprocessing	CLAHE (ClipLimit = 2.0), Normalization

Performance Metrics

To evaluate the anatomical segmentation task, we used the Mean Intersection over Union (mIoU) and the Dice Similarity Coefficient (DSC) as primary metrics. These quantify the overlap between the predicted segmentation masks and the corresponding ground truth across the anatomical classes, excluding the background. The mIoU is computed as the macro-average of the per-class Intersection over Union, as detailed in Equation (5).

$$\mathrm{mIoU} = \frac{1}{C} \sum_{i=1}^{C} \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FP}_i + \mathrm{FN}_i} \tag{5}$$

where C represents the number of anatomical classes ($C = 7$, excluding background), and TP_i, FP_i, and FN_i denote the True Positives, False Positives, and False Negatives for class i, respectively.

For the caries segmentation task, given the clinical importance of minimizing missed diagnoses and the class imbalance inherent in pathological datasets, we prioritized the Dice Similarity Coefficient (Equation (6)) along with Recall, Precision, and Specificity.

$$\text{DSC (F1)} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \tag{6}$$

Sensitivity (Equation (7)) measures the proportion of actual carious pixels correctly identified, while Precision (Equation (8)) assesses the reliability of the positive predictions.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

Additionally, we computed Specificity (Equation (9)) to evaluate the model's ability to correctly distinguish structural tissue from pathology, thereby reducing false-positive diagnoses that could lead to overtreatment, where TN denotes the True Negatives (background or healthy tissue correctly classified as non-carious).

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{9}$$

4. Results

4.1. Implementation Details

The models were implemented using the PyTorch framework (version 2.0) and the Segmentation Models PyTorch library (version 0.3.3). All experiments were conducted using Google Colab Pro+ resources equipped with a single NVIDIA A100-SXM4 GPU (40GB VRAM). The average training time for the proposed Hybrid MTL model was approximately 3 h.

4.2. Validation of the Single-Task Structure Model and Pseudo-Labeling Strategy

Prior to evaluating the Hybrid MTL model, we validated the reliability of the Single-Task Structure Model trained on the labeled subset (see Figure 1, Part 2). The quantitative results for this preliminary validation are detailed in Table 5.

As presented in Table 5, the single-task anatomical model achieved a Global Mean IoU of 0.7071 and a Mean Dice Score of 0.8259. Notably, the model demonstrated high precision in segmenting key tissues relevant to caries progression, such as Dentin (Dice: 0.8668) and Enamel (Dice: 0.8268).

Table 5. Quantitative Results of the Single-Task Structure Segmentation model on the internal validation set. Metrics are reported for each anatomical class along with the global macro-average.

Anatomical Structure	IoU (Jaccard)	Dice Score (F1)
Dentin	0.7649	0.8668
Enamel	0.7047	0.8268
Implant	0.8371	0.9113
Pulp	0.5662	0.7230
Restoration	0.7160	0.8345
Foreign Object	0.7090	0.8297
Endodontics	0.6517	0.7891
Global Mean	0.7071	0.8259

It is pertinent to clarify that these metrics were calculated on an internal validation split, distinct from the final independent test set reserved for the comparative analysis. Since the primary function of this single-task model was to generate pseudo-labels, this validation confirms its capacity to provide reliable anatomical supervision for the unlabeled data without compromising the integrity of the final evaluation.

Qualitatively, the generation of pseudo-labels (See Figure 1, part (3)) proved effective. As illustrated in Figure 3, the model successfully propagated anatomical context to the unlabeled portion of the dataset, delineating structures such as enamel and dentin with high commitment even in the absence of manual ground truth.

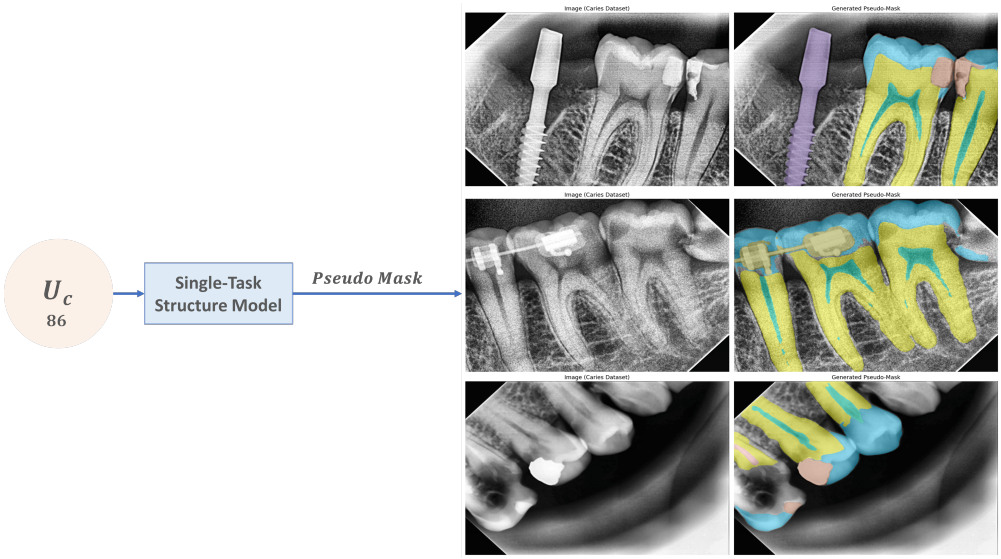


Figure 3. Visual assessment of generated pseudo-labels: The model infers anatomical structures from images lacking structural annotations (randomly selected examples shown).

4.3. Comparative Performance: Hybrid MTL vs. Single-Task Baseline

Training Dynamics and Convergence Stability

To assess the optimization behavior, we analyzed the comparative training history of both models until their respective early-stopping triggers. Figure 4 illustrates the evolution of Validation Loss and DSC, and reveals a distinct contrast in learning efficiency.

As shown in Figure 4a, the Single-Task Baseline (Red line) exhibited clear signs of overfitting. Quantitatively, its validation loss reached a minimum at Epoch 27 before monotonically diverging to 0.734 by Epoch 50 (an 11.8% degradation). This behavior confirms that the encoder began to memorize noise in the limited training samples.

In sharp contrast, the Hybrid MTL framework (Blue line) demonstrated remarkable stability. The caries-specific loss stabilized around 0.655 and remained flat, effectively mitigating the divergence observed in the baseline. This stability is fundamentally linked to the findings in Figure 4b, where the Anatomical Task (Green line) quickly reached a high-performance plateau (mIoU $\approx$ 0.70).

This dense structural supervision served as a regularizing scaffold, leading to the improved MTL Caries DSC (Blue line) relative to the baseline (Red dashed line), validating the hypothesis that simultaneous anatomical learning prevents overfitting in low-data regimes.

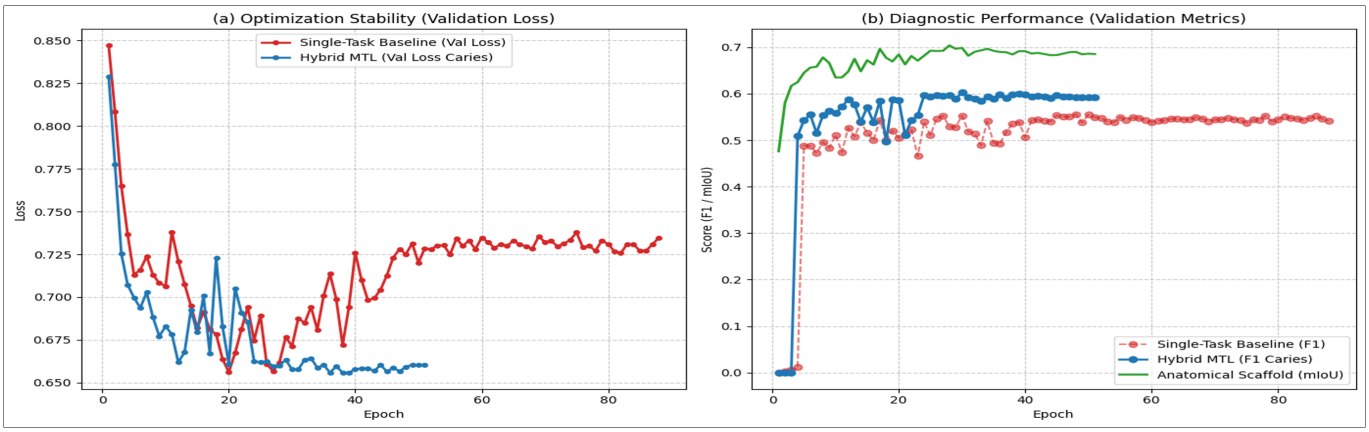


Figure 4. Comparative Training Dynamics. (**a**) Stability in Optimization: The Single-Task Baseline (Red) shows considerable divergence after Epoch 30 owing to overfitting, whereas the Hybrid MTL (Blue) keeps a consistent loss trajectory. (**b**) Performance in Diagnostics: The Anatomical Task (Green) serves as a foundation, enabling the MTL Caries detection (Blue) to surpass the Baseline (Red dashed line) consistently.

4.4. Quantitative Evaluation and Confusion Matrix Analysis

To verify the models' efficacy, we evaluated them on the test set (N = 31) held out during the training and validation phases, as described in Figure 1. Table 6 summarizes the performance metrics for the binary caries detection task.

Table 6. Comparative Performance Metrics for Caries Segmentation on the Test Set: Single-Task Baseline vs. Hybrid Multi-Task Learning (MTL) Model.

Model	TP	FP	FN	TN	DSC	IoU	Sensitivity	Precision	Specificity
Single-Task Baseline	78,969	39,750	47,208	7,960,537	0.6449	0.4759	0.6259	0.6652	0.9950
MTL	88,400	49,089	37,777	7,951,198	0.6705	0.5044	0.7006	0.6430	0.9939

The proposed Hybrid MTL framework consistently outperformed the Single-Task Baseline across the primary segmentation metrics. Specifically, the Hybrid model achieved a DSC of 0.6705, representing a 3.98% relative improvement over the baseline (0.6449). Similarly, the Intersection over Union (IoU) increased by 5.99% (from 0.4759 to 0.5044).

Most notably, the model demonstrated a substantial increase in Sensitivity, rising from 0.6259 to 0.7006. This indicates that incorporating the auxiliary anatomical task provided a crucial contextual scaffold, enabling the model to identify subtle pathological regions that the single-task counterpart had previously missed.

To gain deeper insight into the classification behavior beyond global metrics, we analyzed the pixel-wise confusion matrices for both models. Figure 5 visually depicts the distribution of True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN).

A pivotal finding in this study is the substantial reduction in FN achieved by the proposed Hybrid MTL framework. The Single-Task Baseline model failed to detect pathological tissue in 47,208 pixels. In contrast, the MTL model successfully reduced this figure to 37,777, representing a 19.9% decrease in undetected carious areas. This improvement is clinically significant, as undetected lesions can lead to unchecked progression and irreversible loss of tooth structure.

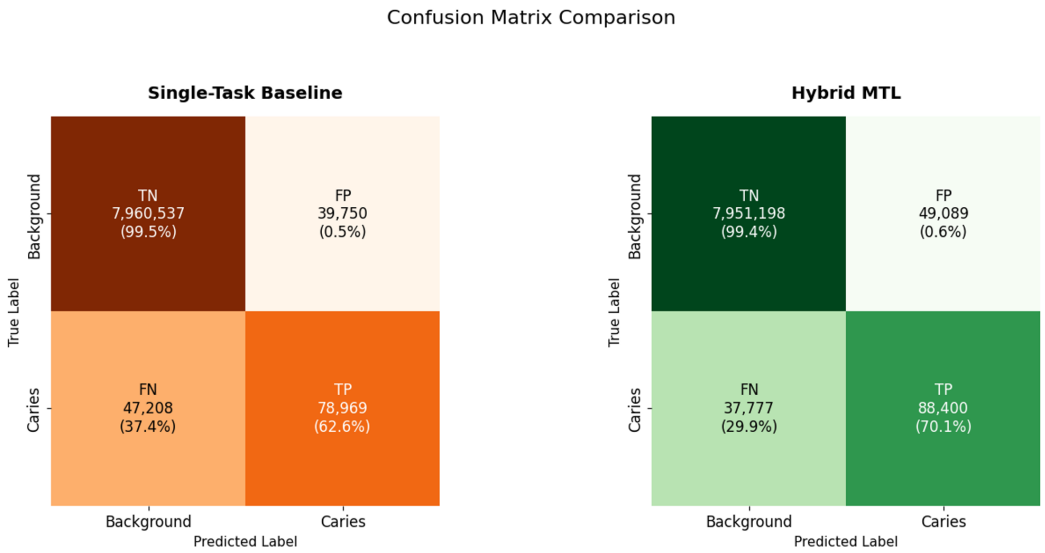


Figure 5. Comparative Confusion Matrices (Pixel-wise counts). (**Left**): Single-Task Baseline. (**Right**): Hybrid MTL. The MTL model significantly shifts pixels from False Negatives (bottom-left quadrant) to True Positives (bottom-right quadrant), indicating superior lesion capture.

However, this enhanced detection capability comes with an important trade-off. The significant increase in Sensitivity (from 0.6259 to 0.7006) was accompanied by an increment in False Positives (from 39,750 to 49,089 pixels) and a slight decrease in Precision (from 0.6652 to 0.6430).

This shift indicates that by leveraging the anatomical structural context as a scaffold, the MTL model became significantly more sensitive to subtle demineralization patterns often overlooked by the baseline. While this enhancement results in a less conservative segmentation profile, characterized by a higher false alarm rate, such a trade-off is often strategically advantageous in clinical screening tools. The primary objective is to maximize the detection of potential pathologies for human verification.

4.5. Qualitative Visual Assessment

To complement the quantitative metrics, a visual analysis of representative cases from the test set, depicted in Figures 6 and 7, is presented to illustrate the strengths and limitations of incorporating anatomical context.

In Figure 6, the first row presents a extended annotated caries lesion, ensuring clarity in the analysis. The MTL model demonstrates consistently high performance, reflecting its stability in fundamental detection capabilities. Rows 2 and 3 in Figure 6 provide visual validation of the anatomical scaffold concept, illustrating how the MTL framework significantly enhances the segmentation boundaries of lesions that were only partially detected by the baseline model. This improvement results in marked enhancements in Dice scores.

Most critically, Figure 6, Row 4, illustrates the marked improvement in sensitivity driving the quantitative results. In this scenario, with a subtle lesion, the single-task baseline fails (False Negative, Dice 0.000). However, the Hybrid MTL model successfully localizes the pathology. This visual evidence supports the premise that incorporating anatomical context facilitates the capture of incipient or complex lesions that specialized models often miss.

Furthermore, we conduct a failure analysis in Figure 7, Row 1, indicating that, in specific instances, the hybrid MTL approach might obscure subtle pathological features that do not conform to anticipated structural patterns. Furthermore, in highly ambiguous or noisy images (Figure 7, Row 2), the increased sensitivity of the MTL model can lead to slight over-segmentation in adjacent healthy tissues compared to the more conservative

baseline, evidencing the observed quantitative trade-off, in which a significant gain in recall accompanies a minor decrease in precision.

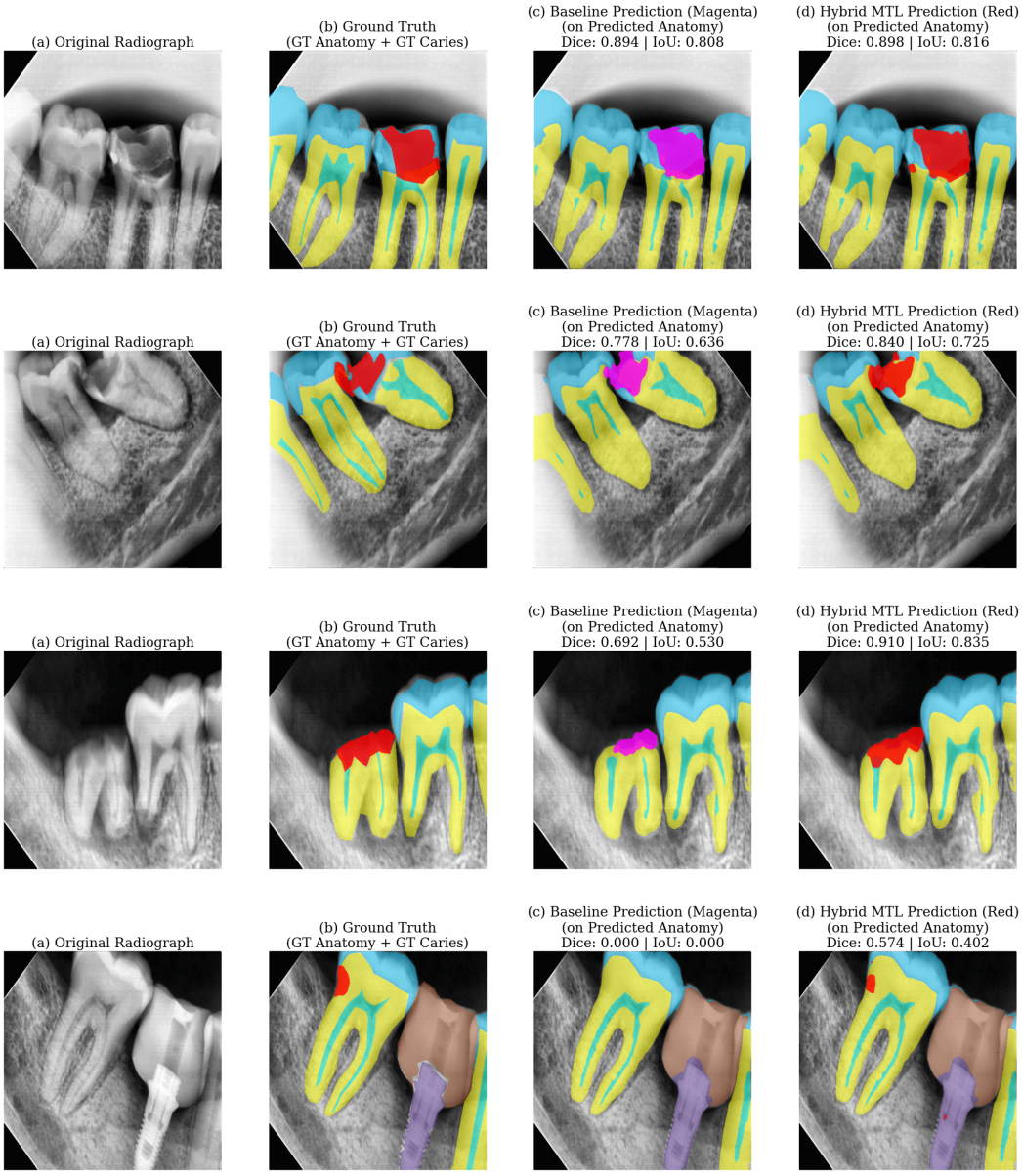


Figure 6. Qualitative assessment demonstrating the impact of the Hybrid MTL framework in varied clinical scenarios. Columns: (**a**) Original Radiograph; (**b**) Complete Ground Truth; (**c**) Single-Task Baseline Prediction; (**d**) Hybrid MTL Prediction. Dice scores for the caries class are provided. (**Row 1**): A lesion where both models perform equivalently well, indicating that the MTL framework maintains robustness in clear-cut cases. (**Rows 2 and 3**): Examples illustrating the MTL model's less conservative segmentation profile driven by higher sensitivity. The simultaneous anatomical learning allows it to capture more extensive pathological areas and improve completeness compared to the predictions of the baseline. (**Row 4**): A scenario where the baseline model completely misses a subtle lesion (Dice 0.000), while the MTL model identifies it (Dice 0.500), aided by the anatomical scaffold. The annotation colors follow the scheme described in Figure 2.

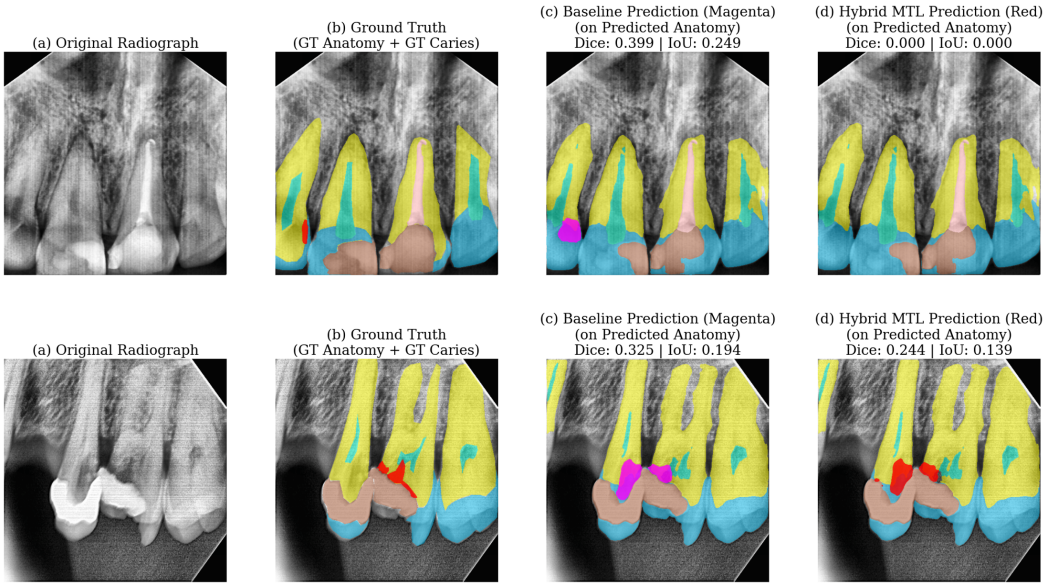


Figure 7. Analysis of challenging cases and limitations in the test set. (**Row 1**): In this example, the Hybrid MTL model fails to identify a lesion that the Single-Task baseline successfully detects. This illustrates a scenario in which strong anatomical priors from the auxiliary task could obscure subtle pathological features that do not conform to learned structural patterns. (**Row 2**): The MTL model exhibits slightly more over-segmentation in ambiguous areas adjacent to the lesion compared to the conservative baseline. This behavior visually reflects the quantitative trade-off between sensitivity and precision, where higher recall comes at the cost of increased false positives in certain regions. The annotation colors follow the scheme described in Figure 2.

5. Discussion

The main goal of this study was to determine whether incorporating anatomical knowledge into a caries detection model via Multi-Task Learning (MTL) could enhance segmentation robustness on complex periapical radiographs. The experimental results obtained provide empirical support for this approach. The comparative analysis (Table 6) demonstrated that integrating simultaneous structural learning enhanced overall diagnostic performance compared to the specialized single-task baseline, raising the DSC from 0.6449 to 0.6706.

The data show that this improvement was not uniform; it primarily resulted from an increase in Sensitivity, which led to a clinically significant 19.9% reduction in false-negative pixels (undetected carious areas). The enhanced detection capability achieved comes with a discernible trade-off. Specifically, the model shifted to a less conservative detection profile, resulting in a significant increase in recall. This improvement, however, was accompanied by a rise in false positives and a slight reduction in precision, as indicated by a decrease from 0.6652 to 0.6430 (see Table 6). This finding indicates that incorporating anatomical context enhances sensitivity at a minor cost to precision and will serve as the foundation for interpreting the mechanisms and clinical utility of our approach.

While the Hybrid MTL model demonstrates strong performance in lesion localization (high sensitivity), the observed reduction in precision necessitates careful clinical interpretation. The model's tendency toward over-segmentation indicates that, although it serves as a robust alert system for detecting demineralization, it may overestimate the actual volumetric extent of lesions. In restorative workflows, exclusive reliance on the model's predicted boundaries can lead to excessive tissue removal if not carefully reviewed by clinicians.

Accordingly, we position this tool primarily as a diagnostic aid to minimize missed diagnoses (false negatives), while the precise determination of excavation margins remains the clinician's responsibility. This approach is consistent with a human-in-the-loop

workflow, in which high sensitivity ensures that subtle lesions are identified for further inspection, and professional judgment is applied to address potential overestimation during treatment planning. As a result, the current version of the model is most appropriate for triage and screening applications, rather than for automated surgical guidance.

We posit that the observed surge in sensitivity is directly attributable to the strategic design of our training objective, specifically the heavy weighting assigned to the auxiliary structural task $W_{Struct} = 0.8$. By prioritizing the learning of anatomical geometry, the shared encoder was forced to develop spatial representations of the tooth structure. This approach enabled the caries detection head to more easily isolate pathological deviations from the established norm. Consequently, the model gained the capacity to detect a significant subset of subtle, incipient lesions that were overlooked by the single-task baseline, as evidenced by the reduction in global False Negatives.

It is important to note a distinct difference in training dynamics. While the Single-Task Baseline exhibited clear signs of overfitting, evidenced by a validation loss that reached a minimum (0.656) and subsequently deteriorated, rising to 0.73 by the end of training. In stark contrast, the Hybrid MTL model demonstrated superior regularization: its validation loss descended and stabilized (0.655), maintaining a consistent performance plateau without the divergence observed in the single-task model. This stability suggests that concurrent structural learning effectively constrained the shared encoder, preventing it from memorizing noise and ensuring stronger generalization.

We argue that the weighted structural task functioned as an implicit regularizer for the shared encoder. The comparative training dynamics empirically support this mechanism (see Figure 4). While the single-task baseline exhibited clear signs of overfitting, the MTL framework successfully constrained the solution space to anatomically plausible features

Table 7 benchmarks our results against existing literature focused on periapical radiography. It is crucial to interpret these metrics in the context of data availability and annotation quality. Recent studies utilizing large, fully annotated private datasets, such as Priya & Raja [33] and Ari et al. [35], have set high performance ceilings (DSC of 95.8% and 82.0%, respectively).

However, our work addresses a fundamentally different challenge: robustness under data scarcity and incompleteness. When compared to Khan et al. [32], who operated under similar constraints of a limited public dataset size (200 images), our Hybrid MTL framework demonstrates a substantial superiority (67.06% vs. 53.20%). This confirms that even without the luxury of massive, perfect ground truth, anchoring the model with anatomical context significantly boosts performance, offering a viable path for real-world clinical implementation where data is often imperfect.

Table 7. Comparison with State-of-the-Art Deep Learning Models focused exclusively on Periapical Radiography.

Study	Dataset Size	Data Condition	Architecture	Metric	Result
Priya & Raja (2025) [33]	1400 imgs	Private/Full GT	Caries-SegNet	DSC	95.80%
Ari et al. (2022) [35]	1169 imgs	Private/Full GT	U-Net + CLAHE	DSC	82.00%
Khan et al. (2021) [32]	200 imgs	Public/Limited	U-Net	DSC	53.20%
Ours (Hybrid MTL)	292 imgs	Partial/Asymmetric	MTL U-Net++	DSC	67.06%

Note: "Data Condition" refers to the completeness of annotations. Our method is designed for scenarios with missing or asymmetric labels.

Having established the competitive standing of our caries detection performance, it is equally vital to verify the stability of the auxiliary task. A critical concern in Multi-Task Learning is the phenomenon of negative transfer, where the addition of a secondary task degrades the performance of the primary one. Table 8 dispels this concern for our framework.

While the Hybrid MTL model exhibited a slight decrease in structural segmentation metrics (mIoU: 75.17% vs. 75.10%) compared to the dedicated single-task model, this reduction is consistent with the strategic reallocation of model capacity.

By imposing a dominant weight on the structural loss ($W_{Struct} = 0.8$), we ensured that the encoder maintained a robust anatomical representation (Dice > 82%) sufficient to serve as a contextual scaffold. Crucially, this slight sacrifice in anatomical precision enabled the shared encoder to accommodate the conflicting features required for the primary caries detection task, ultimately yielding the highest performance in pathology identification (DSC: 67.05%). This demonstrates an effective prioritization where the auxiliary task supports the primary goal without suffering catastrophic degradation.

Table 8. Performance Comparison across Task Strategies: Structural Segmentation vs. Caries Detection.

Model/Strategy	Task Focus	mIoU (Struct)	Dice (Struct)	Caries DSC
Single-Task (Structure)	Anatomy Only	**75.10%**	**85.78%**	-
Single-Task (Baseline)	Caries Only	-	-	64.49%
Hybrid MTL (Ours)	Anatomy + Caries	70.19%	82.48%	67.05%

Paired Wilcoxon signed-rank tests were performed on the test set (N = 31) for both Dice and Sensitivity metrics to evaluate the statistical robustness of the observed improvements. The resulting p-values exceeded 0.05 (Dice: $p = 0.69$; Sensitivity: $p = 0.80$), indicating that the per-image performance differences are not statistically significant with the current sample size.

However, a critical distinction exists between per-image averages and global pixel-level metrics. Statistical tests assign equal weight to each image regardless of lesion size, whereas global metrics, as reported in Table 6, aggregate predictions across the entire dataset. Clinically, the most significant result is the 19.9% reduction in total False Negative pixels. Although pixel-wise metrics quantify segmentation accuracy, this reduction specifically serves as a proxy for lesion-level detection, demonstrating that the Hybrid MTL model recovered substantial pathological volumes missed by the baseline. In a diagnostic context, this improvement reduces the likelihood of overlooking active disease, directly addressing the clinician's primary concern of minimizing missed lesions. While the model may not consistently outperform the baseline on individual images due to high variance, it shows a superior capacity to capture the overall disease burden within the cohort.

Beyond diagnostic metrics, this study makes a primary methodological contribution. The promising results of our Hybrid framework suggest the viability of leveraging heterogeneous, partially annotated datasets. In medical imaging, where obtaining fully annotated data for multiple tasks is often prohibitively expensive, strategies that maximize data utility are critical. By designing a system that effectively leverages images lacking complete Ground Truth through strategic masking and loss calculations, we propose a potential pathway to develop context-aware dental AI models in real-world scenarios where data incompleteness is the norm.

Despite these promising findings, several inherent limitations in the study design must be acknowledged. The independent test set (N = 31) consists exclusively of radiographs with confirmed carious lesions. Although this composition restricts the assessment of specificity in healthy patients, it directly tests the hypothesis that anatomical scaffolding enhances sensitivity in pathological cases. Therefore, the reported metrics reflect the model's ability to detect lesions within diseased cohorts, rather than its performance in a general screening population that includes healthy individuals. Future studies should include caries-free controls to enable a comprehensive evaluation of the model's false-positive rate in broader clinical settings.

Furthermore, our qualitative failure analysis indicated that in highly noisy images or those with complex metallic restorations, the strong anatomical priors might occasionally obscure atypical pathological features or lead to over-segmentation.

Since the dataset is derived from a single clinical center, the model is tailored to the specific noise distribution and intensity histograms of the acquisition device. In cross-center deployment scenarios, differences in sensor sensitivity, voltage (kVp), and exposure time often introduce a domain shift that can reduce model performance. The proposed Hybrid MTL framework provides a theoretical advantage over single-task models in addressing this issue. Anchoring the encoder to anatomical structures, which are geometrically consistent across populations and imaging protocols, allows the auxiliary task to serve as a stabilizing prior.

For clinical translation, the model is initially intended for integration into an existing radiographic visualization software suite. Within this deployment context, the model functions in a controlled software environment, which ensures a standardized workflow and minimizes exposure to extreme acquisition variability. To enhance vendor-agnostic generalizability, future research will expand the training cohort using a large-scale, multi-centric dataset from diverse sources.

Consequently, future research should address two key areas: collecting a balanced dataset with caries-free controls to enable comprehensive evaluation of specificity in broader clinical settings, and exploring uncertainty quantification methods to reduce over-segmentation in ambiguous regions.

6. Conclusions

This study establishes that incorporating anatomical context through a Hybrid Multi-Task Learning framework is an effective strategy for enhancing caries segmentation in periapical radiographs, particularly when training data is incomplete. By prioritizing the learning of healthy dental structures ($W_{Struct} = 0.8$), our approach successfully leveraged anatomical knowledge as a stabilizing scaffold, leading to a marked increase in sensitivity and a reduction in undetected lesions (19.9%) compared to a specialized single-task baseline.

While this gain in detection capability entails a shift towards a less conservative profile with slightly lower precision, we argue that this trade-off is clinically advantageous for diagnostic support, where minimizing false negatives is paramount. Methodologically, our work validates a scalable approach to train sophisticated models using heterogeneous, partially annotated datasets, a critical step forward for medical AI in resource-constrained environments. Future efforts extending this context-aware paradigm to larger, balanced cohorts and to 3D imaging will be essential to further refine the reliability of automated dental diagnostics.

Author Contributions: Conceptualization, H.A.G.-O., L.F.L.-V. and T.I.-P.; methodology, J.A.N.-P. and F.G.-V.; software, F.G.-V., J.R.G.-R. and R.S.-A.; validation, J.A.N.-P., S.I.-D. and R.S.-A.; formal analysis, J.A.N.-P., S.I.-D. and V.I.R.-A.; investigation, J.A.N.-P. and S.I.-D.; resources, H.A.G.-O. and L.F.L.-V.; data curation, F.G.-V. and R.S.-A.; writing—original draft preparation, J.A.N.-P., J.R.G.-R. and S.I.-D.; writing—review and editing, H.A.G.-O., L.F.L.-V. and T.I.-P.; visualization, F.G.-V. and V.I.R.-A.; supervision, H.A.G.-O. and L.F.L.-V.; project administration, H.A.G.-O. and J.R.G.-R.; funding acquisition, T.I.-P. and V.I.R.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Zentia Dental Clinic (protocol code CEI-ZDE-005-2024 on 30 April 2024. All personal data were handled and protected in accordance with the clinic's internal privacy regulations.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: The datasets employed in this study can be found at: https://universe.roboflow.com/dental-7rj2x/zentia_periapicaldataset_cariesonly-cbxbu (accessed on 4 December 2025) for the caries dataset, and https://universe.roboflow.com/dental-7rj2x/zentia_periapicaldataset_structures-ri2aw (accessed on 4 December 2025) for the structure dataset.

Acknowledgments: The authors want to thank the Mexican Secretariat of Science, Humanities, Technology and Innovation (SECIHTI by its initials in Spanish) for its support to the National Laboratory of Embedded Systems, Advanced Electronics Design and Micro Systems (LN-SEDEAM by its initials in Spanish), project numbers 282357, 293384, 299061, 314841, 315947, and 321128 and scholarship number 1012274. Additionally, the authors express their gratitude to the Zentia Dental Clinic (https://www.facebook.com/clinicadentalfernandezynava) (accessed on 4 December 2025) and its specialists for their valuable assistance in creating the dataset used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MTL	Multi-Task Learning
AI	artificial intelligence
D_S	Structure Dataset
D_C	Caries Dataset
U_S	Structure-Only Dataset
U_C	Caries-Only Dataset
U_{Int}	Intersection Dataset
$\mathcal{U}$	Universe Dataset
$\mathcal{T}_{Structure}$	Training Structure Dataset
$\mathcal{T}_{final}$	Final Test Set
GT	Ground Truth Mask
$\mathcal{L}_{Struct}$	Structure Loss (Combined Dice + Focal)
$\mathcal{L}_{Caries}$	Caries Loss (Composite FT + SCE + Lovász)
$\mathcal{L}_{Dice}$	Dice Loss
$\mathcal{L}_{Focal}$	Focal Loss
$\mathcal{L}_{FT}$	Focal Tversky Loss
$\mathcal{L}_{SCE}$	Soft Cross-Entropy Loss
$\mathcal{L}_{Lovasz}$	Lovász-Softmax Loss
$\mathcal{L}_{Total}$	Total Weighted Objective Loss
W_{Struct}	Weight for Structure Task
W_{Caries}	Weight for Caries Task
mIoU	Mean Intersection over Union
DSC	Dice Similarity Coefficient

References

1. Kane, S.F. The effects of oral health on systemic health. *Gen. Dent.* **2017**, *65*, 30–34.
2. Pitts, N.B.; Twetman, S.; Fisher, J.; Marsh, P.D. Understanding dental caries as a non-communicable disease. *Br. Dent. J.* **2021**, *231*, 749–753. [CrossRef] [PubMed]
3. Cheng, L.; Zhang, L.; Yue, L.; Ling, J.; Fan, M.; Yang, D.; Huang, Z.; Niu, Y.; Liu, J.; Zhao, J.; et al. Expert consensus on dental caries management. *Int. J. Oral Sci.* **2022**, *14*, 17. [CrossRef] [PubMed]

4. Paolone, G.; Romo-Sáenz, C.I.; Chavez-Reyes, E.M.; Gomez-Flores, R.; González-Flores, M.N.; Sosa-Martínez, R.; Cruz-Fierro, N.; Elizondo-Elizondo, J.; de la Garza-Ramos, M.A. Food Insecurity and Dental Caries in Rural Mexican Populations. *Int. J. Dent.* **2023**, *2023*, 6694259. [CrossRef]
5. Moreno-Barrera, A.; Morales-Ruiz, P.; Ribas Pérez, D.; Flores-Fraile, J.; Castaño-Seiquer, A. Analysis and Evaluation of Dental Caries in a Mexican Population: A Descriptive Transversal Study. *Int. J. Environ. Res. Public Health* **2023**, *20*, 3873. [CrossRef] [PubMed]
6. Bukhari, O.M. Dental Caries Experience and Oral Health Related Quality of Life in Working Adults. *Saudi Dent. J.* **2020**, *32*, 382–389. [CrossRef]
7. Bhatnagar, D.M. Oral health: A gateway to overall health. *Contemp. Clin. Dent.* **2021**, *12*, 211–212. [CrossRef]
8. Schwarz, C.; Hajdu, A.I.; Dumitrescu, R.; Sava-Rosianu, R.; Bolchis, V.; Anusca, D.; Hanghicel, A.; Fratila, A.D.; Oancea, R.; Jumanca, D.; et al. Link between Oral Health, Periodontal Disease, Smoking, and Systemic Diseases in Romanian Patients. *Healthcare* **2023**, *11*, 2354. [CrossRef]
9. Zanini, L.G.K.; Rubira-Bullen, I.R.F.; Nunes, F.L.D.S. A Systematic Review on Caries Detection, Classification, and Segmentation from X-ray Images: Methods, Datasets, Evaluation, and Open Opportunities. *J. Imaging Inform. Med.* **2024**, *37*, 1824–1845. [CrossRef]
10. Zandona, A.F.; Longbottom, C. Detection and Assessment of Dental Caries: A Clinical Guide. In *Detection and Assessment of Dental Caries: A Clinical Guide*; Zandona, A.F., Longbottom, C., Eds.; Springer: Cham, Switzerland, 2019.
11. Qayyum, A.; Tahir, A.; Butt, M.A.; Luke, A.; Abbas, H.T.; Qadir, J.; Arshad, K.; Assaleh, K.; Imran, M.A.; Abbasi, Q.H. Dental caries detection using a semi-supervised learning approach. *Sci. Rep.* **2023**, *13*, 749. [CrossRef]
12. Pitts, N.B.; Banerjee, A.; Mazevet, M.E.; Goffin, G.; Martignon, S. From 'ICDAS' to 'CariesCare International': The 20-year journey building international consensus to take caries evidence into clinical practice. *Br. Dent. J.* **2021**, *231*, 769–774. [CrossRef]
13. Litzenburger, F.; Lederer, A.; Kollmuß, M.; Hickel, R.; Kunzelmann, K.H.; Heck, K. Near-infrared transillumination with high dynamic range imaging for occlusal caries detection in vitro. *Lasers Med. Sci.* **2020**, *35*, 2049–2058. [CrossRef]
14. Ossowska, A.; Kusiak, A.; Świetlik, D. Artificial Intelligence in Dentistry—Narrative Review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 3449. [CrossRef]
15. Molteni, R. The Way We Were (and How We Got Here): Fifty Years of Technology Changes in Dental and Maxillofacial Radiology. *Dentomaxillofac. Radiol.* **2021**, *50*, 20200133. [CrossRef]
16. Rasnayaka, S.; Leuke Bandara, D.; Jayasundara, A.; Jayasinghe, R.; Wimalasiri, C.; Rathnayake, P.; Wijerathne, S.; Ragel, R.; Thambawita, V.; Nawinne, I. DenPAR: Annotated Intra-Oral Periapical Radiographs Dataset for Machine Learning. *Sci. Data* **2025**, *12*, 1615. [CrossRef] [PubMed]
17. Zhao, Y.; Wang, X.; Che, T.; Bao, G.; Li, S. Multi-task deep learning for medical image computing and analysis: A review. *Comput. Biol. Med.* **2023**, *153*, 106496. [CrossRef] [PubMed]
18. Vinayahalingam, S.; Kempers, S.; Limon, L.; Deibel, D.; Maal, T.; Hanisch, M.; Bergé, S.; Xi, T. Classification of Caries in Third Molars on Panoramic Radiographs Using Deep Learning. *Sci. Rep.* **2021**, *11*, 12609. [CrossRef] [PubMed]
19. Oztekin, F.; Katar, O.; Sadak, F.; Yildirim, M.; Cakar, H.; Aydogan, M.; Ozpolat, Z.; Talo Yildirim, T.; Yildirim, O.; Faust, O.; et al. An Explainable Deep Learning Model to Predict Dental Caries Using Panoramic Radiograph Images. *Diagnostics* **2023**, *13*, 226. [CrossRef]
20. Mirzaei, O.; Bilgehan, B.; Abduljalil, M.; Saleh, M.; Kayssoun, A.; İlhan, A. Performance Evaluation of Deep Learning Models for Dental Caries Classification via Panoramic Radiograph Images. *Clin. Exp. Health Sci.* **2024**, *14* , 868–875. [CrossRef]
21. Imak, A.; Celebi, A.; Siddique, K.; Turkoglu, M.; Sengur, A.; Salam, I. Dental Caries Detection Using Score-Based Multi-Input Deep Convolutional Neural Network. *IEEE Access* **2022**, *10*, 18320–18329. [CrossRef]
22. Megalan Leo, L.; Kalapalatha Reddy, T. Learning Compact and Discriminative Hybrid Neural Network for Dental Caries Classification. *Microprocess. Microsyst.* **2021**, *82*, 103836. [CrossRef]
23. Liu, J.; Liu, X.; Shao, Y.; Gao, Y.; Pan, K.; Jin, C.; Ji, H.; Du, Y.; Yu, X. Periapical Lesion Detection in Periapical Radiographs Using the Latest Convolutional Neural Network ConvNeXt and Its Integrated Models. *Sci. Rep.* **2024**, *14*, 25429. [CrossRef]
24. Pérez de Frutos, J.; Holden Helland, R.; Desai, S.; Nymoen, L.C.; Langø, T.; Remman, T.; Sen, A. AI-Dentify: Deep Learning for Proximal Caries Detection on Bitewing X-ray—HUNT4 Oral Health Study. *BMC Oral Health* **2024**, *24*, 344. [CrossRef]
25. Ayhan, B.; Ayan, E.; Karadağ, G.; Bayraktar, Y. Evaluation of Caries Detection on Bitewing Radiographs: A Comparative Analysis of the Improved Deep Learning Model and Dentist Performance. *J. Esthet. Restor. Dent.* **2025**, *37*, 1949–1961. [CrossRef] [PubMed]
26. Alsolamy, M.; Nadeem, F.; Azhari, A.A.; Ahmed, W.M. Automated Detection, Localization, and Severity Assessment of Proximal Dental Caries from Bitewing Radiographs Using Deep Learning. *Diagnostics* **2025**, *15*, 899. [CrossRef] [PubMed]
27. Sheng, C.; Wang, L.; Huang, Z.; Wang, T.; Guo, Y.; Hou, W.; Xu, L.; Wang, J.; Yan, X. Transformer-Based Deep Learning Network for Tooth Segmentation on Panoramic Radiographs. *J. Syst. Sci. Complex.* **2023**, *36*, 257–272. [CrossRef]
28. Deb, M.; Deb, M.; Dhar, M.K. A Deep Learning Approach to Teeth Segmentation and Orientation from Panoramic X-rays. *Signals* **2025**, *6*, 40. [CrossRef]

29. Yu, H.; Ye, X.; Hong, W.; Shi, R.; Ding, Y.; Liu, C. A Cascading Learning Method with SegFormer for Radiographic Measurement of Periodontal Bone Loss. *BMC Oral Health* **2024**, *24*, 325. [CrossRef]
30. Wang, C.-W.; Huang, C.-T.; Lee, J.-H.; Li, C.-H.; Chang, S.-W.; Siao, M.-J.; Lai, T.-M.; Ibragimov, B.; Vrtovec, T.; Ronneberger, O.; et al. A Benchmark for Comparison of Dental Radiography Analysis Algorithms. *Med. Image Anal.* **2016**, *31*, 63–76. [CrossRef]
31. Lee, S.; Oh, S.I.; Jo, J.; Kang, S.; Shin, Y.; Park, J.W. Deep Learning for Early Dental Caries Detection in Bitewing Radiographs. *Sci. Rep.* **2021**, *11*, 16807. [CrossRef] [PubMed]
32. Khan, H.; Haider, M.; Ansari, H.; Ishaq, H.; Kiyani, A.; Sohail, K.; Muhammad, M.; Khurram, S.A. Automated Feature Detection in Dental Periapical Radiographs Using Deep Learning. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **2020**, *131*, 600–608. [CrossRef]
33. Priya, J.; Raja, S.K.S. Caries-SegNet: Multi-Scale Cascaded Hybrid Spatial Channel Attention Encoder–Decoder for Semantic Segmentation of Dental Caries. *Biomed. Tech.* **2025**, *70*, 241–258. [CrossRef]
34. Ibraheem, W.I.; Jain, S.; Ayoub, M.N.; Namazi, M.A.; Alfaqih, A.I.; Aggarwal, A.; Meshni, A.A.; Almarghlani, A.; Alhumaidan, A.A. Assessment of the Diagnostic Accuracy of Artificial Intelligence Software in Identifying Common Periodontal and Restorative Dental Conditions (Marginal Bone Loss, Periapical Lesion, Crown, Restoration, Dental Caries) in Intraoral Periapical Radiographs. *Diagnostics* **2025**, *15*, 1432. [CrossRef]
35. Ari, T.; Sağlam, H.; Öksüzoğlu, H.; Kazan, O.; Bayrakdar, İ.Ş.; Duman, S.B.; Çelik, Ö.; Jagtap, R.; Futyma-Gąbka, K.; Różyło-Kalinowska, I.; et al. Automatic Feature Segmentation in Dental Periapical Radiographs. *Diagnostics* **2022**, *12*, 3081. [CrossRef]
36. Jagtap, R.; Samata, Y.; Parekh, A.; Tretto, P.; Roach, M.D.; Sethumanjusha, S.; Tejaswi, C.; Jaju, P.; Friedel, A.; Briner Garrido, M.; et al. Clinical Validation of Deep Learning for Segmentation of Multiple Dental Features in Periapical Radiographs. *Bioengineering* **2024**, *11*, 1001. [CrossRef] [PubMed]
37. Agrawal, P.; Nikhade, P. Artificial Intelligence in Dentistry: Past, Present, and Future. *Cureus* **2022**, *14*, e27405. [CrossRef] [PubMed]
38. Ali, M.; Irfan, M.; Ali, T.; Wei, C.R.; Akilimali, A. Artificial Intelligence in Dental Radiology: A Narrative Review. *Ann. Med. Surg.* **2025**, *87*, 2212–2217. [CrossRef] [PubMed]
39. Zhou, Z.; Rahman, M.; Tajbakhsh, N.; Liang, J. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* **2023**, *42*, 777–787. [CrossRef]
40. Khurshid, Z.; Faridoon, F.; Silkosessak, O.C.; Trachoo, V.; Waqas, M.; Hasan, S.; Porntaveetus, T. Multi-regional deep learning models for identifying dental prostheses on panoramic radiographs. *Sci. Rep.* **2025**, *15*, 11955.
41. Yeung, M.; Sala, E.; Schönlieb, C.B.; Rundo, L. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.