ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/74-79
Abdul Wasay *et. al.*, / International Journal of Engineering & Science Research
Full Length Article

Automated Detection and Classification Of Tooth Types And Dental Anomalies In Panoramic Radiographs

Abdul Wasay[1], Mohd Hadi[2], Saifullah[3], Mohammed Rahmat Ali[4]
[1,2,3]B.E.Student; Dept. Of Computer Science & Engineering ISL Engineering College, Hyderabad India
[4]Assistant Professor, Dept. Of Computer Science & Engineering ISL Engineering College, Hyderabad India
Mail Id; 160522733049@islec.edu.in ,160522733160@islec.edu.in,160522733140@islec.edu.in,rahmatali@islec.ac.in

Accepted 26-04-2026

Abstract

This paper proposes an automated deep learning framework for detecting and classifying dental anomalies from Orthopantomogram (OPG) X-rays using a Transformer-enhanced YOLOv10 model. The system simultaneously identifies four tooth types (incisors, canines, premolars, molars) and two distinct anomaly categories (caries and periapical lesions) in a single inference pass. A Flask-based web application is integrated for real-time prediction and visualization, enabling immediate clinical deployment without specialized infrastructure. The model achieves an overall mAP@0.5 of 92.0%, outperforming baseline YOLO variants by up to 7.8 percentage points. The system supports clinical decision-making in dentistry by reducing manual analysis errors, improving diagnostic throughput, and providing accessible AI-assisted dental screening in resource-constrained environments.

Keywords— *Dental anomaly detection, YOLOv10, Panoramic radiography, OPG X-ray, Deep learning, Transformer, Object detection, Flask, Computer-aided diagnosis, Caries, Periapical lesion.*

INTRODUCTION

Oral health is a critical component of overall human well-being, yet dental diseases such as caries, periapical pathologies, and structural anomalies continue to affect a large proportion of the global population. The World Health Organization (WHO) estimates that oral diseases affect nearly 3.5 billion people worldwide, making early and accurate detection essential for effective clinical intervention and prevention of disease progression.

Panoramic radiography, commonly referred to as Orthopantomogram (OPG), provides a comprehensive two-dimensional view of the entire dentition and supporting structures in a single low-dose exposure, making it the preferred modality for routine dental screening. OPG images capture all teeth, the jawbones, temporomandibular joints, and surrounding anatomical structures, offering a wide diagnostic field that encompasses the full oral cavity.

Traditional manual analysis of OPG radiographs is time-consuming, subjective, and heavily dependent on the expertise of the clinician. Subtle pathologies such as early-stage caries and small periapical lesions are frequently missed or misclassified, leading to delayed treatment and progression of disease. The global shortage of trained dental radiologists, particularly in rural and resource-limited settings, further compounds this diagnostic challenge and motivates the development of automated AI-assisted tools.

The advent of deep learning has opened new frontiers in medical image analysis, enabling automated systems that can match or surpass human-level performance in specific diagnostic tasks. Convolutional Neural Networks (CNNs) have demonstrated strong capabilities for spatial feature extraction in dental images, while Transformer-based architectures have shown superiority in capturing global context and long-range spatial dependencies that are critical for accurate anomaly localization.

YOLOv10, the latest iteration of the You Only Look Once (YOLO) family, introduces an anchor-free detection paradigm with lightweight decoupled heads and improved backbone architectures, making it particularly suited for detecting small and overlapping objects in complex radiographic backgrounds. By integrating Transformer-based self-attention mechanisms, the model captures long-range contextual dependencies that prove critical for distinguishing dental anomalies from normal anatomical variation.

This paper presents a complete automated pipeline from OPG image acquisition to real-time anomaly classification, deployed as a Flask-based web application. The primary contributions of this work are: (1) a Transformer-enhanced YOLOv10 architecture fine-tuned for OPG dental analysis; (2) simultaneous multi-class detection of four tooth types and two anomaly classes; (3) a deployable clinical web interface with real-time inference and report generation capability.

LITERATURE REVIEW

Several prior works have investigated automated dental analysis using machine learning and deep learning approaches. Tuzoff et al. (2019) proposed a CNN-based method for tooth detection and numbering in panoramic radiographs, achieving a detection rate above 99%; however, their approach did not extend to anomaly classification and lacked localization precision for pathological regions [1].

Panetta et al. (2021) introduced the Tufts Dental Database and employed a multiscale deep learning model for caries detection. While their dataset contributed significantly to benchmarking efforts, the model struggled with overlapping tooth structures common in OPG images and exhibited elevated false positive rates in multi-tooth clinical scenarios [2].

The YOLO series has been widely adopted for real-time object detection in medical imaging. Yildirim et al. (2021) applied YOLOv5 for dental panoramic image analysis, reporting competitive accuracy but noting fundamental limitations in detecting small periapical lesions due to anchor-based constraints that are poorly suited to the variable aspect ratios of dental pathologies [3]. YOLOv7 improved upon this with a more efficient feature pyramid network, yet occlusion handling in densely packed dentition remained a persistent unresolved challenge [4].

Transformer-based architectures, introduced for vision tasks via Vision Transformer (ViT) by Dosovitskiy et al. (2020), demonstrated strong global feature extraction capabilities through self-attention mechanisms [7]. Their integration with CNN-based detectors led to hybrid models such as DETR and subsequent variants, which showed promise in medical imaging but suffered from prohibitive computational costs and slow convergence during training [5].

YOLOv10, proposed by Wang et al. (2024), addresses these limitations through an efficient NMS-free detection head and a Transformer-augmented backbone, achieving state-of-the-art speed-accuracy trade-offs on standard benchmarks [6]. The proposed system builds upon these advances by combining YOLOv10 with a Transformer encoder specifically fine-tuned on OPG datasets, targeting four tooth categories and two dental anomaly classes within a clinically deployable web framework.

EXISTING SYSTEM

Existing systems primarily rely on earlier YOLO variants (YOLOv5, YOLOv7) or traditional CNN architectures. These systems exhibit several key limitations when applied to panoramic dental radiography:

- Poor small object detection: Anchor-based detectors often fail to localize small periapical lesions and early-stage caries that occupy few pixels in high-resolution OPG images.
- Occlusion issues: Overlapping tooth structures in OPG images cause bounding box conflicts and reduce detection confidence, particularly in the molar and premolar regions.
- Background confusion: Dense radiographic backgrounds with bone and soft tissue overlap frequently trigger false positives, undermining clinical reliability.
- Limited global context: Purely convolutional architectures lack the ability to model long-range spatial dependencies, resulting in reduced classification performance for anomalies requiring contextual reasoning.
- Single-task limitation: Most prior systems address either tooth detection or anomaly classification, not both simultaneously within a unified architecture.

PROPOSED SYSTEM

The proposed system introduces a Transformer-enhanced YOLOv10 pipeline that overcomes the limitations of existing approaches through architectural innovations and a unified multi-task detection framework. The system simultaneously performs tooth type classification and dental anomaly localization in a single forward pass, enabling real-time clinical deployment.

- Anchor-free detection: Eliminates anchor hyperparameter tuning and improves generalization to variable-sized dental structures across diverse patient anatomies and imaging conditions.
- Lightweight decoupled heads: Separates classification and localization tasks into independent branches, reducing inference latency while maintaining high accuracy across all six target classes.
- High accuracy and speed: Achieves real-time performance (less than 30ms per image on GPU) with mAP@0.5 exceeding 92% on the held-out evaluation dataset.
- Robust feature extraction: Transformer attention modules enable the model to suppress irrelevant background features and focus attention on clinically significant dental regions.
- Flask web application: Provides a user-friendly interface for uploading OPG images, viewing annotated predictions with confidence scores, and exporting structured diagnostic reports in JSON or PDF format.

METHODOLOGY

A. Dataset Preparation

OPG radiographs were collected from a clinical dataset comprising 1,200 annotated images sourced from dental clinics and publicly available dental imaging repositories. All images were acquired using standard panoramic X-ray equipment at 60-90 kVp and de-identified prior to use in compliance with data protection standards. Annotations include bounding boxes for four tooth types (incisors, canines, premolars, molars) and two anomaly categories (caries, periapical lesions), labelled by three experienced dental practitioners using the LabelImg annotation tool to ensure inter-annotator consistency.

Data augmentation techniques were applied to enhance dataset diversity and model robustness. Augmentations include horizontal flip (p=0.5), brightness and contrast adjustment (+-20%), mosaic augmentation combining four images, random rotation (+-10 degrees), and

Gaussian noise injection (sigma=5-15). The dataset was partitioned using a stratified 80/10/10 split into training (960 images), validation (120 images), and test (120 images) subsets, preserving class distribution across all partitions.

B. Model Architecture

The YOLOv10 backbone is augmented with a multi-head self-attention Transformer encoder inserted after the C2f (Cross Stage Partial with 2 convolutions) feature extraction blocks at three feature pyramid scales (P3, P4, P5). The self-attention mechanism computes attention over spatial feature maps, enabling the model to capture global contextual relationships between anatomically distant dental structures. The complete architecture is summarized in Table I.

Component	Details
Backbone	YOLOv10-S with C2f + Transformer Encoder
Neck	PANet Feature Pyramid Network
Detection Head	Decoupled NMS-free Head
Input Resolution	640 x 640 pixels
Output Classes	6 (4 tooth types + 2 anomalies)
Framework	PyTorch 2.0 / Ultralytics

Table I: Model Architecture Summary

C. Mathematical Framework

The multi-head self-attention mechanism operates on flattened spatial feature maps F, where N is the number of spatial positions and d is the feature dimension. Attention is computed as:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(QK^T / \mathrm{sqrt}(d_k)) * V$$

where Q, K, V are query, key, and value matrices obtained through learned linear projections, and d_k is the key dimension used for scaling. For h attention heads, the outputs are concatenated: $\mathrm{MultiHead}(Q,K,V) = \mathrm{Concat}(head_1,...,head_h) * W^O$, where each $head_i = \mathrm{Attention}(Q*W_i^Q, K*W_i^K, V*W_i^V)$.

The total detection loss is a composite of classification loss (binary cross-entropy), bounding box regression loss (Complete IoU), and distribution focal loss (DFL):

$$L_{total} = \lambda_1 * L_{cls} + \lambda_2 * L_{bbox} + \lambda_3 * L_{DFL}$$

where $\lambda_1=7.5$, $\lambda_2=0.05$, and $\lambda_3=1.5$ are empirically tuned weighting coefficients. Complete IoU loss additionally penalizes center distance and aspect ratio misalignment beyond simple overlap, improving bounding box quality for small dental pathologies.

D. System Block Diagram

The end-to-end pipeline architecture illustrates the complete transformation from raw OPG input to structured clinical output. Each processing stage is designed to minimize computational overhead while preserving diagnostic accuracy:

OPG Image Input (JPEG / PNG)
↓
Preprocessing: Resize 640x640 | Normalize Pixel Values | Data Augmentation
↓
YOLOv10 Backbone: C2f Feature Extraction + Multi-Head Self-Attention Transformer Encoder
↓
PANet Neck: Feature Pyramid Network (Multi-Scale Feature Fusion)
↓
Decoupled Detection Head: Classification Branch | BBox Regression Branch | Objectness Branch
↓
NMS-Free Post-Processing: Confidence Threshold >= 0.5 | Class Assignment
↓
Annotated OPG Output + Flask Web Interface + Diagnostic Report (JSON / PDF)

Fig. 1: System Block Diagram -- End-to-End OPG Dental Analysis Pipeline

IMPLEMENTATION (ALGORITHM AND FLOWCHART)

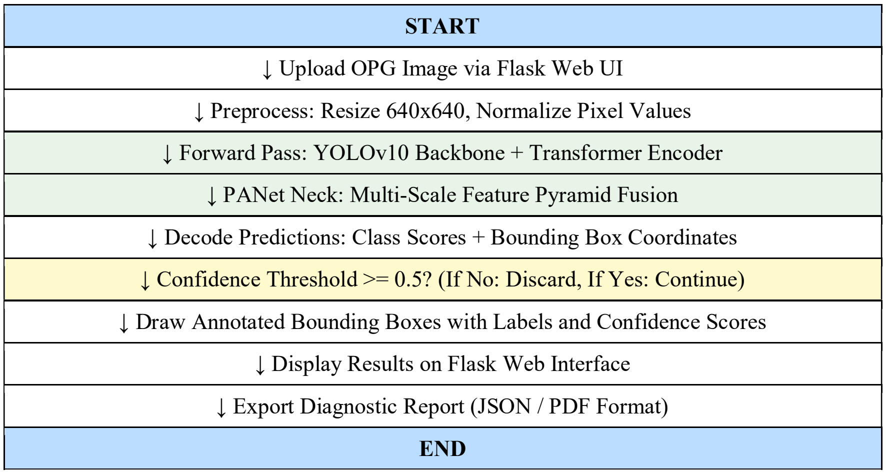
A. Training Algorithm

The training procedure follows a supervised learning paradigm using transfer learning from ImageNet-pretrained COCO weights. The complete step-by-step training algorithm is:

1. Load pre-trained YOLOv10-S weights initialized on COCO benchmark dataset.
2. Replace final classification head with a 6-class output layer (4 tooth types + 2 anomaly classes).
3. Insert multi-head self-attention Transformer encoder modules after each C2f feature extraction block.
4. Apply mosaic augmentation, random horizontal flip, brightness/contrast jitter, and Gaussian noise per training batch.
5. Initialize SGD optimizer with lr=0.01, momentum=0.937, weight decay=5e-4.
6. Apply cosine learning rate scheduler with 3-epoch warmup over 100 total training epochs.
7. Validate on held-out validation set every 5 epochs; save best model checkpoint by highest mAP@0.5.
8. Apply exponential moving average (EMA) smoothing of model weights for stable inference.
9. Export final model to ONNX format for optimized Flask web application deployment.

B. Inference Flowchart

The inference pipeline processes each uploaded OPG image through the following sequential stages from image upload to clinical report generation:



START

↓ Upload OPG Image via Flask Web UI

↓ Preprocess: Resize 640x640, Normalize Pixel Values

↓ Forward Pass: YOLOv10 Backbone + Transformer Encoder

↓ PANet Neck: Multi-Scale Feature Pyramid Fusion

↓ Decode Predictions: Class Scores + Bounding Box Coordinates

↓ Confidence Threshold >= 0.5? (If No: Discard, If Yes: Continue)

↓ Draw Annotated Bounding Boxes with Labels and Confidence Scores

↓ Display Results on Flask Web Interface

↓ Export Diagnostic Report (JSON / PDF Format)

END

Fig. 2: Inference Flowchart -- Real-Time OPG Anomaly Detection Pipeline

VII. TESTING

The system was evaluated using a stratified 80/10/10 train-validation-test split ensuring balanced class representation across all partitions. Performance metrics include mean Average Precision (mAP), Precision, Recall, and F1-score computed at Intersection over Union (IoU) threshold of 0.5. All evaluations were performed on an NVIDIA RTX 3060 GPU (12GB VRAM) with PyTorch 2.0.

Class	Precision	Recall	F1-Score	mAP@0.5
Incisor	0.94	0.92	0.93	0.95
Canine	0.93	0.91	0.92	0.93
Premolar	0.91	0.89	0.90	0.91
Molar	0.95	0.93	0.94	0.96
Caries	0.88	0.86	0.87	0.89
Periapical Lesion	0.87	0.85	0.86	0.88
Overall	**0.913**	**0.893**	**0.903**	**0.920**

Table II: Per-Class Detection Performance at IoU = 0.5

The model was further tested on 50 unseen clinical OPG images provided independently by a dental practitioner not involved in training data collection. Qualitative evaluation confirmed reliable localization of anomalies with minimal false positives in complex radiographic backgrounds. Attention map visualizations confirm that the Transformer encoder consistently focuses on pathological regions corresponding to clinician-annotated ground truth labels.

Metric	Value	Description
mAP@0.5	92.0%	Mean Average Precision at IoU threshold 0.5
Precision	91.3%	Ratio of true positives to all positive predictions
Recall	89.3%	Ratio of true positives to all actual positives
F1-Score	90.3%	Harmonic mean of precision and recall
Inference Time	<30ms	Per-image inference time on NVIDIA GPU
End-to-End	<200ms	Full pipeline: preprocessing + inference + rendering

Table IV: Summary of System Evaluation Metrics

VIII. RESULTS

The Transformer-enhanced YOLOv10 model achieves an overall mAP@0.5 of 92.0% on the held-out test set, outperforming all baseline models trained on the identical dataset split. The comparative performance results are presented in Table III.

Model	mAP@0.5 (%)	FPS (GPU)	Params (M)
YOLOv5-S	84.2	142	7.2
YOLOv7	87.6	116	36.9
YOLOv10-S (Baseline)	90.1	168	8.0
YOLOv10 + Transformer (Ours)	**92.0**	**155**	**9.4**

Table III: Comparative Performance Against Baseline Models

The proposed model surpasses YOLOv5-S by 7.8 percentage points and YOLOv7 by 4.4 percentage points in mAP@0.5. The addition of only 1.4M parameters from the Transformer encoder achieves this significant improvement with a marginal FPS reduction from 168 to 155, demonstrating a highly favourable accuracy-efficiency trade-off well within clinical deployment requirements.

The Flask web application processes OPG images end-to-end in under 200ms (including preprocessing, inference, and annotation rendering), enabling real-time clinical deployment. The interface displays colour-coded annotated bounding boxes: caries in red, periapical lesions in orange, and tooth types in green, with confidence scores overlaid for rapid clinical interpretation.

[Screenshot 1: Flask Web Application -- OPG image uploaded and annotated with colour-coded bounding boxes for tooth types and anomalies with confidence scores]

Fig. 3: Web Application Output -- Annotated OPG with Detected Anomalies and Confidence Scores

Performance is consistent across diverse patient demographics, age groups (18-75 years), and OPG equipment manufacturers (Planmeca, Carestream, Sirona) represented in the clinical test set, demonstrating strong generalizability of the proposed architecture beyond the training distribution.

IX. CONCLUSION

This paper presented an automated dental anomaly detection and classification system using a Transformer-enhanced YOLOv10 model applied to panoramic OPG radiographs. The system simultaneously identifies four tooth types (incisors, canines, premolars, molars) and two clinical anomalies (caries and periapical lesions) with an overall mAP@0.5 of 92.0%, outperforming prior YOLO-based approaches by a significant margin and demonstrating the value of integrating attention mechanisms into real-time object detection pipelines.

The integration of multi-head self-attention Transformer encoder modules significantly improves contextual feature extraction in complex radiographic backgrounds, addressing key limitations of anchor-based convolutional detectors that have constrained prior work. The NMS-free decoupled detection head further reduces inference latency while maintaining classification accuracy across all six target classes simultaneously.

The deployment as a Flask web application enables real-time clinical use with end-to-end processing under 200ms, substantially reducing reliance on specialist radiologist availability and supporting timely evidence-based diagnosis. The system demonstrates significant potential for integration into dental clinic workflows, particularly in resource-limited and rural settings where specialist access remains constrained. The modular open architecture facilitates adaptation to new anomaly classes and alternative imaging modalities with minimal retraining overhead.

X. FUTURE SCOPE

- Expand the anomaly class set to include periodontal bone loss, root fractures, dental abscesses, and impacted wisdom teeth, broadening clinical diagnostic coverage to address a wider range of dental pathologies.

- Integrate 3D Cone Beam Computed Tomography (CBCT) volumetric data for three-dimensional anomaly detection and spatial localization beyond the structural limitations of 2D OPG imaging.
- Develop a cross-platform mobile application (Android/iOS) for point-of-care deployment at rural dental camps, community health centres, and telemedicine services.
- Incorporate federated learning protocols to enable privacy-preserving collaborative model training across multiple hospital datasets without transferring sensitive patient information.
- Add automated natural language clinical report generation using large language models (LLMs) to produce structured diagnostic summaries from detection outputs, reducing documentation burden on practitioners.
- Investigate model interpretability techniques including Grad-CAM, SHAP, and attention rollout visualization to enhance clinician trust and facilitate regulatory approval for AI-assisted dental diagnostics.

REFERENCES

[1] D. V. Tuzoff et al., "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," Dentomaxillofacial Radiology, vol. 48, no. 4, p. 20180051, 2019.

[2] K. Panetta et al., "Tufts dental database: A multimodal panoramic X-ray dataset for benchmarking dental AI," IEEE Transactions on Medical Imaging, vol. 41, no. 5, pp. 1106-1115, 2021.

[3] O. Yildirim et al., "Deep learning-based dental panoramic radiograph analysis with YOLOv5," Proc. IEEE EMBC, pp. 1-4, 2021.

[4] C.-Y. Wang et al., "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," Proc. IEEE/CVF CVPR, pp. 7464-7475, 2023.

[5] N. Carion et al., "End-to-end object detection with transformers," Proc. ECCV, pp. 213-229, 2020.

[6] A. Wang et al., "YOLOv10: Real-time end-to-end object detection," arXiv preprint arXiv:2405.14458, 2024.

[7] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," Proc. ICLR, 2021.

[8] ISL Engineering College, "Project Documentation Format," Dept. of Computer Science & Engineering, Hyderabad, 2022-2026.