



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: <https://doi.org/10.22214/ijraset.2026.85144>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Dual-Model Deep Learning Framework for Tooth Segmentation and Multi-Condition Detection in Panoramic Radiographs

Ashray Chaitanya¹, Balaji M², Akshath Rai³, Achyutha Nayaka DN⁴, Dr. Roopa H⁵, Dr. Devi Ganesan⁶

^{1, 2, 3, 4}Student, Dept of ISE Bangalore Institute of Technology, Bengaluru, India

⁵Professor, Dept of ISE Bangalore Institute of Technology, Bengaluru, India

⁶Oral D Research Bengaluru, India

Abstract: Orthopantomogram (OPG) radiographs are one of the main pillars of dental diagnosis, yet their interpretation is subjective, time consuming, and expensive depending on the radiologist. We suggest a deep learning structure to solve this by adding U-Net-based tooth segmentation and YOLOv8-based object detection to analyze the dental abnormalities in an automated manner.

The segmentation model deals with anatomical localization and gives estimates of the regions of focus, and the detection model deals with clinically significant findings like cavities, impacted teeth, fillings and implants. The U-Net has 97.23% accuracy on validation data in the segmentation of teeth and the YOLOv8 model has a total mAP at 0.5 of 78.7% at 640×640 , with the difference between individual classes representing imbalances in our limited dataset. One of the design considerations was clinical usability via continuous learning approach. The framework has an interactive course-correction interface where clinicians can manipulate segmentation masks and refine detection bounding boxes. An error-correction loop is built on top of these corrections, where both models are refined at each annotated case, as the number of annotated cases increases.

The final output is an automatically generated PDF that displays patient details along with the dental analysis summary. The dental anomaly and the count of each type is displayed. Confidence scores for prediction can also be displayed depending on the user's choice. This end-to-end pipeline is practically feasible and cost effective while being useful in real-time for clinicians.

Index Terms: Dental image analysis, orthopantomogram (OPG), medical image segmentation, object detection, U-Net, YOLOv8, deep learning, active learning, clinical decision support.

I. INTRODUCTION

Dental diseases are among the most common global health concerns, affecting a substantial portion of the population worldwide. Orthopantomogram panoramic radiographs are extensively used in dental practice because they provide a comprehensive two-dimensional view of the maxillofacial region and surrounding structures. Common pathologies observable in OPG images include dental caries (cavities), impacted teeth, dental restorations (fillings), and implants requiring accurate assessment. Precise identification of these conditions is essential for normal treatments, surgeries, and preventive care.

However, manual interpretation of these OPG images has several challenges. Radiographs often contain structures of the overlapping teeth, gums, unclear boundaries, and metallic distortions caused by dental restorations. The interpretation depends heavily on clinician expertise while also being more expensive and time consuming.

Deep learning has proven to be successful in medical image analysis tasks, particularly in segmentation and object detection. The U-Net architecture [1] has become a standard framework for medical image segmentation and masking due to its trademark ability to preserve spatial resolution and prevent loss of features via the encoder-decoder structure with skip connections. Similarly, real-time object detection models such as YOLO are quick but need improvement in finding area of focus for accurate detection. YOLOv8 when paired with U-Net, provides improved speed and increased area of focus which makes it suitable for practical deployment.

In dental imaging research, segmentation and detection tasks are often looked at independently. Some approaches apply object detection directly to panoramic images without explicit anatomical localization. They do not have an area of focus for the model to learn, which increases sensitivity to background structures and unwanted features. Other studies focus solely on tooth segmentation without integrated pathology classification.

In this work, we propose a dual-model deep learning framework for dental OPG analysis that integrates U-Net-based tooth segmentation with YOLOv8-based pathology detection. Both models operate on the same input radiograph, providing complementary outputs: segmentation enables anatomical localization (area of focus) and uncertainty estimation, while YOLO identifies anomalies including cavities, impacted teeth, fillings, and implants. In addition, this framework adopts a continuous learning approach for the whole pipeline. An interactive correction interface allows doctors to refine segmentation masks and detection bounding boxes. A threshold-based retraining pipeline enables fine-tuning of both models when sufficient corrected samples are accumulated. The system also generates structured PDF reports summarizing AI findings, confidence metrics, and clinical recommendations, supporting practical clinical workflows.

The primary contributions of this work are as follows:

- 1) Two-stage framework integrating U-Net-based anatomical segmentation with YOLOv8-based multi-pathology classification.
- 2) Correction-driven continuous learning mechanism that enables iterative model improvement using clinician-provided annotations in a real-world facility.
- 3) A prediction of error/uncertainty that will help clinicians to help the model in relearning difficult cases.
- 4) Threshold-based incremental fine-tuning with controlled updates, incorporating Early Stopping and adaptive learning rate scheduling for stable convergence.
- 5) End-to-end workflow of pre-processing, segmentation, detection, post-processing (connected component analysis), and automated clinical report generation.

The proposed system demonstrates that deployable, correction-aware AI solutions can be used in real-time to reduce cost and time, while also being more accurate in the detection of dental anomalies.

II. LITERATURE REVIEW AND RELATED WORK

In dental imaging, U-Net and its variants (like U-Net++) have been successfully applied for tooth segmentation and structural analysis tasks [6], [7]. But there are no frameworks that include a two step process for anomaly classification as the end goal.

Object detection in medical imaging has been advanced by models such as YOLO [2], which perform real-time multi-object localization. YOLOv8 has improved accuracy and computational efficiency, making them suitable for clinical applications. In dental history, independent YOLO-based approaches have been used to detect cavities, implants, and impacted teeth, although performance is often affected by class imbalance in training dataset and low-contrast features [6], [7]. One of the reasons it suffers in accuracy is because of the lack of area of focus. Although segmentation and detection are typically approached independently, very limited work integrates both tasks within a unified framework. Existing approaches lack mechanisms for incorporating user feedback or enabling model refinement in real-world clinical settings. A human-in-the-loop procedure would solve the hard cases that only humans are capable of solving. Active learning strategies have been explored to improve model performance by prioritizing uncertain or difficult samples for retraining [8]. However, fully deployable systems that combine segmentation, detection, and correction-driven continuous learning remain limited in current dental imaging applications.

III. METHODOLOGY

A. Overall Pipeline Architecture

The proposed system is designed as a two-stage deep learning framework for automated dental OPG analysis. A single Orthopantomogram (OPG) image is processed by two neural networks: a U-Net-based segmentation model and a YOLOv8-based object detection model. The input image is preprocessed and passed through the U-Net model to generate a segmentation mask. This masked output is the input for the YOLO model. The YOLO model gets the region of focus from the masked regions of the OPG. Then it detects anomalies based on a classification with probability values. Let the input radiograph be represented as:

$$I \in \mathbb{R}^{H \times W} \quad (1)$$

Two inference functions are defined:

$$M = f_{seg}(I) \quad (2)$$

$$D = f_{det}(I) \quad (3)$$

where f_{seg} produces a binary segmentation mask M , and f_{det} produces a set of detected bounding boxes

$$D = \{(b_i, c_i, p_i)\} \quad (4)$$

with bounding box coordinates b , class labels c , and confidence scores p_i .

The masked output represents anatomical tooth regions, while the detection output represents pathological classification including cavity, impacted tooth, fillings, and implants. The outputs are then combined for visualization and clinical interpretation. This separation enhances interpretability and modularity, allowing each model to be independently retrained or updated.

B. Dataset Preparation

The segmentation dataset consists of approximately 1,200–1,500 Orthopantomogram (OPG) radiographs with corresponding binary masks, derived from the Teeth Segmentation on Dental X-ray Images dataset [9]. The data was split into training, validation, and test sets (70:15:15), with images resized to 512×512 and augmented using standard techniques.

The detection dataset comprises 780 annotated OPG images with bounding boxes for four pathological classes—implants (20.4%), fillings (69.2%), impacted teeth (4.9%), and cavities (5.5%), based on dataset [10] and adapted to YOLO format. A similar 70:15:15 stratified split was used, with images resized to 640×640 . The dataset exhibits class imbalance, particularly for cavity detection.

1) Stage 1: U-Net Segmentation Model

The segmentation component is based on the U-Net architecture, which performs pixel-wise classification to identify tooth-bearing regions in panoramic OPG radiographs. Input images are resized to 512×512 pixels and converted to grayscale prior to training and inference.

The model follows an encoder–decoder structure with skip connections, enabling effective extraction of both low-level and high-level features while preserving spatial information. The encoder captures contextual representations through convolution and pooling operations, while the decoder restores spatial resolution and prevents loss of data for precise segmentation.

A sigmoid activation function is applied at the output layer to generate probability maps, which are run through thresholds to obtain binary segmentation masks. The model is trained using a combination of Binary Cross-Entropy and Dice loss [4] to improve both pixel-wise accuracy and region overlap. Training is performed using the Adam optimizer [3]. During correction-driven refinement, transfer learning is applied with a reduced learning rate to adapt the model using newly corrected samples while preserving previously learned features.

2) Stage 2: YOLOv8 Object Detection Model

The pathology detection component employs YOLOv8, a single-stage object detection architecture for identifying dental abnormalities in OPG radiographs. The model processes input images resized to 640×640 pixels and predicts bounding boxes with associated class labels and confidence scores.

The detection task includes four classes: cavities, impacted teeth, fillings, and implants. YOLOv8 enables efficient multi-object localization and classification in a single forward pass, making it suitable for real-time clinical applications.

The model is trained using standard YOLO loss functions that jointly optimize localization, objectness, and classification. Performance is evaluated using precision, recall, and mean average precision (mAP) [5].

The trained model achieved an overall precision of 0.739 and recall of 0.773, with mAP@0.5 of 0.787 and mAP@[0.5:0.95] of 0.535. Performance varies across classes due to dataset imbalance, with cavity detection being the most challenging.

C. Two-Stage Pipeline Integration and Continuous Learning

The proposed framework integrates segmentation and detection within a unified dual-model architecture in which both models process the same original OPG radiograph independently. Let the segmentation output be defined as

$$M = f_{seg}(I) \quad (5)$$

while the detection network, implemented using YOLOv8, simultaneously predicts pathological findings

$$D = f_{det}(I) \quad (6)$$

where

$$D = \{(b_i, c_i, p_i)\} \quad (7)$$

represents the set of predicted bounding boxes b_i , class labels c_i , and confidence scores p_i .

The segmentation output M provides anatomical localization of tooth-bearing regions, whereas the detection output D identifies pathological conditions.

Post-processing and Structural Analysis: To derive clinically meaningful insights, connected component analysis is applied to the binary segmentation mask:

$$T_{seg} = CC(M) \quad (8)$$

where $CC(\cdot)$ denotes the connected component extraction function. This enables estimation of individual tooth regions and supports structural quantification. The segmentation masks are processed to extract region of focus, which is used for input to YOLO and for bounding box approximation and visualization.

1) Inference Preprocessing: During deployment, up-loaded radiographs undergo a standardized preprocessing pipeline prior to inference. This includes:

- Automatic region-based cropping to isolate the primary diagnostic area.
- Resizing to fixed input dimensions (512×512 for segmentation, 640×640 for detection).
- Intensity normalization to ensure consistent input distribution.

This preprocessing ensures robustness against variations in acquisition conditions and improves inference stability.

2) Correction-Driven Continuous Learning: To enable progressive system improvement, the framework incorporates a correction-driven refinement mechanism grounded in transfer learning principles. Clinicians may manually edit predicted segmentation masks or adjust detection bounding boxes through the graphical interface. Corrected samples are stored in a structured database and used for subsequent model refinement. Let N_{seg} denote the number of corrected segmentation masks and N_{det} denote the number of corrected detection annotations.

Retraining is triggered when predefined thresholds are satisfied:

$$N_{seg} \geq 20 \quad (9)$$

$$N_{det} \geq 10 \quad (10)$$

3) Hard-Sample Selection Strategy: Instead of uniformly using all corrected samples, the system employs a hard-sample selection strategy. Each prediction is associated with an error or uncertainty score derived from model outputs (e.g., segmentation loss or detection confidence deviation). These error/uncertainty scores are stored alongside predictions and reused during retraining to prioritize informative samples.

During retraining, samples with higher error scores are prioritized:

$$S_{hard} = \{x_i \mid \text{loss}(x_i) > \tau\} \quad (11)$$

where τ is a predefined threshold.

This approach ensures that the model focuses on difficult or misclassified cases, improving learning efficiency and generalization, consistent with active learning principles.

4) Fine-Tuning Configuration: Model refinement is performed using transfer learning on the previously trained model rather than training from scratch. Let the updated model be defined as:

$\text{Model}_{t+1} = \text{FineTune}(\text{Model}_t, \text{Corrected Samples})$ (12) Fine-tuning is performed using the following configuration:

- Learning rate: 1×10^{-4}
- Batch size: 4
- Epochs: 5
- Validation split: 0.2

To ensure stable convergence and prevent overfitting, the following strategies are applied:

- EarlyStopping: monitors validation loss and stops training when no improvement is observed
- ReduceLROnPlateau: dynamically reduces learning rate when validation performance stagnates

These mechanisms improve training efficiency while maintaining model generalization.

- 5) Existing Model Refinement: Unlike conventional re-training approaches, the system does not reinitialize model weights. Instead, it loads the current deployed model and incrementally updates its parameters using corrected samples. This ensures:
 - retention of previously learned knowledge
 - faster convergence
 - reduced computational cost
- 6) Closed-Loop Learning Workflow: The entire system operates as a continuous learning loop, defined as:
 Prediction → User Correction → Data Storage → Selective Retraining → Model Update
 This iterative pipeline enables the model to adapt to real-world clinical variations over time while maintaining controlled updates.
- 7) Summary of Integration: The integration of segmentation, detection, correction, and retraining within a unified framework enables:
 - simultaneous anatomical and pathological analysis
 - adaptive learning from user feedback
 - efficient handling of difficult samples
 - scalable deployment in real-world clinical environments

D. Deployment on Hugging Face Spaces

The complete system is deployed on the Hugging Face Spaces platform, providing a cloud-hosted environment for interactive access. The frontend is implemented using the Gradio framework, enabling users to upload OPG images, visualize segmentation masks and detection bounding boxes with confidence scores, and optionally provide corrections for model refinement.

The backend consists of Python-based services integrating TensorFlow for U-Net segmentation inference and the Ultralytics YOLO framework for object detection. The application dynamically detects available hardware and utilizes GPU acceleration when provisioned, otherwise defaulting to CPU execution. Model weights are stored with version control, ensuring reproducibility and maintainability. Users access the system through a web interface without requiring local installation of dependencies or manual model execution. This deployment strategy facilitates accessibility while maintaining controlled model updates and computational flexibility.

IV. RESULTS AND DISCUSSION

The U-Net segmentation model achieved a peak validation accuracy of 97.23% at Epoch 14, demonstrating strong anatomical localization capability for tooth-bearing regions. Training progression showed rapid convergence, with validation accuracy exceeding 95% by the second epoch. Subsequent epochs exhibited diminishing improvements, indicating stabilization of learned representations. A slight increase in validation loss after Epoch 14 suggested the onset of mild overfitting, confirming the effectiveness of early stopping mechanisms in preserving generalization performance.

Qualitative evaluation demonstrates accurate delineation of tooth structures while excluding surrounding anatomical regions such as the mandible, sinus cavities, and cranial structures. The predicted masks exhibit strong spatial correspondence with ground truth annotations, particularly along occlusal boundaries and inter-dental regions. Minor discrepancies are observed primarily at low-contrast boundaries, which is consistent with inherent radiographic variability.

The high segmentation accuracy supports the suitability of the U-Net architecture for dental anatomical localization tasks. The combined BCE and Dice loss

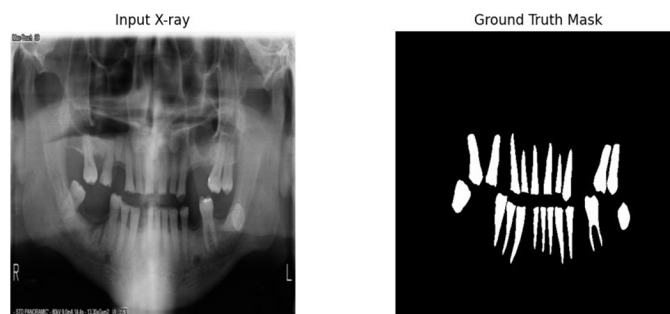


Fig. 1. Example Dental X-ray and Ground Truth Segmentation Mask. A panoramic dental X-ray (left) and its corresponding ground truth segmentation mask (right), used for training tooth segmentation.

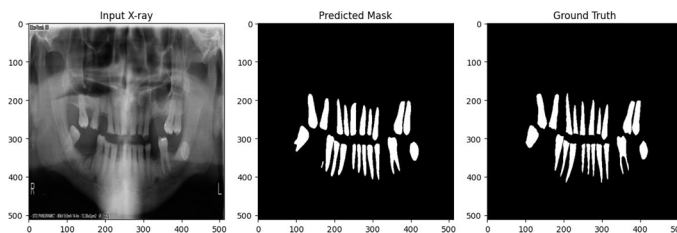


Fig. 2. U-Net Sample Segmentation Output. Input dental X-ray (left), predicted mask by trained U-Net (middle), and ground truth mask (right).

formulation contributed to improved boundary adherence and robustness against class imbalance.

The YOLOv8 detection model achieved 78.7% mAP@0.5, 73.9% precision, 77.3% recall, and 53.5% mAP@[0.5:0.95] on the validation dataset. These results indicate reliable multi-pathology detection performance across the four classes: cavity, impacted tooth, fillings, and implants.

Qualitative results demonstrate accurate localization of impacted teeth within complex panoramic anatomical structures. The model effectively identifies dental restorations and implants with high confidence scores, while cavity detection remains comparatively more challenging due to limited training samples and subtle radiographic contrast.

The achieved mAP@0.5 of 78.7% reflects strong detection capability under moderate IoU thresholds, while the lower mAP@[0.5:0.95] value indicates sensitivity to stricter localization requirements. This performance gap is expected in medical detection tasks involving small or low-contrast lesions.

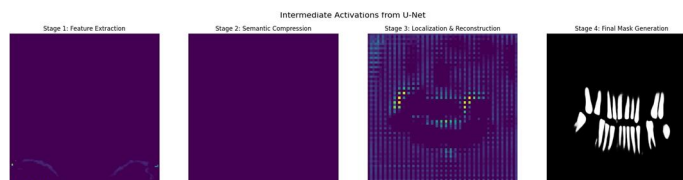


Fig. 3. Intermediate activations from U-Net showing feature extraction, semantic compression, localization & reconstruction, and final mask generation.

Per-class performance trends correlate with dataset distribution, where classes with higher sample counts exhibit stronger detection accuracy. This observation highlights class imbalance as a primary factor influencing detection performance.

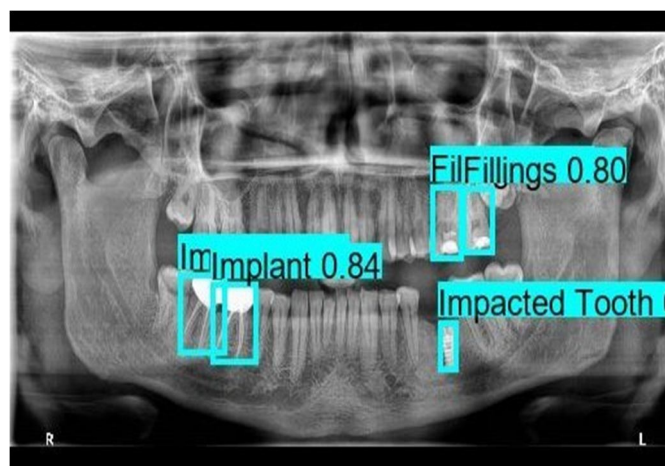


Fig. 4. YOLOv8 Results – Impacted Tooth and Fillings Detection. YOLOv8 detection of impacted teeth and fillings with bounding boxes and probability scores on a panoramic OPG image.

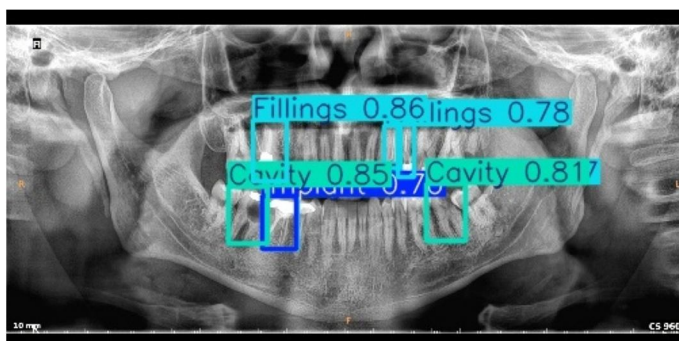


Fig. 5. YOLOv8 Multi-Pathology Detection. YOLOv8 output showing bounding box predictions for cavities, fillings, and implants, with confidence scores overlaid.

TABLE I
Overall and Per-Class Detection Results (YOLOv8)

Metric	Value
Overall Precision	0.739
Overall Recall	0.773
mAP@0.5	0.787
mAP@0.5:0.95	0.535

Class	Precision	Recall	mAP
Implant	0.912	0.950	0.984
Fillings	0.843	0.878	0.903
Impacted Tooth	0.747	0.684	0.795
Cavity	0.453	0.581	0.466

Summary of global and class-wise precision, recall, and mean average precision for the YOLOv8 dental detection.

The YOLOv8 detection results presented in Table I demonstrate strong overall performance, achieving 78.7%

TABLE II
U-Net Segmentation Training Progression

Metric	Start	End (Epoch 15)
Training Accuracy	0.8745	0.9717
Training Loss	0.8490	0.2241
Validation Accuracy	0.8908	0.9500
Validation Loss	1.4380	0.3461

Key segmentation metrics (accuracy, loss) at start and completion of U-Net training for tooth segmentation.

mAP@0.5 with balanced precision (0.739) and recall (0.773). Class-wise analysis reveals that implant and fillings detection exhibit high reliability, while impacted tooth detection shows moderate sensitivity. Cavity detection remains the most challenging class, primarily due to limited sample representation and subtle radiographic contrast characteristics.

The observed performance distribution correlates with dataset imbalance, where classes with higher instance counts exhibit stronger localization accuracy. This highlights the importance of balanced data collection in medical detection tasks. The lower mAP@[0.5:0.95] compared to mAP@0.5 reflects stricter localization sensitivity under higher IoU thresholds, which is expected in small-object detection scenarios.

The U-Net segmentation results in Table II demonstrate stable convergence behavior, with progressive reduction in training and validation loss across epochs. The segmentation model effectively isolates tooth-bearing regions, supporting reliable anatomical localization. The high validation accuracy confirms suitability of U-Net for dental segmentation tasks.

From a clinical perspective, high implant and filling detection performance supports potential use in treatment planning and documentation workflows. However, cavity detection performance requires further dataset expansion before clinical deployment. Class imbalance remains the primary limitation, particularly for underrepresented pathological categories.

The system deployment on Hugging Face Spaces enables browser-based access without requiring users to manage local model execution. While inference performance depends on provisioned compute resources, the deployment framework supports interactive testing and correction-driven refinement in a cloud-hosted environment.

V. CONCLUSION

This paper presents a dual-model deep learning framework integrating U-Net-based segmentation with YOLOv8-based detection for automated analysis of OPG radiographs. The segmentation model achieved 97.23% validation accuracy, while the detection model reached 78.7% mAP@0.5, demonstrating effective anatomical localization and multi-pathology detection. The combined framework provides complementary anatomical and pathological insights within a unified system. The inclusion of a correction-driven learning mechanism enables progressive model improvement through user feedback, supporting practical clinical applicability.

While the system shows promising performance for screening and documentation, challenges such as dataset imbalance, particularly for cavity detection remain. Further dataset expansion and clinical validation are required for reliable real-world deployment.

VI. FUTURE WORK

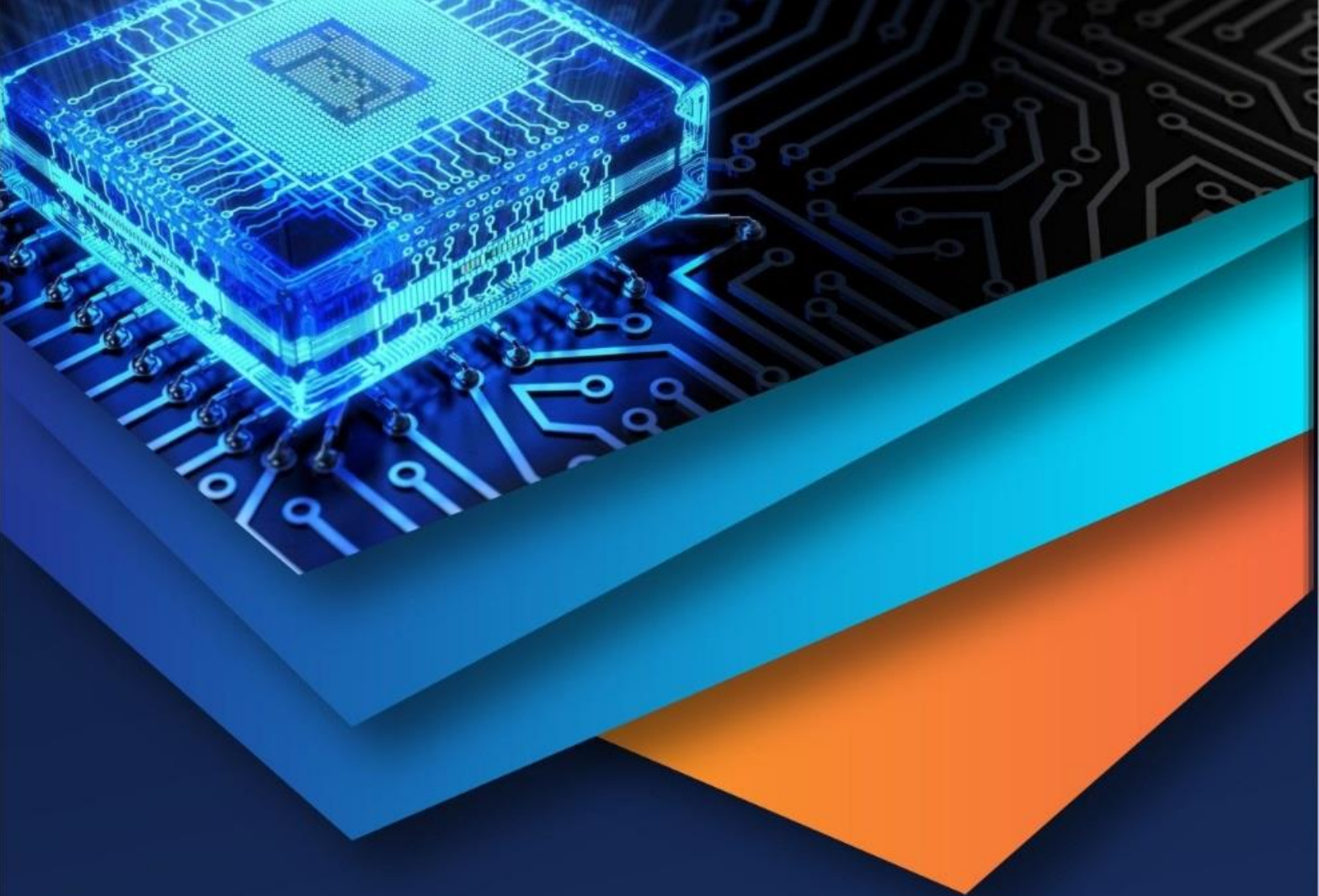
Future work will focus on addressing dataset imbalance, particularly for cavity detection, through expansion of underrepresented classes and improved augmentation strategies. Further improvements may include model optimization, architectural refinements, and exploration of advanced detection variants to enhance performance and generalization. Clinical validation using independent datasets and multi-center studies will be essential to ensure reliability. Additionally, future extensions may explore three-dimensional dental imaging and integration into real-world clinical workflows.

VII. AUTHOR CONTRIBUTIONS

Mr. Ashray Chaitanya, Mr. Balaji M, Mr. Akshat Rai and Mr. Achyutha Nayaka proactively collaborated and contributed to methodological brainstorming, experimental analysis, model development and deployment along with drafting of the manuscript. Dr. Devi Ganesan streamlined the problem definition and advised on the deep learning methodology. Dr. Roopa H advised and supervised the conduct of overall research.

REFERENCES

- [1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 2015, pp. 234–241.
- [2] G. Jocher, A. Chaurasia, and A. Qure, "YOLOv8: State-of-the-art real-time object detector," Ultralytics, 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [3] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [4] L. R. Dice, "Measures of the amount of ecologic association between species," Ecology, vol. 26, no. 3, pp. 297–302, 1945.
- [5] T.-Y. Lin et al., "Microsoft COCO: Common objects in context," in European Conference on Computer Vision (ECCV), Springer, 2014, pp. 740–755.
- [6] J. Lee, H. Kim, and J. Park, "Deep learning in dental radiography: Detection of dental caries using convolutional neural networks," Journal of Dental Research, vol. 97, no. 11, pp. 1245–1252, 2018.
- [7] Y. Chen, L. Zhang, and Z. Li, "Automated tooth detection and numbering in panoramic radiographs using deep neural networks," Computerized Medical Imaging and Graphics, vol. 80, 2020.
- [8] S. M. Khan, A. Rahman, and M. U. Akram, "Artificial intelligence in dental radiology: A review of applications and challenges," Dentomaxillofacial Radiology, vol. 49, no. 5, 2020.
- [9] Humans in the Loop, "Teeth Segmentation on Dental X-ray Images," Kaggle, 2023. [Online]. Available: <https://doi.org/10.34740/KAGGLE/DSV/5884500>
- [10] IMT Kaggle Team, "Dental Radiography Dataset," Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/imtkaggleteam/dentalradiography/data>. [Accessed: Apr. 2026].



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)