RESEARCH ARTICLE

Dental Classes Classification using TEYOLOv8 Network

Victor Kumar Chilaka[1], Pavan Sandula[2,*] and Jagadeesh Thati[2]

[1] Department of Electronics and Communication Engineering, Tirumala Institute of Technology and Sciences, Narasaraopet 522549, India

[2] Department of Electronics and Communication Engineering, Tirumala Engineering College, Narasaraopet 522601, India

Abstract

Accurate and automated detection of dental anatomy is essential for diagnosis and treatment planning. This study proposes a Transformer-Embedded YOLOv8 (TEYOLOv8) network to improve detection and classification of seven dental classes. The model enhances localization and segmentation accuracy by embedding an attentive transformer mechanism into the YOLOv8 framework. The TEYOLOv8 architecture integrates data augmentation, feature localization, segmentation, and classification. It employs C2f (Cross-Stage Partial Focus) convolutional blocks to preserve partial segmentation features and introduces a C2fAttn (Cross-Stage Feature Attention) module to capture fine-grained spatial details such as shape and position, addressing feature dissimilarity and loss. The model was trained and evaluated on a custom dental dataset with metrics including mean Average Precision (mAP), precision, and recall. Experimental results demonstrate that TEYOLOv8 achieves superior performance, with precision of 0.954, recall of 0.973, mAP@0.5 of 0.981, and mAP@0.5:0.95 of 0.794. The integration of the attention mechanism substantially improves feature representation and segmentation quality, enabling precise localization and classification of complex dental structures. Clinical Significance: The TEYOLOv8 model provides an efficient and accurate tool for dental image analysis, supporting precise identification of dental classes. It has strong potential to streamline clinical workflows, improve diagnostic accuracy, and enhance patient outcomes.

Keywords: attention mechanism, classification, cross-stage feature attention, mean average Precision (mAP), YOLOv8.

1 Introduction

The classification of dental anatomy is essential in oral healthcare [1], significantly influencing diagnosis [2], treatment planning, and forensic analysis. The precise identification and categorization of dental features are crucial for optimal clinical decision-making. Historically, dental practitioners have depended on the manual assessment of radiographs [3] and clinical pictures, a procedure that is labor-intensive and prone to inter-observer variability. The emergence of artificial intelligence (AI) and deep learning [4] has transformed medical imaging. These technologies provide automated, efficient, and accurate analytical tools.

In recent years, deep learning models, especially

Submitted: 13 September 2025
Accepted: 16 September 2025
Published: 08 November 2025

Vol. 1, **No.** 1, 2025.
10.62762/FBSP.2025.431958

***Corresponding author:**
✉ Pavan Sandula
pavannit4@gmail.com

Citation
Chilaka, V. K., Sandula, P., & Thati, J. (2025). Dental Classes Classification using TEYOLOv8 Network. *Frontiers in Biomedical Signal Processing*, 1(1), 37–48.

convolutional neural networks (CNNs) [5], have exhibited exceptional efficacy in many medical imaging applications, including dental diagnostics. The You Only Look Once (YOLO) series has become prominent for real-time object recognition[6]. YOLOv8 features architectural improvements that enhance detection precision and velocity. Dental images pose distinct obstacles, including diverse tooth shape, overlapping features, and poor contrast, which might impede the efficacy of conventional detection methods.

Attention mechanisms have been incorporated into deep learning models to enable the network to concentrate on salient features within an image, thereby addressing these challenges. Attention modules, including the Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation (SE) blocks [7], have demonstrated their effectiveness in improving feature representation by accentuating informative regions and suppressing irrelevant ones. The integration of attention mechanisms into YOLOv8 has the potential to enhance its performance in intricate dental imaging scenarios.

Recent research has investigated the utilization of attention-enhanced YOLO models in dental diagnosis. A study by [8] presented YOLOv8-AM, which incorporates multiple attention modules into the YOLOv8 framework for the identification of pediatric wrist fractures, resulting in enhanced mean Average Precision (mAP) scores. In [9] introduced a multiclass teeth segmentation model that integrates a Teeth Attention Block (TAB) into a Swin Transformer architecture, yielding enhanced segmentation accuracy. These developments highlight the capability of attention mechanisms to improve the efficacy of deep learning models [10] in medical imaging.

Building upon these developments, this study proposes a novel approach for dental anatomy classification by integrating attention mechanisms into the YOLOv8 framework. The proposed model aims to leverage the real-time detection capabilities of YOLOv8 and the feature enhancement properties of attention modules to accurately classify dental structures in radiographic images [11]. By focusing on critical regions and suppressing background noise, the attention-enhanced YOLOv8 model is expected to achieve higher classification accuracy, particularly in complex imaging conditions.

This study introduces an innovative method for classifying dental anatomy by incorporating attention mechanisms within the YOLOv8 framework, building on recent advancements in the field. The proposed model seeks to utilize the real-time detection capabilities of YOLOv8 alongside the feature enhancement properties of attention modules to effectively classify dental structures in radiographic images. By concentrating on key areas and minimizing background interference, the attention-enhanced YOLOv8 model aims to attain improved classification accuracy, especially in intricate imaging scenarios.

Incorporating attention mechanisms into the YOLOv8 architecture requires adjustments to the network, specifically by adding attention modules at key locations, such as between the backbone and neck or within the detection head. This modification enables the model to dynamically adjust the importance of feature maps, highlighting significant anatomical structures while reducing the impact of extraneous factors. This approach's effectiveness will be assessed through a carefully selected dataset of dental radiographs [12], which have been annotated for different anatomical structures. Performance metrics, including precision and recall, will evaluate the model's classification abilities.

In conclusion, the integration of YOLOv8's strong detection framework with attention mechanisms offers a compelling path forward for enhancing the classification of dental anatomy. The expected results of this study encompass enhanced diagnostic precision, diminished dependence on manual analysis, and the possibility for immediate implementation in healthcare environments. With the ongoing advancements in dental imaging, these innovations are set to significantly improve patient care and treatment results.

2 Motivation and Problem Formulation

The classification of dental anatomy from general dental images such as pictures, intraoral scans, or radiographs is a fundamental problem in digital dentistry. It supports several applications, such as orthodontic planning [13], automated charting [14], educational tools, and AI-assisted diagnosis [15]. Nonetheless, obstacles such as inconsistent lighting conditions, anatomical overlaps, occlusions, and inter-patient variability impede precise classification. Traditional image processing and shallow machine learning techniques encounter difficulties with such complexity and exhibit insufficient stability across varied datasets.

Recent deep learning models, especially object identification frameworks such as YOLOv8, provide substantial advantages in speed and accuracy. However, when utilized for intricate oral images, ordinary YOLOv8 may misidentify or neglect subtle anatomical components owing to its constrained capacity to preferentially concentrate on fine-grained elements. This constraint drives the incorporation of visual attention techniques into the YOLOv8 design to dynamically highlight salient regions (e.g., cusp tips, grooves, or crown shapes) and mitigate background noise. The problem is formulated as follows:

- Given a set of dental images $X = \{x_1, x_2, ..x_n\}$ containing various tooth types and positions,
- The goal is to train a deep model $f_\theta(x)$ augmented with attention modules
- To classify and localize anatomical structures $Y = \{y_1, y_2, ..y_n\}$ exhibiting elevated precision and broad applicability across many image categories and demographics.

The suggested resolution includes

- Cross-stage attention-enhanced layers (e.g., C2fAttn, ImagePoolingAttn) within the backbone and neck of YOLOv8,
- A multi-scale feature fusion mechanism to retain both global and local dental attributes.
- An enhanced detecting head tailored for precise anatomical classification.

This attention-based YOLOv8 approach seeks to enhance performance on generic dental anatomy datasets by delivering contextually aware, feature-selective, and real-time predictions.

3 Proposed Methodology

In this section, we introduce our innovative method for the classification of general dental anatomy radiograph images, which were collected from Kaggle. Our approach is based on an enhanced YOLOv8 architecture that is incorporated with attention mechanisms, specifically the C2fAttn module that has been extended with spatial attention. Our goal is to create a detection and classification model that is both lightweight and highly accurate. This model will be capable of identifying subtle morphological variations in dental anatomy that are frequently overlooked by conventional models.

3.1 Proposed Model Overview

The fundamental framework of YOLOv8, a recent evolution in the YOLO family, is the foundation of the proposed architecture, as shown in Figure 1. This framework is renowned for its efficient one-stage detection. We improve this backbone by incorporating Cross-Stage Partial Fusion (C2f) blocks that are supplemented with a spatial attention mechanism (referred to as C2fAttn-SA). These modules are incorporated into the model at various phases to enhance interdental class discrimination, emphasize salient regions (e.g., cusps, root curvatures, and enamel patterns), and refine contextual features.

The complete pipeline comprises the following:

- Backbone with Spatial-Aware C2fAttn Blocks
- Neck Layer with Multi-scale Feature Aggregation
- Pooling Attention Module for Contextual Enhancement
- YOLOv8 Detection Head Modified for Classification Tasks

The Proposed spatial attention mechanism is shown in Figure 2 which depicts the transformer network with the following criteria below

3.2 C2fAttn-SA Module: Spatial Awareness and Cross-Stage Feature Attention

The C2f block, which was introduced in YOLOv8, maintains representational power while reducing computational burden by partially merging outputs from multiple convolutional stages. The C2fAttn-SA module is formed by incorporating spatial attention to extend this block.

3.2.1 Mathematical Modeling:

Let $X \in \mathbb{R}^{C\times H\times W}$ be the input feature map to the C2f block, where C, H, and W represent the number of channels, height, and width, respectively. The C2fAttn-SA module executes the subsequent phases:

Channel-wise Grouping & Convolution:

- Divide X into k groups: $\{X_1, X_2, ..., X_k\}$
- Apply the convolution f_θ to each group:

$$Y_i = f_\theta(X_i), \quad i = 1, 2, ..., k$$

- Concatenation is employed to merge the data:

$$Y = \text{Concat}(Y_1, Y_2, ..., Y_k)$$

Generation of Spatial Attention Maps:

- Average and maximum aggregation across the channel dimension generate spatial attention (SA):

$$M_{\text{SA}} = \sigma(f^{7\times7}([\text{AvgPool}(Y); \text{MaxPool}(Y)]))$$

where σ is the sigmoid activation and $f^{7\times7}$ is a convolution with a 7×7 kernel.

Refinement: The output that is enhanced for attention is

$$Y' = Y \cdot M_{\text{SA}}$$

Residual Connection: The input is multiplied by the output to ensure stability.

$$Z = Y' + X$$

This module enables the network to accentuate spatially significant features, including molar grooves, crown ridges, and root canal curvature.

3.2.2 *Pooling Attention for Global Context*

The global structural context is essential in dental classification tasks due to the high inter-class similarities, and we implement an ImagePooling Attention (IPA) mechanism to incorporate it. This module implements a global self-attention map and aggregates multi-scale features from various backbone levels. Figures 1 and 2 shown the flow model of proposed network with spatial attention inclusion.

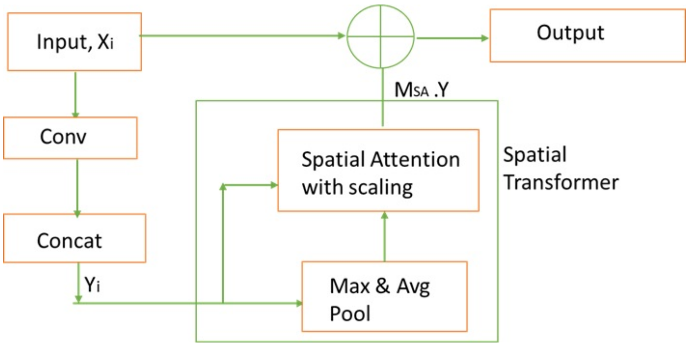

Figure 1. Proposed TEYOLOv8 architecture

Mathematical Formulation: Using the features from three layers F_1, F_2, F_3, we operate as follows:

- Global average pooling:

$$G_i = \text{GAP}(F_i), \quad i = 1, 2, 3$$

- Softmax fusion and MLP projection:

$$A = \text{softmax}(W[G_1, G_2, G_3])$$

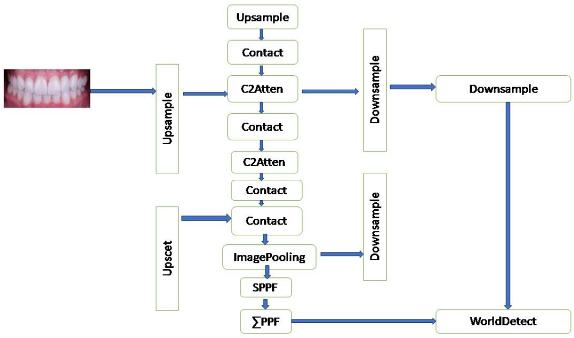

Figure 2. Proposed Spatial Attention Transformer

- Contextual refinement:

$$F' = \sum_i A_i \cdot F_i$$

The principal path is concatenated with this context vector to direct the classifier head.

3.2.3 *Dental Classification Neck and Head Design*

Neck In order to consolidate high-level and low-level features, we implement a modified Feature Pyramid Network (FPN—top-down approach) + Path Aggregation Network (PAN—bottom-up approach) structure in the neck. Spatially precise superficial features are combined with semantically rich features from deeper layers through the use of concatenation and upsampling operations.

Modification of the Detection Head We modify the detection head to facilitate dense classification across dental categories (incisors, canines, premolars, molars, etc.), despite the fact that YOLOv8 is inherently a detection architecture.

The feature vector input to the head should be $F \in \mathbb{R}^{C\times H\times W}$. Flattening it to $F' \in \mathbb{R}^{C\times(H\cdot W)}$ the application of a linear classifier and global average pooling:

$$\hat{y} = \text{Softmax}(W_c \cdot \text{GAP}(F') + b_c)$$

where W_c and b_c are learnable parameters, and $\hat{y} \in \mathbb{R}^K$ is the number of dental classes.

3.2.4 *Loss Functions*

A combination of classification and auxiliary attention consistency losses is implemented:

i Classification Loss:
Categorical Cross Entropy

$$\mathcal{L}_{cls} = -\sum_{i=1}^{K} y_i \log(\hat{y}_i)$$

ii Attention Consistency Loss:
Guarantees that attention maps are consistent across layers.

$$\mathcal{L}_{attn} = \sum_{l=1}^{L} \|M_{\text{SA}}^{(l)} - M_{\text{SA}}^{(l-1)}\|^2$$

iii Total loss:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{attn}$$

4 Experimental Results and Discussion

The experimentation is carried out in Google Colab L4GPU on the Dental Anatomy Classes dataset. The dataset consists of images of dental teeth, each with different resolutions. The data constraint with resolution is to resize its original dimension to 640x640, where the label information will correlate. The label information includes fixed tooth positions and specifies the corresponding dental class. The dataset is split into 80% for training, 10% for validation, and 10% for testing. Both subjective and objective results have been analyzed.

4.1 Subjective Analysis

The Transformer-embedded YOLOv8 model will be used to identify the position of a dental image that has been provided. The 7 dental classes will be classified based on their characteristic features through localization and segmentation as they undertake training on various images. The output of a molar classification model applied to real-world dental images is shown in Figure 3, which is the validated image set. Each tooth is identified and labeled with its class, such as 1st Molar, 2nd Premolar, Central Incisor, etc., as well as a confidence score that indicates the model's confidence in the prediction. The detected tooth regions are represented by the bounding boxes in a variety of colors, and the majority of predictions exhibit high confidence levels (ranging from 0.7 to 0.99), which underscores the model's exceptional accuracy in localization and classification. The model consistently functions across a wide range of dental orientations and lighting conditions and effectively differentiates between similar classes. The visual results illustrate a generalizable, high-performing model that is capable of accurately identifying various tooth types in diverse validation samples, despite the presence of some overlapping boxes and misclassifications (particularly for background noise or neighboring tooth types).

4.2 Objective Analysis

The objective analysis is carried out using different loss metrics, namely precision, recall, mAP-0.5, and

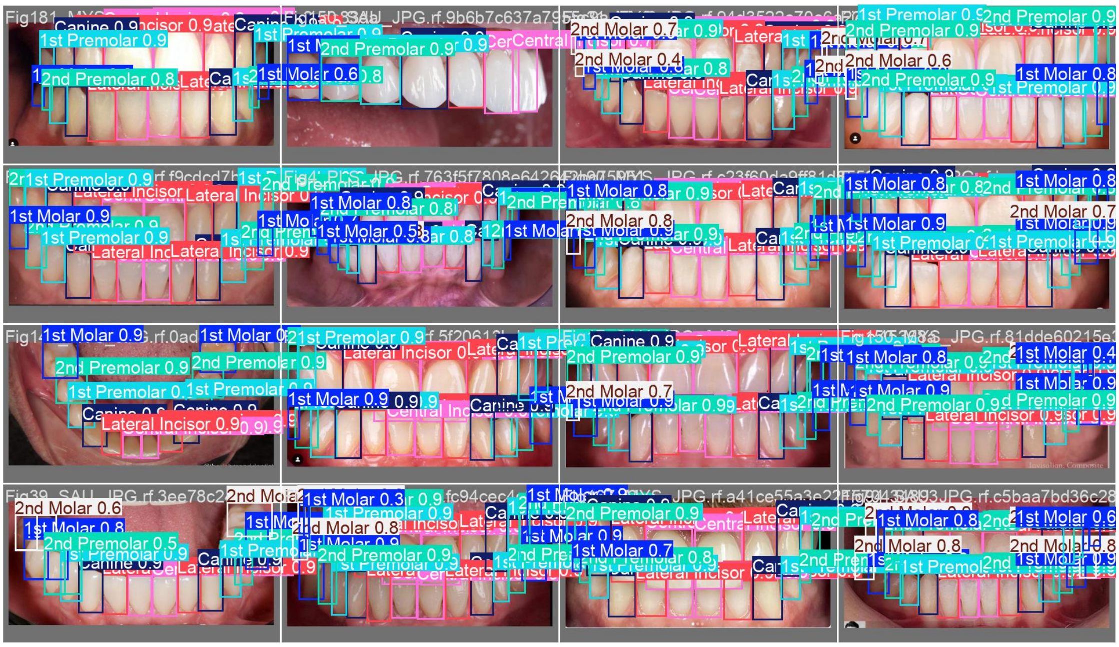

Figure 3. Demonstrating output for different tooth categories.

Table 1. Comparison of different Performance metrics.

S.No	Class	Images	Instances	Precision (P)	Recall (R)	mAP@0.5	mAP@0.5:0.95
1	All	112	2437	0.954	0.973	0.981	0.794
2	1st Molar	106	323	0.906	0.95	0.969	0.697
3	1st Premolar	111	386	0.982	0.995	0.99	0.805
4	2nd Molar	74	157	0.882	0.908	0.946	0.637
5	2nd Premolar	109	364	0.945	0.989	0.985	0.747
6	Canine	112	401	0.979	0.99	0.991	0.873
7	Central Incisor	112	403	0.995	0.993	0.995	0.909
8	Lateral Incisor	112	403	0.985	0.985	0.993	0.89

Table 2. Comparison of different stateof art methods m->min, s->seconds.

Method	Time(Total)	Images	Instances	Precision (P)	Recall (R)	mAP@0.5	mAP@0.5:0.95
YOLOv8 [16]	33s	112	2437	0.816	0.857	0.897	0.678
YOLOv11 [17]	25m 16s	112	2437	0.822	0.799	0.84	0.657
Proposed Method	1m 7s	112	2437	0.879	0.885	0.95	0.729

mAP-0.5-0.9. These loss metrics are investigated with changes in epochs. It means to say all these metrics are plotted between loss vs. epochs. The confusion matrix and its normalized evaluation are also computed with respect to 7 different dental classes. Not only losses, but also accuracy estimation is computed with an ablation study. The provided image visualizes various statistical aspects of a dental image dataset that appears to be annotated for object detection or classification tasks involving different types of teeth.

The provided Table 1 appears to present the evaluation results for an object detection model, specifically targeting the identification of various types of teeth (e.g., molars, premolars, incisors) in dental images.

The comparative analysis is also provided in Table 2 where the YOLO11 has not produced a benchmark within a time bound and cannot proceed to real-time processing. However, our proposed method reaches an outstanding result with only 50 epochs and in a short duration. YOLO11 may reach the best, but it had a higher computational cost leveraged with 10 epochs itself. In order to bring uniformity, all the results shown in Table 2 observed at 10 epochs.

The primary metrics employed to assess the model's performance for each class (tooth type) are:

1. Class: This denotes the various categories or types of teeth (e.g., 1st molar, 2nd premolar, central incisor, etc.).
2. Images: The quantity of images in the dataset associated with each class.
3. Instances: The quantity of distinct objects (teeth) that the model must identify in the images for each category.
4. Precision (P): This metric quantifies the proportion of detected objects that are true positives. The ratio of true positive detections to the total number of positive detections, encompassing both true and false positives. Increased precision results in a reduction of false positives. For instance, the 1st Molar exhibits a precision of 0.906, indicating that 90.6% of the model's detections for this category are accurate.
5. Recall (R): This metric quantifies the number of actual objects in the dataset that were accurately identified by the model. The ratio of true positives to the total number of ground-truth objects is defined as the proportion of correctly identified instances relative to the actual objects present. Increased recall results in a reduction of false negatives. For instance, the 1st Molar exhibits a recall of 0.95, indicating that 95% of the actual 1st molars in the dataset were accurately identified

by the model.

6. mAP@0.5: This metric represents the mean Average Precision calculated at an Intersection over Union (IoU) threshold of 0.5. This metric indicates the average precision across all classes, defining a detection as correct when the Intersection over Union (IoU) with the ground-truth object exceeds 0.5. An increased mAP indicates superior overall detection performance. For instance, the central incisor exhibits an mAP@0.5 of 0.995, signifying strong performance by the model in this category.

7. mAP@0.5:0.95: This metric represents a more rigorous assessment of mean Average Precision, as it averages precision across various Intersection over Union (IoU) thresholds ranging from 0.5 to 0.95. This metric provides a comprehensive evaluation of the model's performance, especially in scenarios where accurate localization is critical. A greater value signifies a model that demonstrates strong performance across various localization thresholds. For instance, the central incisor exhibits a high mAP@0.5:0.95 of 0.909, indicating that the model effectively detects and accurately localizes it across multiple thresholds.

Figure 4 shows four different representations of dental anatomy with its following characteristics.

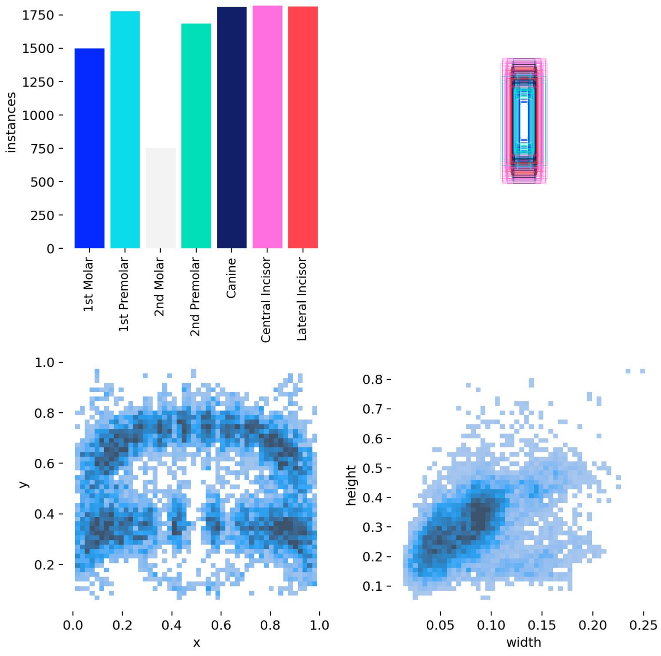

Figure 4. Data representation with its heatmap.

Top-Left The Top-Left is a bar chart that shows the number of instances for different tooth categories on the x-axis, namely 1st molar (blue), 1st premolar, 2nd molar (fewer instances than others), 2nd premolar, canine, central incisor, and lateral incisor. The y-axis represents the number of bounding box annotations for each category.

Top-Right On the Top-Right Bounding Box Overlay (Shape Distribution) This subplot shows a composite image of all bounding boxes likely drawn on a normalized grid. It represents the shape, orientation, and distribution of bounding boxes for all classes combined. The overlapping colored rectangles indicate the common region where most teeth are located, showing their general alignment and size range.

Bottom-Left On the Bottom-Left, the 2D histogram heatmap (X vs Y center positions) is generated with the X-axis as the normalized horizontal center coordinate of bounding boxes. Whereas with the Y-axis as the normalized vertical center coordinate. The color mapping indicates frequency (more instances = darker), which signifies that the heatmap forms a dental arch shape, representing how teeth are distributed across a panoramic X-ray. Most bounding boxes are centered in a curved region—matching typical dental structure.

Bottom-Right On the Bottom-Right, the 2D histogram heatmap (width vs height) represents X-axis as bounding box width (normalized) and the Y-axis as bounding box height (normalized). Most bounding boxes are within a typical width range of 0.05 to 0.15 and height range of 0.2 to 0.5. This shows a cluster of common tooth sizes, useful for model anchor box generation in object detection.

The Precision-Confidence Curve illustrated in the diagram represents the correlation between a dental classification model's precision (the accuracy of positive predictions) and its confidence (the certainty in predictions) across various tooth classes, including molars, premolars, canines, and incisors, as shown in Figure 5. The Y-axis denotes precision, ranging from 0.0 to 1.0, whereas the X-axis indicates confidence thresholds, also spanning from 0.0 to 1.0. Each curve represents a distinct tooth class, demonstrating that precision enhances with increased confidence—higher confidence thresholds result in fewer yet more accurate predictions.

The designation "all classes 1.00 at 0.993" signifies that the model attains perfect precision (100%) across all

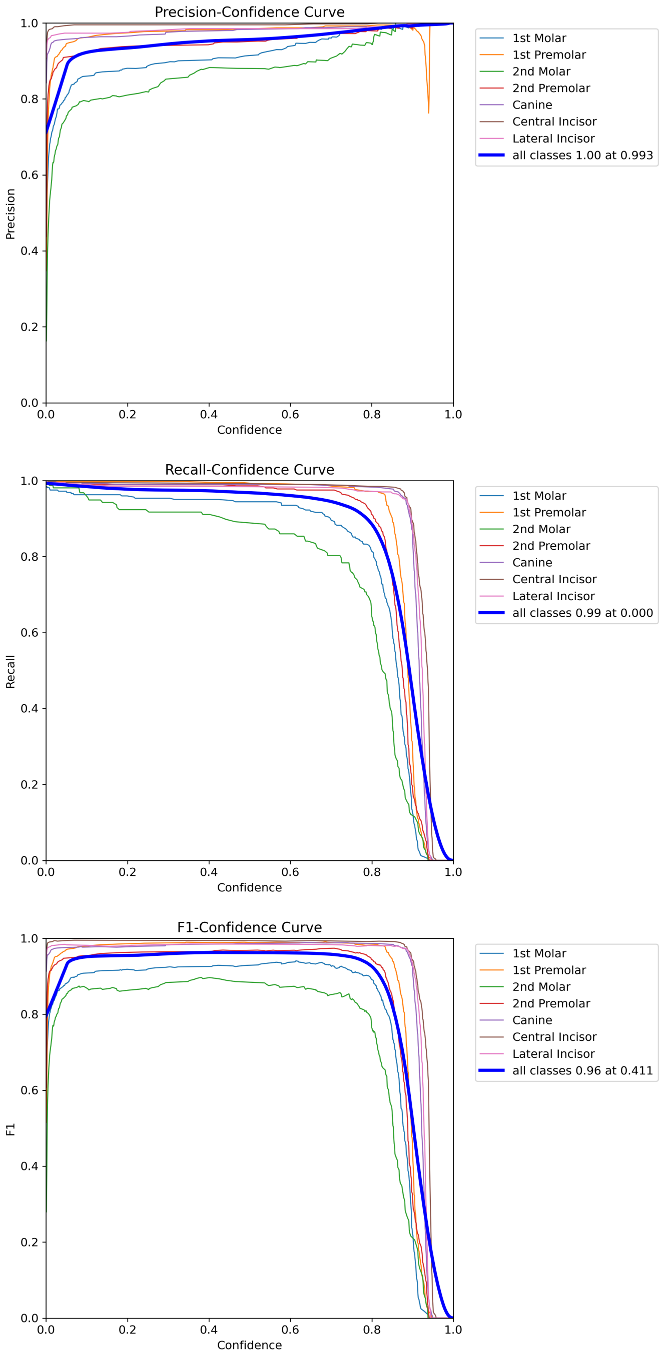

Figure 5. Precision-Recall Confidence curves.

classes when the confidence threshold is elevated to 0.993, although this may result in diminished recall. The curves indicate variations in performance among classes, with certain classes sustaining high precision at lower confidence levels, whereas others necessitate greater certainty for dependable predictions. This visualization assists in determining an optimal confidence threshold to balance precision and recall, tailored to the specific requirements of the application, such as emphasizing accuracy in medical diagnostics.

The recall-confidence curve demonstrates the variation in a dental classification model's recall, which reflects its capacity to identify all pertinent cases, as the confidence threshold, indicating the model's prediction certainty, fluctuates across various tooth classes such as molars, premolars, and incisors, as depicted in Figure 5. The Y-axis indicates recall values ranging from 0.0 to 1.0, whereas the X-axis denotes confidence thresholds also spanning from 0.0 to 1.0. At low confidence thresholds (approximately 0.0), recall is elevated (around 1.0) due to the model generating a greater number of predictions, thereby identifying most true positives; however, this frequently results in an increase in false positives. As confidence rises, recall diminishes due to the model's increased selectivity, which may result in the omission of certain true positives. The curves indicate performance variations by class: certain teeth, such as the central incisor, exhibit stable recall across thresholds, whereas others, like the 2nd molar, experience a significant decline. The statement "all classes 0.99 at 0.000" signifies that at a confidence threshold of 0.0, the model attains 99% recall across all classes, although this is likely associated with reduced precision. This curve facilitates the equilibrium between recall and precision; lower thresholds emphasize the detection of all positives, which is essential in medical screening, whereas higher thresholds enhance precision by minimizing false positives. The graph facilitates the selection of optimal thresholds in accordance with diagnostic priorities.

Figure 5 presents the F1 score where X-axis is Confidence threshold — the minimum confidence score necessary for a model to validate a prediction and Y-axisis F1 score — a balanced metric that integrates precision and recall. Each line denotes a distinct tooth class (e.g., 1st Molar, 2nd Premolar) illustrating the progression of its F1 score with increasing threshold values. The bold blue line indicates the mean F1 score for all classes. Confidence Threshold Range: 0.0 to 1.0 At low thresholds (e.g., 0.1–0.4), the model incorporates numerous detections, including false positives, potentially diminishing precision while enhancing recall. At elevated thresholds (e.g., 0.9–1.0), the model adopts a more stringent approach, dismissing uncertain predictions. This typically enhances precision while diminishing recall, which may adversely affect the F1 score. Optimal Confidence: The highlighted blue annotation on the curve indicates that the highest overall F1 score (0.96) is achieved at a confidence threshold of 0.411. Setting the model to filter out predictions with confidence below 0.411 optimally balances precision and recall across all classes. By class Behavior: The majority of curves exhibit a peak followed by a prolonged plateau, subsequently declining sharply within the range of confidence = 0.8–0.9. The 2nd Molar (green line) exhibits inferior performance relative to other classes, consistently demonstrating lower F1 scores across all thresholds. This aligns with your previous confusion matrix, which indicated a higher frequency of misclassification for molars. The 1st premolar, canine, and central incisor exhibit high performance, with F1 scores approaching 1.0 across a broad confidence range, signifying reliable detection. The lateral incisor and first molar perform adequately, although there are minor declines observed at elevated thresholds.

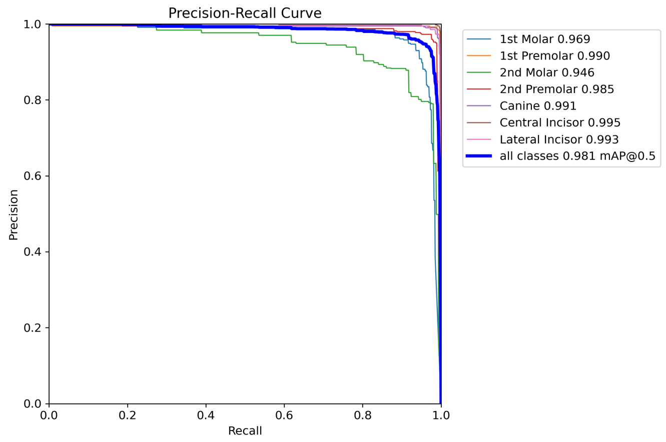

Figure 6. Precision and Recall Curves.

Figure 6 presents a Precision-Recall (PR) Curve utilized for assessing the performance of an object detection model, particularly in the identification of various types of teeth in panoramic radiographs. The x-axis denotes recall, quantifying the number of actual positives accurately identified, whereas the y-axis denotes precision, reflecting the proportion of predicted positives that are correct. Each curve in the plot represents a distinct tooth class, with the average precision (AP) values adjacent to each class name reflecting the model's performance for that particular category. The AP values are high across all classes, with the central incisor (0.995) and lateral incisor (0.993) attaining the highest scores, followed by the canine (0.991) and first premolar (0.990). The lowest AP is recorded for the second molar (0.946), yet it still reflects strong performance. The thick blue line indicates the overall mean Average Precision (mAP) across all classes at an IoU threshold of 0.5, recorded at 0.981, which is a commendable result. The proximity of the curves to the top-right corner of the plot indicates that the model attains high precision alongside elevated recall levels, signifying a reduction in false positives and false negatives.

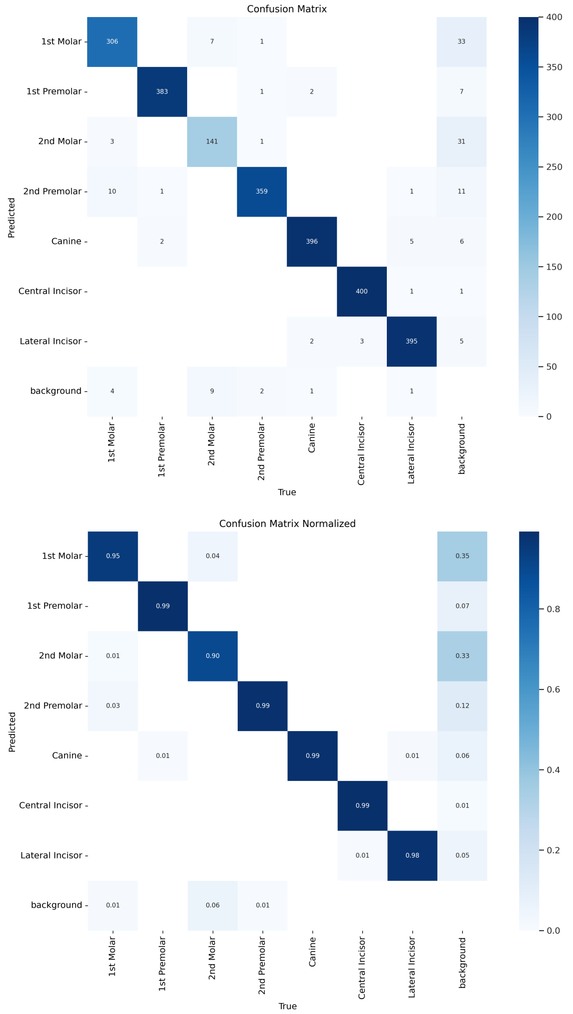

Figure 7. Standard and Normalized confusion matrix summarizes the results of predictions in a classification.

The confusion matrix is a tool used to evaluate the performance of a classification model by summarizing the correct and incorrect predictions made by the model. It provides insights into the types of errors made and the overall accuracy of the model. Figure 7 illustrates the performance of the multi-class tooth classification model, demonstrating high accuracy across the majority of tooth categories. The model demonstrates proficiency in identifying central incisors, canines, and lateral incisors, achieving 400, 396, and 395 correct predictions, respectively, which reflects a low rate of misclassification. The 2nd premolar demonstrates favorable outcomes with 359 accurate predictions, although there is some minor confusion with the 1st molar and background.

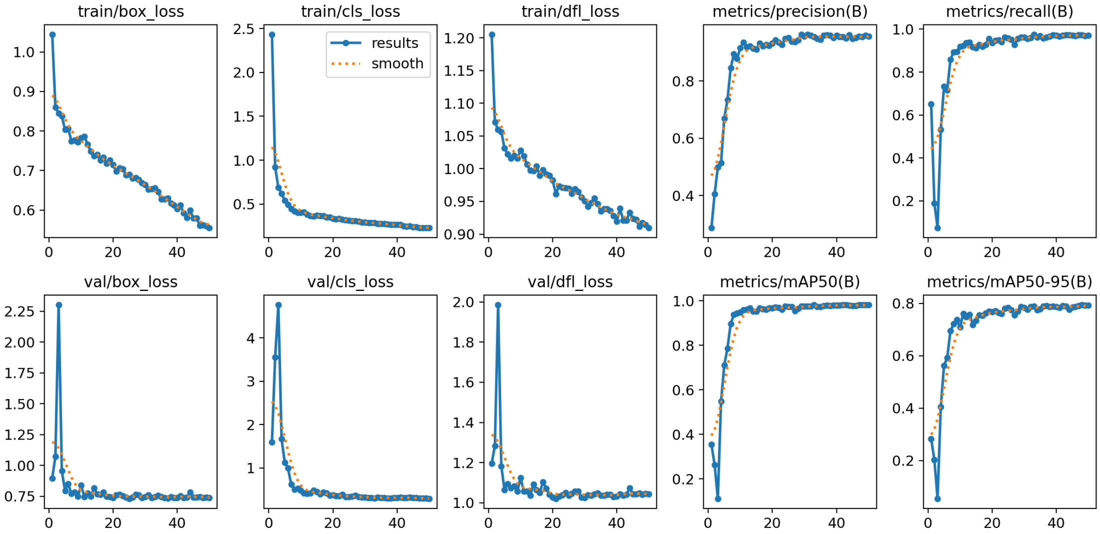

Figure 8. Training and validation curves of the proposed model.

Moderate misclassifications are noted for the 1st and 2nd molars, especially concerning the background class, with 33 instances of the 1st molar misclassified as background and 31 instances for the 2nd molar. Moreover, several background samples were inaccurately classified as dental structures. Despite areas for improvement, especially in differentiating molar types and minimizing background confusion, the model exhibits reliable classification performance and effectively distinguishes most tooth types. Improvements should prioritize the enhancement of feature distinctions and the more effective management of class imbalances.

Figure 7 also presents the normalized confusion matrix. A confusion matrix summarizes the results of predictions in a classification problem. Each row of the matrix corresponds to the instances predicted to belong to a specific class, whereas each column indicates the actual class. Standardized: The values are normalized to a range of 0 to 1, ensuring that the sum of each row equals 1. This facilitates the interpretation of class-wise performance, independent of class imbalance. The diagonal values indicate the proportion of accurate predictions for each class. These should ideally approach 1.00 (i.e., 100%). For instance: The first molar is accurately predicted 95% of the time. The 1st premolar, 2nd premolar, canine, and central incisor exhibit approximately 99% accuracy. The performance of the lateral incisor is slightly lower at 98%. The prediction accuracy for the 2nd molar is 90%. These indicate the locations of errors, specifically where a true class is misidentified as a different predicted class. Significant misclassifications: First Molar → Background (35%): The elevated value may suggest that the model occasionally fails to identify first molars, potentially due to factors such as occlusion, inadequate contrast, or dataset imbalance. Second Molar → Background (33%): A high error rate indicates a potential confusion between molars and background elements. The second premolar constitutes 12% of the background. The observed errors indicate a higher likelihood of misclassification of molars and premolars, particularly those located posteriorly, as background. This misclassification may be attributed to inadequate lighting or contrast in the X-ray images. Structures that overlap. Reduced sample sizes for those classes during training.

Figure 8 presents the training and validation curves of the proposed model over 50 epochs. This document presents an analysis of the train curves (top row) and their implications.

- train/box_loss: This curve illustrates the progression of bounding box regression loss, demonstrating a consistent decline from above 1.0 to approximately 0.55. This indicates that the model is improving its accuracy in localizing objects (teeth) over time.
- train/cls_loss: The classification loss initiates at approximately 2.5 and subsequently decreases to below 0.3, demonstrating the model's efficacy in learning to classify various tooth types, such as molars, canines, and incisors.
- train/dfl_loss: This refers to the Distribution Focal Loss, which is associated with accurate bounding box localization. The curve exhibits

a consistent decline, indicating improved localization precision.

- Metric/precision(B): The precision metric, which measures the correctness of predicted objects, increases rapidly and stabilizes around 0.95. This indicates that the model is achieving greater accuracy in its predictions while reducing the occurrence of false positives.
- metrics/recall(B): The recall, which measures the proportion of correctly detected ground truth objects, exhibits a significant increase and subsequently stabilizes around 0.95. This indicates that the model effectively identifies the majority of actual objects while minimizing false negatives.

The training curves indicate convergence and robust performance, characterized by a steady decline in losses and a gradual increase in performance metrics, specifically precision and recall. This suggests that the model is acquiring knowledge efficiently from the training dataset.

5 Conclusion

In this paper, a deep learning architecture designed for the classification of dental teeth utilizing panoramic raw data, with a focus on pediatric dental analysis. The model integrated advanced components, including C2F blocks, C2F attention mechanisms, and SPPF modules, into the transformer-embedded YOLOv8 network, which significantly improved feature extraction and multi-scale representation. Data augmentation techniques were utilized to enhance generalization and robustness under diverse radiograph conditions. The proposed method exhibited significant classification accuracy and detection performance across various tooth categories. The model attained an overall precision of 0.954, a recall of 0.973, a mean Average Precision (mAP) at 0.5 of 0.981, and a mAP at 0.5:0.95 of 0.794. Specific tooth classes, notably the Central Incisor, Canine, and Lateral Incisor, demonstrated enhanced detection metrics, thereby validating the efficacy of the architecture.

Clinical Significance:

The proposed method of approach provides accurate and precise information in classifying the teeth, which provides predominantly an oral representation of the pediatric dentistry. It also inculcates the real streaming analysis for practitioners to improve with better patient treatment, and connect to remote advice analysis.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] MUHAMAD, D. A. H. (2016). Dental and Oral Health. *Sciences, 14*(4), 124-130.

[2] Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., ... & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The lancet digital health, 1*(6), e271-e297. [CrossRef]

[3] Rozylo-Kalinowska, I. (2020). Introduction to Dental Radiography and Radiology. In *Imaging Techniques in Dental Radiology: Acquisition, Anatomic Analysis and Interpretation of Radiographic Images* (pp. 1-5). Cham: Springer International Publishing. [CrossRef]

[4] Tsuneki, M. (2022). Deep learning models in medical image analysis. *Journal of Oral Biosciences, 64*(3), 312-320. [CrossRef]

[5] Schwendicke, F., Golla, T., Dreher, M., & Krois, J. (2019). Convolutional neural networks for dental image diagnostics: A scoping review. *Journal of dentistry, 91*, 103226. [CrossRef]

[6] Fang, W., Wang, L., & Ren, P. (2019). Tinier-YOLO: A real-time object detection method for constrained environments. *Ieee Access, 8*, 1935-1944. [CrossRef]

[7] Ganguly, A., & Ruby, A. U. (2023). Evaluating CNN architectures using attention mechanisms: Convolutional Block Attention Module, Squeeze, and Excitation for image classification on CIFAR10 dataset.

[8] Chien, C. T., Ju, R. Y., Chou, K. Y., Xieerke, E., & Chiang, J. S. (2025). YOLOv8-AM: YOLOv8 Based on Effective Attention Mechanisms for Pediatric Wrist Fracture Detection. *IEEE Access*. [CrossRef]

[9] Ghafoor, A., Moon, S. Y., & Lee, B. (2023). Multiclass Segmentation Using Teeth Attention Modules for Dental X-Ray Images. *IEEE Access, 11*, 123891-123903. [CrossRef]

[10] Razaghi, M., Komleh, H. E., Dehghani, F., & Shahidi, Z. (2024, March). Innovative diagnosis of dental diseases using YOLO V8 deep learning model. In

2024 13th Iranian/3rd International Machine Vision and Image Processing Conference (MVIP) (pp. 1-5). IEEE. [CrossRef]

[11] Sukegawa, S., Yoshii, K., Hara, T., Tanaka, F., Yamashita, K., Kagaya, T., ... & Furuki, Y. (2022). Is attention branch network effective in classifying dental implants from panoramic radiograph images by deep learning?. *PLoS One, 17*(7), e0269016. [CrossRef]

[12] George, J., Hemanth, T. S., Raju, J., Mattapallil, J. G., & Naveen, N. (2023, August). Dental radiography analysis and diagnosis using YOLOv8. In *2023 9th International Conference on Smart Computing and Communications (ICSCC)* (pp. 102-107). IEEE. [CrossRef]

[13] Rischen, R. J., Breuning, K. H., Bronkhorst, E. M., & Kuijpers-Jagtman, A. M. (2013). Records needed for orthodontic diagnosis and treatment planning: a systematic review. *PLoS one, 8*(11), e74186. [CrossRef]

[14] Davila, K., Setlur, S., Doermann, D., Kota, B. U., & Govindaraju, V. (2020). Chart mining: A survey of methods for automated chart analysis. *IEEE transactions on pattern analysis and machine intelligence, 43*(11), 3799-3819. [CrossRef]

[15] Kong, M., He, Q., & Li, L. (2018). AI assisted clinical diagnosis & treatment, and development strategy. *Strategic Study of Chinese Academy of Engineering, 20*(2), 86-91.

[16] Wang, H., Diao, K., Wu, L., Li, X., Zhou, X., & Liu, X. (2024, December). An Automatic Tooth Position and Dental Disease Detection Algorithm Based on YOLOv8. In *2024 10th International Conference on Computer and Communications (ICCC)* (pp. 2058-2062). IEEE. [CrossRef]

[17] Akdoğan, S., Öziç, M. Ü., & Tassoker, M. (2025). Development of an AI-Supported clinical tool for assessing mandibular third molar tooth extraction difficulty using panoramic radiographs and YOLO11 Sub-Models. *Diagnostics, 15*(4), 462. [CrossRef]

Chilaka Victor Kumar was born in Guntur, Andhra Pradesh, India, in 1990. After finishing his schooling in 2006, he received B.Tech. degree in Electronics and Communication Engineering from P.N.C & VIJAI Institute of Engineering & Technology (JNTU Kakinada) in 2018. (victorych401@gmail.com)

Pavan Sandula was born in Srikakulam, Andhra Pradesh, India, in 1990. After he completed under graduation in 2007, received B.Tech Electronics and Communication Engineering at UCE-VZM in 2012 and the M.Tech degree in Signal and Image Processing at NIT Rourkela. Later, He joined as JRF in Nov., 2016 with Ph.D. admission at NIT Rourkela and with awarded Doctorate on Compressed Video Zoom Motion Analysis and Saliency Estimation in ECE stream Dec., 2021. His research interests include camera motion analysis, Pattern recognition, Image Segmentation and Machine Learning. (Email: pavannit4@gmail.com)

Jagadeesh Thati was born in Guntur, Andhra Pradesh, India, in 1986. After finishing his schooling in 2003, he received B.Tech. degree in Electronics and Communication Engineering from SCR Engineering College (JNTU Kakinada) in 2007, and the M.Tech. degree in Signal Processing from JNTU College of Engineering, Hyderabad (JNTUH), in 2009. He has been an Associate Professor at Tirumala Engineering College, Narasaraopet, since 2011. He received the Ph.D. degree in Electronics and Communication Engineering from the National Institute of Technology, Rourkela. His research interests include remote sensing, multispectral image processing, change detection, image processing techniques, and machine learning. (Email: jagadeeshthati@gmail.com)