Automated Multi-Task Deep Learning Framework for Simultaneous Identification of Dental Pathologies and Arch Localization in Orthopantomograms (OPG)

Hatim Elhag[1], Norehan Binti Mokhtar[2], Siti Noor Fazliah Binti Mohd Noor[3],
Gururajaprasad Kaggal Lakshmana Rao[4], Nordin Zakaria[5], Fakhitah Ridzuan[6], Mohammed Bin Salah[7]
Dental Simulation and Virtual Learning Research Excellence Consortium, Department of Dental Science,
Advanced Medical and Dental Institute, Universiti Sains Malaysia, Bertam, Penang, Malaysia[1, 2, 3]
Department of Orthodontics, Penang International Dental College, NB Tower 5050,
Jalan Bagan Luar, Butterworth, Penang, Malaysia[4]
Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Seri Iskandar 32160, Malaysia[5]
Faculty of Data Science and Computing, Universiti Malaysia Kelantan, Pengkalan Chepa, 16100 Kota Bharu, Kelantan[6]
Department of Oral and Maxillofacial Surgery-College of Dentistry, King Saud University, Riyadh, Saudi Arabia[7]

Abstract—**Interpreting orthopantomograms (OPGs) accurately and efficiently remains challenging due to overlapping anatomical structures and the subtle presentation of many dental pathologies. While deep learning (DL) has shown considerable potential in dental imaging, most existing models focus on single-task detection and do not address the need for simultaneous analysis of both structural and pathological features within a unified system. To fill this gap, an automated multi-task deep learning framework is proposed that concurrently identifies ten dental pathology classes such as fractured roots and periodontally compromised teeth alongside arch-region classification of edentulous spans in OPGs. This method makes use of the YOLOv11 object detection architecture, which is powered by the C2PSA attention module and sophisticated post-processing algorithms that were trained on a larger collection of ten classes from the Dental OPG Kennedy dataset. With a mean Average Precision (mAP) 0.854 and an F1-score of 0.961 for every class assessed, the model demonstrated strong overall performance. Even in more complicated situations like root fractures (F1 = 0.928), the model was able to recognize well-defined edentulous areas with almost perfect sensitivity (Recall = 0.997) and good mAP. By integrating diagnostic analysis and anatomical techniques in a single optimized inference process, the YOLOv11 system offers interpretable and scalable solutions for comprehensive screening, thereby representing an advance for Computer-Assisted Dentistry Diagnostics.**

Keywords—Orthopantomogram (OPG); deep learning (DL); multi-task learning; YOLOv11; object detection; dental pathology; arch classification; Computer-Aided Diagnosis (CAD)

I. INTRODUCTION

Nowadays, panoramic radiography, in the form of an orthopantomogram (OPG), is an important imaging modality in dental practice as a fundamental examination that covers the maxilla and mandible along with complete dentition and related anatomic structures [1]. Therefore, the entire process of screening, treatment planning, and follow-up can be realized in different environments, including general dental care and specialist treatment. Nevertheless, the OPG is by nature a complex and time-consuming image to be interpreted. Specifically, anatomical structures overlapped with each other, and the image quality varied from one case to another; hence, subtle signs of pathological processes were not easy to identify accurately by human beings [2].

Recent developments in Artificial Intelligence (AI) and Deep Learning (DL) have been finding a lot of promise in creating diagnostic tools in medical imaging areas such as dental radiology [3]. There have been several sources that reveal a rapidly increasing trend for implementing AI systems in the dental field, including detection or analysis of teeth, segmentation analysis, dental caries detection, analysis of endodontic lesions, and implant assessment. For instance, AbuSalim et al. [4] highlight the use of DL technology in dental informatics and the importance of implementing complex and explainable artificial intelligence systems that can improve efficiency in the healthcare sector. Similarly, Ali et al. [5] also include a detailed explanatory study related to the use of AI in dental radiology, highlighting its potential for increasing diagnostic accuracy, reducing radiation exposure, and improving patient outcomes.

Among DL approaches, Convolutional Neural Networks (CNNs) as well as those based on transformers have been utilized to identify dental health conditions at a similar level of human expertise. Especially in terms of identifying dental images, it has been noted that the use of Single-Stage object detection techniques such as You Only Look Once (YOLO) is a highly promising one because of its fast detection speed and efficient spatial localization technique [6]. For example, the application of the YOLOv3 model was proposed by Almalki et al. [7] for the detection of dental pathologies from the panoramic OPG X-ray images. In their approach, the Darknet-53 framework was proposed with the identification of four dental pathologies, namely cavity, root canal, dental crown, and broken-down roots, with the achievement of over 90% mAP on four different dental pathologies. Furthermore, Beser et al. [8] developed a YOLOv5 model with automatic segmentation and numbering of deciduous and permanent teeth in mixed-dentition radiographs. Their YOLOv5 model scored high sensitivity, specificity, and accuracy (compared with other architectures like Faster R-CNN [9]), while also showing superior computational efficiency. Similarly, Makarim et al. [10] compared YOLOv5 with other models for classification of dental pathological images and reported that YOLOv5 achieved higher accuracy as well as faster processing times than alternative methods, demonstrating its suitability for multi-class dental object detection. Ismail and Taseen [11] proposed a new YOLOv8 architecture for the detection of 14 classes of dental anomalies. Their solution included the use of the anchor-free detection head, CSPDarknet53 as the backbone network, and the C2f feature fusion module. Although their model achieved high accuracy above 0.85 for structural tooth abnormalities such as implants and crowns, it failed to demonstrate similar performance in detecting subtle conditions such as caries and periapical lesions. Based on this perspective, Şevik and Mutlu in [12] proved that it is possible to use the YOLOv11 model in identifying pediatric dental anomalies in panoramic radiographs. According to their findings, YOLO's fast learning and strong small object detection capabilities make it ideal for clinical dental image processing tasks.

Despite these advances, a number of significant gaps still exist in applying DL to panoramic dental images. One is that most of the literature focuses on single-task problems (eg., detect certain pathology, count teeth, classify implant brands) rather than multi-task approaches that merge several related tasks together within a single framework. For example, a multi-task learning study on mandibular third-molar classification found that learning multiple outputs simultaneously didn't always outperform single-task models [13]. In addition, multi-task frameworks that perform both anatomic classification and pathology detection are rarer still. Combining these two problems might yield more robust and meaningful clinical predictions. Publicly available OPG data used for research purposes is often too small or lacks diversity in terms of molecular composition, leading to issues of generalizability for models trained on such limited datasets and adaptation between fields outside their original domain [14]. In order to overcome the above-mentioned challenges, the following study introduces a comprehensive framework for multi-object detection in OPG radiography that can detect multiple dental pathologies while simultaneously classifying arch topology. The proposed framework adopts YOLOv11-based object detection and OpenCV-driven image analysis to jointly identify pathology localization and arch type in panoramic radiographs. A ten-class taxonomy is established to describe a variety of pathological and positional characteristics, including root fractures, periapical states, and different arch types. This framework is intended to advance Computer-Aided Diagnosis by delivering interpretable, efficient, and scalable analysis suitable for clinical screening workflows.

The remainder of this study is organized as follows: Section II reviews the related work on deep learning-based dental image analysis and identifies the research gap addressed in this study. Section III presents the proposed methodology, including dataset preparation, the YOLOv11 architecture, and the evaluation metrics. Section IV discusses the experimental results and analyzes the performance of the proposed framework. Section V provides a qualitative comparison with existing deep learning-based dental detection methods. Section VI discusses the clinical significance of the proposed framework, while Section VII presents the clinical interpretation of prediction errors and their implications for diagnostic decision-making. Section VIII outlines the limitations of the current study and directions for future research. Finally, Section IX concludes the study and highlights potential future research directions.

II. Related Work

Deep learning has significantly advanced automated dental image analysis by enabling accurate detection, classification, segmentation, and localization of various dental conditions from different imaging modalities. Recent studies have demonstrated the effectiveness of convolutional neural networks (CNNs), object detection frameworks, and multi-task learning strategies for improving diagnostic accuracy while reducing the need for manual interpretation.

Several studies have focused on automated dental pathology detection using deep learning. ToothNet, a YOLOX-based framework, was developed for the simultaneous detection of dental caries and fissure sealants from intraoral photographs, achieving image-level AUC values of 0.925 for caries detection and 0.902 for sealant detection. At the tooth level, the framework achieved a sensitivity of 0.807, precision of 0.814, and an F1-score of 0.810, demonstrating the effectiveness of object detection models for dental disease identification [15]. Similarly, a hybrid deep neural network integrating CNN, BiLSTM, and attention mechanisms was proposed for predicting dental treatment outcomes by combining radiographic, clinical, and behavioral information, achieving an overall accuracy of 94.2% and an AUC of 0.96 [16]. Although these approaches demonstrated promising performance, they primarily focused on individual diagnostic tasks rather than providing comprehensive analysis of panoramic dental radiographs.

Recently, multi-task learning (MTL) has attracted considerable attention because of its ability to simultaneously solve multiple related clinical tasks while learning shared feature representations. A CBAM-based multi-task framework jointly performed age estimation and gender classification from CBCT panoramic slices, achieving a mean absolute error of 1.08 years for age estimation and 95.3% classification accuracy for gender prediction, outperforming conventional single-task approaches [17]. Likewise, a hybrid MTL framework combining U-Net++, EfficientNet-B4, pseudo-label generation, and selective loss masking simultaneously performed anatomical segmentation and caries detection under partially annotated datasets, achieving a Dice Similarity Coefficient (DSC) of 0.6706, an Intersection over Union (IoU) of 0.5044, and improving sensitivity by 7.47% while reducing false-negative pixels by 19.9% [18]. These studies demonstrate the advantages of multi-task learning; however, they mainly address classification or segmentation problems rather than simultaneous pathology detection and anatomical localization.

Several researchers have also investigated tooth detection and localization using deep learning. A relation-based framework incorporating prior dental knowledge employed a multi-task CNN and a label reconstruction strategy to improve tooth recognition and localization in dental periapical radiographs, outperforming conventional object detection approaches [19]. Similarly, a cascade CNN framework utilizing hierarchical label-tree representations achieved 95.8% precision and 96.1% recall for identifying all 32 tooth positions, effectively handling challenging cases involving missing, decayed, and restored teeth [20]. Despite these encouraging results, these approaches were primarily designed for tooth identification and localization rather than detecting multiple dental pathologies from panoramic OPG images.

Despite the rapid progress of deep learning in dental imaging, several important limitations remain. First, most existing studies focus on a single clinical task, such as caries detection, tooth localization, implant classification, age estimation, or treatment outcome prediction, rather than providing a comprehensive assessment of panoramic dental radiographs. Second, many methods are developed for intraoral photographs, CBCT images, or periapical radiographs, limiting their direct applicability to panoramic OPG imaging. Third, existing studies employ different datasets, annotation protocols, class definitions, and evaluation methodologies, making direct comparison difficult. Moreover, few studies simultaneously integrate dental pathology detection and anatomical arch localization within a unified object detection framework.

Motivated by these limitations, this study proposes a YOLOv11-based multi-task framework for simultaneous detection of multiple dental pathologies and localization of maxillary and mandibular arch regions using an expert-annotated extension of the Dental OPG Kennedy dataset. Unlike previous studies that primarily address individual diagnostic tasks, the proposed framework combines comprehensive pathology detection and anatomical localization within a single inference pipeline, providing a more clinically relevant solution for computer-aided analysis of panoramic dental radiographs.

III. Proposed Methodology

This section describes the methodology adopted to achieve the main objective of this study, which is to perform multitask dental pathology feature detection and dental arch classification on OPG panoramic X-ray images using the YOLOv11 framework integrated with OpenCV-based analysis. The overall operating procedure of the proposed system is shown in Fig. 1. This involves four main steps: preparing the dataset, developing the YOLOv11 model, and evaluating and analyzing the results.

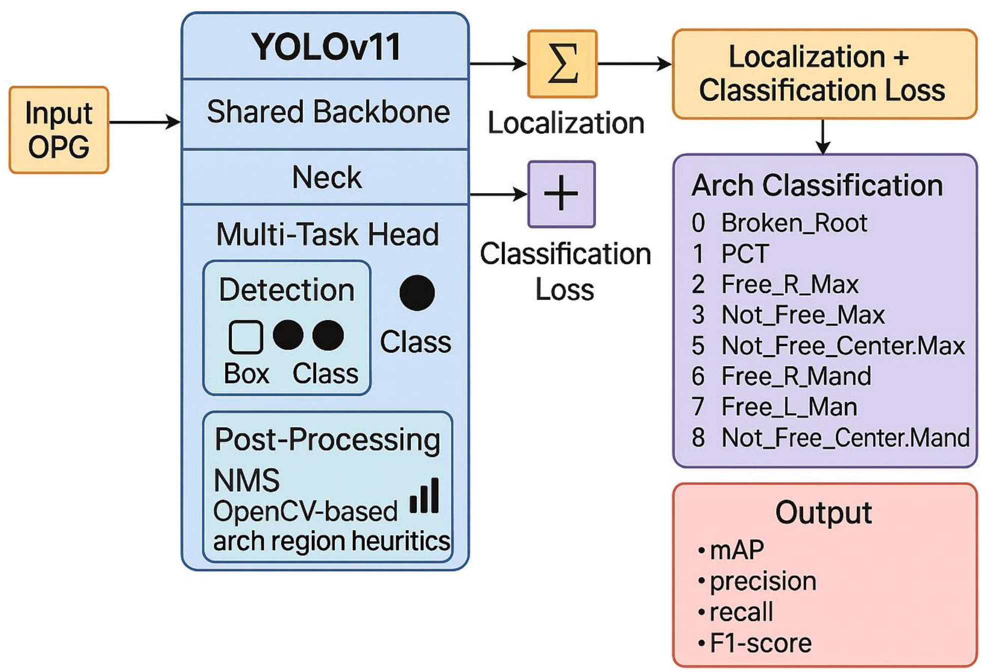

Fig. 1. Overview of the proposed multi-task object detection framework for dental pathologies and arch classification in panoramic OPG X-rays.

The following subsections explain each stage in detail.

A. Dataset Preparation

In this study, the publicly available Dental OPG Kennedy dataset was used, which was created by Orville Kennedy et al. and published on Kaggle [21]. This dataset contains panoramic dental radiographs called orthopantomograms (OPGs). It was originally collected to classify partially edentulous dental arches based on the Kennedy system. All the images are annotated according to root fracture, periodontally compromised tooth (PCT), and edentulous arch classes. These annotations are in the form of YOLO-ready text files, comparable to many other datasets, with each line representing a bounded box coordinate in addition to a class identifier. There are 500 dental X-ray images in PNG format, each with an average size of approximately 1500x800 pixels.

To support the objectives of this study, the original three-class annotation scheme was extended to ten clinically meaningful categories, enabling the detection of a broader range of dental conditions. The additional annotations were manually performed by a qualified dental expert according to predefined clinical diagnostic criteria. Each panoramic radiograph was carefully reviewed to ensure that the assigned labels accurately represented the corresponding dental pathology or anatomical region. This expert-guided annotation process enhanced the clinical relevance and consistency of the extended dataset, providing reliable ground-truth labels for training and evaluating the proposed framework. The resulting annotation scheme comprises ten clinically meaningful categories, including root fractures, periodontally compromised teeth (PCT), and edentulous conditions affecting both the maxillary (upper jaw) and mandibular (lower jaw) arches, as summarized in Table I. Although the annotations were performed by a qualified dental expert, formal quantitative assessment of inter-annotator agreement and labeling uncertainty was beyond the scope of the present study and will be considered in future work.

TABLE I. FINAL 10-CLASS TAXONOMY USED FOR MULTI-TASK DETECTION.

Class ID	Class Name	Anatomical Region	Clinical Description / Meaning
0	Broken_Root	Both arches	Tooth root fracture or retained root fragment visible on OPG.
1	PCT (Periodontally Compromised Tooth)	Both arches	Tooth affected by advanced periodontal bone loss or mobility.
2	Free_R_Max	Maxilla – Right	Free-end edentulous space on the right side of the maxillary arch.
3	Free_L_Max	Maxilla – Left	Free-end edentulous space on the left side of the maxillary arch.
4	Not_Free_Max	Maxilla – Posterior	Bounded edentulous region (teeth missing between adjacent teeth) in maxilla.
5	Not_Free_Center_Max	Maxilla – Center	Central bounded edentulous area in anterior maxillary region.
6	Free_R_Mand	Mandible – Right	Free-end edentulous space on the right side of the mandibular arch.
7	Free_L_Mand	Mandible – Left	Free-end edentulous space on the left side of the mandibular arch.
8	Not_Free_Mand	Mandible – Posterior	Bounded edentulous region (non-free space) in posterior mandibular area.
9	Not_Free_Center_Mand	Mandible – Center	Central bounded edentulous space in anterior mandibular region.

This configuration efficiently enables the model to detect and classify pathological classes (classes 0-1) and arch types (classes 2-9), simultaneously. All the labels were double-checked by dental professionals to guarantee they are all clinically correct. Data pre-processing was carried out with OpenCV and other image processing libraries. All training images were resized to the 640 × 640 pixels expected by YOLOv11 input. The original PNG images were converted to grayscale in order to save computational cost and highlight radiographic features. The contrast limited adaptive histogram equalization (CLAHE) [22] was used to enhance the contrast, and the bidirectional filtering was applied to suppress noise while preserving edge sharpness, then pixel values were normalized between 0 and 1. Finally, the dataset was divided into 70% training, 20% validation, and 10% testing. This partition was maintained consistently throughout all experiments to ensure a fair and reproducible evaluation of the proposed framework. While a fixed data split facilitates consistent comparison under identical training conditions, it does not fully capture variability across different data partitions.

B. YOLOv11 Model Architecture

To achieve high computation efficiency with high precision feature extraction, the YOLOv11 model was selected as a detection framework. As illustrated in Fig. 2, the network consists of three blocks: 1) Backbone, 2) Neck, and 3) Detection Head.

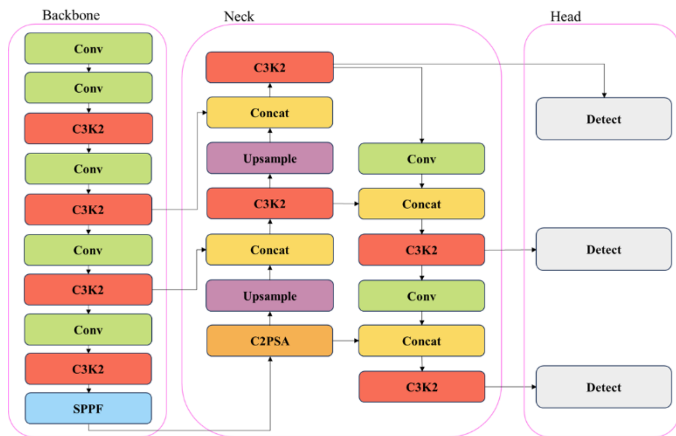

Fig. 2. YOLOv11 architecture

It leverages a series of convolutional layers equipped with Cross-Stage Partial (CSP) and C3K2 modules to promote the gradient flow and feature reutilization, without introducing extra computation overhead [23]. The Spatial pyramid pooling-fast (SPPF) block enhances the generalization capability of the model for object size changed in images by incorporating multi-scale context information from several receptive field. The neck incorporates *C2PSA* (Cross-Stage Partial with Spatial Attention) blocks, enabling spatial-attention refinement to emphasize diagnostically relevant regions [24], such as periapical lesions or edentulous margins, within panoramic X-rays. The detection head operates on three scales: P_3, P_4, and P_5 and outputs a vector:

$$[x, y, w, h, o, p_1, p_2, \dots, p_C] \quad (1)$$

For each anchor, where (x, y) denote the normalized box-center offsets, (w, h) denote the width and height, o represents the objectness confidence, and p_i are class probabilities for $C = 10$ dental categories. The total loss function combines localization, confidence, and classification losses as:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{CIoU}} + \beta\mathcal{L}_{\text{obj}} + \gamma\mathcal{L}_{\text{cls}} \quad (2)$$

where, $\mathcal{L}_{\text{CIoU}}$ represents the Complete-IoU regression loss for bounding-box optimization, and $\mathcal{L}_{\text{obj}}$ and $\mathcal{L}_{\text{cls}}$ denote the binary and categorical cross-entropy losses, respectively. Hyperparameters α, β, and γ were empirically tuned to balance localization precision and class discrimination across both pathological and positional categories.

To adapt YOLOv11 to panoramic dental imagery, the model was fine-tuned via transfer learning using the pre-trained YOLOv11s checkpoint initialized on the COCO dataset. The anchor-box dimensions were recalculated using k-means clustering on the bounding-box distributions of the dental dataset to better accommodate small and irregular anatomical structures. Training was conducted at an input resolution of 640×640 pixels with a batch size of 16 and a learning rate of 1×10^{-3} employing a cosine decay schedule and warm-up phase. The optimizer used SGD with momentum 0.937 and weight decay 5×10^{-4}, while early stopping was applied after 100 epochs to prevent overfitting. During inference, Non-Maximum Suppression (NMS) was performed with an IoU threshold of 0.45 and a confidence threshold of 0.25 to eliminate duplicate predictions.

C. Model Evaluation and Performance Measures

The model was evaluated on standard object detection metrics, precision, recall, mean average precision, mAP, F1-score, to check how well the detector is performing to classify the various objects with different confidence levels.

Precision (P) indicates how many positives detected are actually true and gives an indication of the false positive rate. It is calculated as:

$$P = \frac{TP}{TP+FP} \quad (3)$$

Here, TP and FP indicate the number of true positives and false positives, respectively. A higher accuracy value indicates that the detector is making accurate predictions and correctly identifying real objects with minimal false positives.

Recall (R) evaluates the sensitivity of a model to all real objects in an image by quantifying the proportion of correctly identified objects to the total number of valid instances. It is given by:

$$R = \frac{TP}{TP+FN}, \quad (4)$$

Here, FN indicates a false negative, i.e. an object missed by the detector. High recall means the detector detects most objects in the image.

The F1 score provides a smooth average between precision and recall, providing a balanced view of the detector's overall performance in correctly identifying objects. It is expressed as:

$$F_1 = 2 \times \frac{P \times R}{P+R} \quad (5)$$

where, $F_1 = 1$ indicates perfect detection performance and $F_1 = 0$ represents complete failure. This metric is especially useful in evaluating trade-offs between precision and recall when optimizing detection thresholds.

The mean Average Precision (mAP) is the main metric used in object detection to measure accuracy across different Intersection-over-Union (IoU) thresholds. For each class, an Average Precision (AP) score is calculated by measuring the area under the Precision–Recall (PR) curve, defined as:

$$AP = \int_0^1 P(R)dR \quad (6)$$

and the mAP is then calculated as the mean of AP scores over all object classes:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (7)$$

where, N is the total number of classes. The mAP@50–95 metric is reported by averaging the AP scores across several IoU thresholds, ranging from 0.5 to 0.95 in steps of 0.05. This approach offers a thorough assessment of how accurately the model localizes and detects objects.

These metrics combined give a comprehensive view of how well the detection system is performing. They balance the trade-offs between precision, recall, and localization accuracy, and are widely accepted as the standard criteria for evaluating modern deep learning object detection frameworks like YOLOv11.

IV. Results and Discussion

Fig. 3 shows the evaluation of class-wise performance using mean Average Precision and Recall metrics at IoU ranges between 0.50 and 0.95 (mAP@50-95) highlighting the discriminative, as well as structural limitations of the multi-task framework based on YOLOv11 architecture.

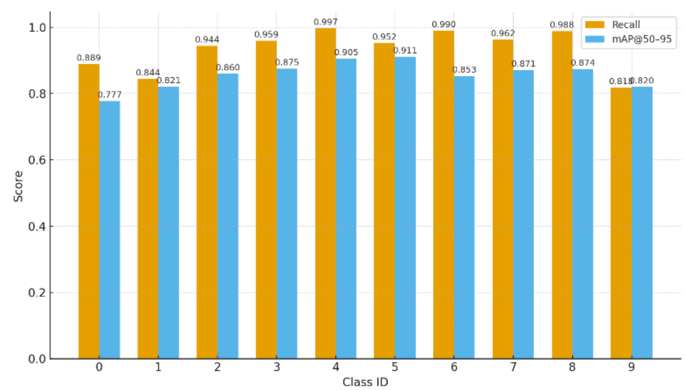

Fig. 3. Per-class comparison of Recall and mAP@50–95 for YOLOv11.

One core limitation of the model is observed in the Broken_Root (Class 0) category, with a Recall of 0.889 and mAP@50–95 of 0.777. Even though the YOLOv11x model benefits from the channel-dependent Cross-Stage Partial Spatial Attention (C2PSA) modules, the detection of minor, irregular-shaped fractures are still a struggle. The radio-image lucent and anisotropic orientation of fractured roots with overlying implant trabecular bone greatly degrades the distinct mid-level convolutional feature. Therefore, the destined abstraction is reduced within the feature pyramid, partially detecting and under-segmenting what is needed. Similarly, the Recall (0.844) and mAP@50-95 (0.821) of the Periodontal Compromised Teeth (PCT) exhibits the moderate sensitivity of the model in detecting pathological patterns – limiting the spatial precision. The ability of the anchor-free detector to regress high-IoU bounding boxes is reduced by the boundaries and margins of the diffuse lesion and irregular bone-loss. According to these observations, it is suggested that even though the attention selectivity of the model is improved by the C2PSA mechanism, relying on localized regions with low contrasts is still restricted by the radiographic contrasts and limited spatial information.

Conversely, Free_R_Max (Class 2) and Free_L_Max (Class 3) categories exhibit strong localization performance (Recalls = 0.944 and 0.959; mAP@50–95 = 0.860 and 0.875). This demonstrates the ability of the model to extract persistent signs of ridge-morphology using multi-scale feature collections. The path Aggregation Network (PAN) of the YOLOv11 two-way feature fusion functionally integrates coarse and fine-grained contextual features, allowing accurate identification of balanced edentulous areas despite uneven illumination in wide-angle projections. The Not_Free_Max (Class 4) category (Recall = 0.997, mAP@50–95 = 0.905) further emphasizes the ability of the architecture to describe well-defined ridges in the mouth. The brightness gradient between dentate and edentulous sections is efficiently discriminated by the attention-guided backbone, proving the high spatial discrimination ability of the model. Similarly, a fine-grained identification of subtle edentulous structures within dense preceding mouth bones (usually limited by low contrasts) is also observed through the Not_Free_Center_Max (Class 5) category, having a Recall of 0.952 and mAP@50-95 of 0.911.

In the jawbone regions, both Free_L_Mand (Class 7) (Recall = 0.962, mAP@50-95 = 0.871) and Free_R_Mand (Class 6) (Recall = 0.990, mAP@50-95 = 0.853) both show high sensitivity, although slight localization inconsistencies persist near the posterior curvature. These disparities emerge from variations in cortical densities and uneven geometries in the jawline, which reduce feature spreading in the Feature Pyramid Network (FPN) pathway. A strong detection and localized accuracy is observed in the Not_Free_Mand (Class 8) category, having a Recall of 0.988 and mAP@50-95 of 0.874. This demonstrates the model's ability to monitor dorsal region containing irregular geometries and overlying impositions. On the other hand, the Not_Free_Center_Mand (Class 9) category displays great difficulty in detection with Recall of 0.818 and mAP@50-95 of 0.820. Incisors exhibiting low illumination, shadows from overlapping soft tissues, and thin cortical boundaries cause features to dilute in early convolutional layers, causing insignificant bounding box regressions to be produced.

The values of Fig. 3 evidently quantify accurate relationship between the plots and evaluation metrics, as described in the text above. Average values of Recall and mAP@50-95 specify the persisting detection ability of the YOLOv11 model across most segments, with some limitations faced in complicated and radiologically complex areas. In conclusion, accurate analysis of orthopantomogram (OPG) is improved by integrating contextual attention and multi-scale fusion techniques. In spite of that, some constraint still persists for minute, less-illuminated, and dark radio-imagery in dental domain, particularly in regions of deceiving bone-lesions. Future improvements of the model may involve hybrid transformer-CNN encoders to help improve localization accuracy in complex radiographic scenarios.

Fig. 4 quantitatively exhibits the balance of the model through F1 scores of each class. The sensitivity and precision values can be seen across all ten categories of dental pathology. This per-class evaluation boosts the stability of the model in accurate detection of multi-class performance.

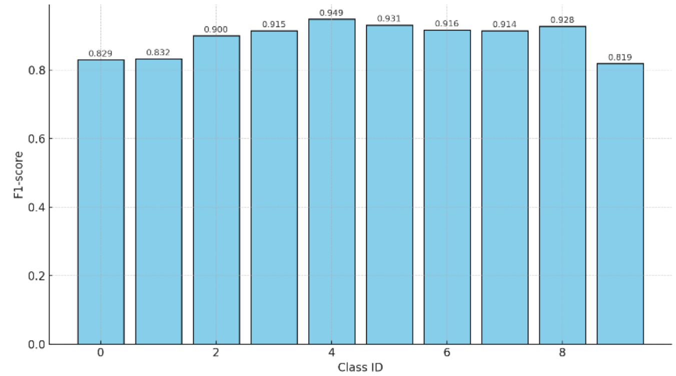

Fig. 4. Per-class comparison of F1-score for YOLOv11.

The Broken_Root (Class 0) (F1-score = 0.829) and PCT (Periodontally Compromised Tooth) (Class 1) (F1-score = 0.832) categories exhibit the lowest F1 scores. These values may be the result of previously-discussed challenges such as complex radiographs, uneven arrangement, and overlying trabecular patterns that limit accurate results. Similarly, the F1-score of the Not_Free_Center_Mand (Class 9) category (0.819) – low due to less-illuminated incisors, shadows, etc., is another example of restricted lesion discrimination. On the other hand, the YOLOv11 model achieves high F1-scores for the Free_R_Max (Class 2) (0.900) and Free_L_Max (Class 3) (0.915) categories, proving the model's ability in detecting even edentulous configurations within wide projections. The highest performance of 0.949, 0.931, and 0.928 for the Not_Free_Max (Class 4), Not_Free_Center_Max (Class 5), and Not_Free_Mand (Class 8) categories exhibit the localization ability of the model in well-defined and high contrast regions. The classes usually have distinct boundaries and even background densities, allowing the model to accurately discriminate spatial features and bounding box regressions.

On average, the F1-scores of all classes remain above 0.80, determining the model's accurate generalization and diagnosing ability across varying dental patterns. The YOLOv11 outperforms prior architectures of YOLOv5 and YOLOv8 that achieved an average clinical F1-score of 0.78 [25], [26]. This improvement may be attributed to the feature recalibration attention mechanism of C2SPA, the multi-scale fusion technique of the model, and boosted accuracy through Mosaic and MixUp pipelines to illuminate panoramic X-rays.

The normalized confusion matrix differentiating between true and predicted dental pathology classifications can be seen in Fig. 5 for all ten classes. The 'background' label in the confusion matrix does not represent an actual class in the dataset but corresponds to false-positive or unmatched predictions generated during evaluation. The discriminative reliability of the YOLOv11 model can be observed by analyzing consistent classification and inter-class confusion areas.

The diagonal matrix proves the harmony among true and predicted variables. This indicates that the proposed model successfully learns discriminative radiographic representations for most dental classes. High precision with 489 correct classifications and one misclassification in *Class 0* emphasizes YOLOv11 detection ability for well-defined structures. Similarly, 290 and 342 correct predictions and minimum offdiagonal occurrences for *Classes 4 and 8* highlight the capacity of the network in capturing regular textures and uniform geometries in the regions. Some confusion in *Classes 1, 3, and 6* is observed with slight misclassifications. These minor occurrences may have emerged due to common grayscale attributes, scattered lesion borders, and overlying radiographic patterns in similar regions. These anomalies usually occur in panoramic imaging due to imposed complexes and distorted projections, which can cause restrictions even in advanced detectors.

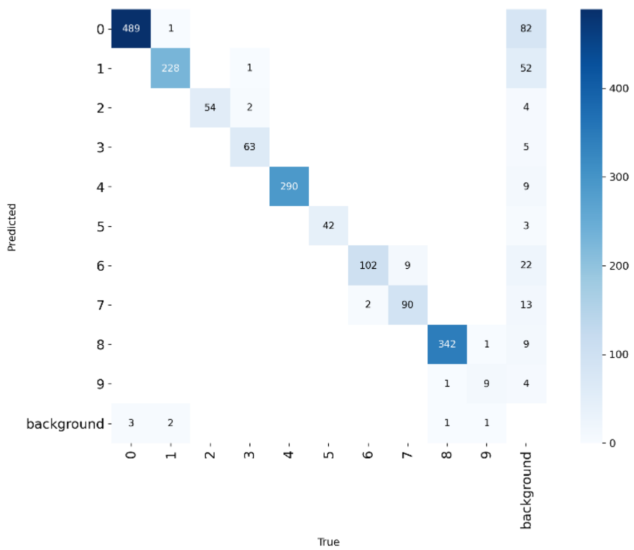

Fig. 5. Confusion matrix

The small number of misclassifications refers to weak and incomplete feature samples where the detector failed to place them into correct target classes. Such occurrences typically result from subtle, unclear radiographic segments or low-contrast lesions with incomplete information. It can be noted that no class displays dominating misclassifications in the system, proving a balanced learning behavior and no bias in the dataset. Overall, the confusion matrix validates the discriminative ability of the YOLOv11 model in complex diagnosing scenarios.

The residual errors may have occurred due to similar and overlapping neighboring entities. These limitations may be overcome by using augmentations specific to the domain (e.g., normalizing radiographic contrasts, spatial deformations), feature identification using high resolutions, or refining feature discrimination.

Training and Validation box regression loss of the YOLOv11 model across 100 epochs is given in Fig. 6. The curves for both training and validation datasets evolve to a consistent and stable downward trend, proving the model's ability in achieving spatial orientation between true and predicted bounding boxes.

For the training dataset, both curves display losses on high magnitudes, given random initial weights and raw spatial features. A steep decline in the first 10-15 epochs explains the learning of the network of localization features that are coarse. In succession, a gradual, close parallel decrease in both losses verifies smooth and accurate gradient traveling. The training box shows a consistently descending loss to about 0.6 by the 100th epoch, successfully reducing localization error by continuous learning. Loss for validation begins at a higher magnitude – indicating hidden data presence – then gradually converges towards the training loss after about 30 epochs. The combination of this behavior and minor oscillations exhibits the ability to generalize accurately without severely overfitting the outcome. The training and validation loss curves nearly overlap in the ending epochs, indicating a balance between bias and variance in the model.

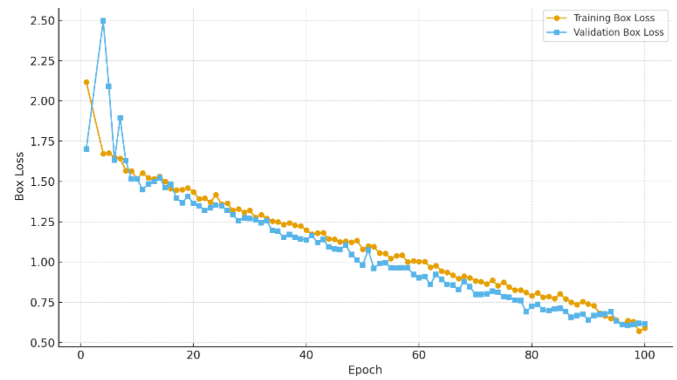

Fig. 6. Training vs. validation box loss over epochs.

The YOLOv11 model proves to be stable and efficiently optimized. The intensified backbone and multi-feature fusion of the network deeply facilitate the inclination traveling and reusability of spatial information across layers, preventing gradients from disappearing in cases of long training. Meanwhile, the enhanced multi-feature accumulation within the bidirectional PAN ensures efficient communication among top-down and bottom-up cues. This approach enables consistent spatial and contextual awareness, allowing correct bounding-box regression across physical diversity and varying dental patterns. All in all, the consistent loss curves indicate efficient learning rates when needed along with normalization and stable back propagations. The convergence pattern proves reliable optimization and precise spatial location in varying dental structures in OPG imagery.

V. Comparative Analysis with Existing Models

Recent advances in automated dental radiography analysis have demonstrated promising results using various YOLO-based object detection frameworks. However, the published studies differ considerably in terms of datasets, annotation protocols, class definitions, and evaluation methodologies. Consequently, direct quantitative comparison of reported performance metrics is challenging. The purpose of this section is, therefore, to position the proposed framework within the existing literature by highlighting methodological differences and overall performance trends rather than providing a direct benchmark against previous methods. Peker and Kurtoglu [27] implemented YOLOv10 model performed well in tooth detection on pediatric panoramic images (F1 = 0.919, mAP@50–95 = 0.696) but was limited to a single task and showed reduced precision at higher IoU thresholds. Furthermore, Almalki et al. [28] proposed YOLOv3 approach reached 99.33% accuracy in identifying broad dental disease types but lacked fine localization and modern attention mechanisms. In contrast, the proposed YOLOv11 model achieved an overall mAP@50–95 of 0.854 on the extended Dental OPG Kennedy dataset while incorporating C2PSA attention and refined Cross-Stage Partial blocks to jointly detect dental pathologies and arch-region characteristics within a unified multi-task framework. Although these results demonstrate strong detection performance, direct quantitative comparison with previous studies should be interpreted with caution because the published methods were evaluated on different datasets, annotation schemes, and experimental settings.

TABLE II. QUALITATIVE COMPARISON OF REPRESENTATIVE YOLO-BASED DENTAL DETECTION STUDIES AND THE PROPOSED YOLOV11 FRAMEWORK.

Study	YOLO Version	Main Task	Key Findings / Remarks
Peker & Kurtoglu [27]	YOLOv10	Tooth detection & numbering	Robust tooth numbering; limited fine localization accuracy
Almalki et al. [28]	YOLOv3	Common dental disease detection	High accuracy; limited to coarse pathology detection
Adnan et al. [29]	YOLOv5m	Six pathology detection (OPG)	Strong diagnostic accuracy; single-task framework
Putra DDS et al. [30]	YOLOv4	Tooth numbering	Excellent speed; limited to structural detection only
AbuSalim et al. [31]	YOLOv5, v6, v7	Multi-granularity tooth classification	Accuracy declines with class granularity; multi-model setup
AbuSalim et al. [32]	HDL (YOLOv6 +CBAM-YOLOv5)	Hierarchical tooth classification	High accuracy & recall; increased computational complexity
Proposed YOLOv11 (This Study)	YOLOv11	Multi-task detection (pathologies + arch classification)	Simultaneous multi-class and multi-region detection with superior localization and balanced metrics

Adnan et al. [29] used YOLOv5m to detect six dental pathologies with strong accuracy (Precision 94.22%, Recall 90.42%, mAP 93.71%), though its application was limited to pathology detection alone. In comparison, the proposed framework extends this capability by simultaneously integrating lesion identification with multi-region arch classification within a single detection model. On the extended ten-class Dental OPG Kennedy dataset, the proposed YOLOv11 framework achieved an overall F1-score of 0.961, demonstrating balanced detection performance across pathological and anatomical categories. Putra DDS et al. [30] employed YOLOv4 for automated tooth numbering, achieving high recall (100%) and rapid inference (20.58 ms per image), but their system did not incorporate pathological analysis. Likewise, AbuSalim et al. [31] [32] investigated multi-granularity detection using YOLOv5–YOLOv7 and later developed a hierarchical deep learning (HDL) approach combining YOLOv6 with a CBAM-enhanced YOLOv5, achieving mAP = 0.88, Recall = 0.999, and F1 = 0.94. Although effective, their multi-stage framework increased computational complexity. In contrast, the proposed YOLOv11 framework unifies structural and pathological detection within a single-stage architecture and achieved an overall mAP@50–95 of 0.854 and F1-score of 0.961 on the extended Dental OPG Kennedy dataset. In addition to these quantitative results, the framework performs simultaneous pathology detection and arch classification in a single inference pipeline, reducing computational complexity compared with previously reported multi-stage approaches. Table II shows comparative analysis of YOLO-based dental detection studies and the proposed YOLOv11 multi-task model.

A sample of the output results from the proposed framework are shown in Fig. 7.

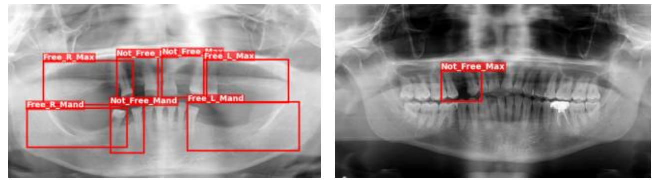

Fig. 7. Sample ot the output results.

VI. CLINICAL SIGNIFICANCE

The proposed multi-task YOLOv11-based framework has potential clinical value as a computer-aided decision support system for the analysis of panoramic dental radiographs. By simultaneously detecting multiple dental pathologies and identifying the corresponding maxillary and mandibular arch regions within a single inference pipeline, the framework can assist dentists in identifying abnormalities during routine clinical examinations, thereby supporting timely clinical decision-making and improving workflow efficiency. Automated localization of abnormalities may reduce the time required for radiographic interpretation, improve diagnostic consistency, and decrease the likelihood of overlooked findings, particularly in busy clinical environments. In addition, the framework may facilitate patient screening by automatically highlighting radiographs with suspected abnormalities, enabling clinicians to prioritize cases requiring further evaluation. The proposed framework may also serve as an educational aid for dental students and less experienced practitioners by highlighting clinically relevant regions of interest and reinforcing the interpretation of panoramic radiographs. Accordingly, the framework is intended to complement, rather than replace, professional clinical expertise, while final diagnosis and treatment planning remain the responsibility of qualified dental practitioners.

VII. CLINICAL INTERPRETATION OF PREDICTION ERRORS

In clinical practice, different prediction errors may have different consequences depending on the underlying dental condition. A false-negative prediction, in which an existing pathology is not detected, may delay diagnosis and treatment, potentially allowing disease progression. For example, missed root fractures or periodontally compromised teeth may result in continued structural damage, infection, or tooth loss if not identified during routine examination. Conversely, a false-positive prediction may lead to additional clinical examination or unnecessary radiographic review, increasing clinician workload but generally posing a lower clinical risk than missed pathology. Therefore, in screening applications, maintaining high sensitivity is particularly important to minimize overlooked abnormalities, while an appropriate balance between sensitivity and precision is required to avoid excessive false alarms. The proposed framework is intended to assist clinicians by highlighting suspicious regions for further review rather than providing definitive diagnoses. Consequently, all AI-generated findings should be interpreted alongside clinical examination and professional judgment before treatment decisions are made.

VIII. Limitations

Even with its outstanding performance, there are still some drawbacks. The model experienced a reduction in accuracy when it comes to detecting small dental issues such as root fractures and periodontal problems because panoramic radiographs create low image contrast and overlapping anatomical structures. Furthermore, the proposed framework was evaluated exclusively on the extended Dental OPG Kennedy dataset. Validation on larger and more diverse datasets collected from multiple institutions and imaging systems is required to further assess its robustness, clinical applicability, and generalizability across different patient populations and acquisition conditions.

Although the proposed framework demonstrated strong detection performance (mAP@50–95 = 0.854 and F1-score = 0.961), the individual contribution of each design component—including the preprocessing pipeline, C2PSA attention mechanism, and multi-task learning strategy—was not evaluated independently through ablation experiments. Consequently, the performance improvements reported in this study should be interpreted as the result of the integrated framework rather than any single component.

The experimental evaluation was conducted using a single train-validation-test split and a fixed training configuration. Although this protocol provides a consistent and reproducible evaluation, it does not quantify the variability associated with different data partitions or training initializations. Consequently, confidence intervals, standard deviations, repeated experimental trials, and statistical significance analyses were not included in the present study. Future work will include repeated experiments using multiple random seeds, k-fold cross-validation, and external validation on independent multi-center datasets to further evaluate the robustness, reproducibility, statistical significance, and generalizability of the proposed framework.

IX. Conclusion and Future Work

This study introduced an automated multi-task deep learning framework based on the YOLOv11 architecture, designed for simultaneous detection of dental pathologies and arch-region classification in orthopantomograms (OPGs). To promote the system's efficiency while dealing with the extended Dental OPG Kennedy set comprising ten classes, the C2PSA attention modules coupled with the improved preprocessing techniques made the system produce outstanding results on all ten classes. The system reported an average mAP of 0.854, along with an F1-score of 0.961, exhibiting nearly flawless sensitivity while detecting well-demarcated edentulous areas. The upcoming work will focus on incorporating panoramic images from different centers into the database to boost the generality of the model instead of relying on domain specificity. Moreover, the application of the transformer modules, depending on self-supervised techniques, could advance the capture of higher-level semantics, especially in the identification of minor pathologies.

References

[1] D. A. Deenadayalan, K. Karthick, and N. Tamilarasu, “Orthopantomogram as a tool for morphometric study: Advancing dental diagnostics with a novel approach,” Int. J. Adv. Res. Med., vol. 6, no. 3, pp. 94–97, 2024, doi: 10.22271/27069567.2024.v6.i3b.573.

[2] S. Hegde, J. Gao, R. Vasa, and S. Cox, “Factors affecting interpretation of dental radiographs,” Dentomaxillofacial Radiol., vol. 52, no. 2, p. 20220279, Feb. 2023, doi: 10.1259/dmfr.20220279.

[3] S. AbuSalim, N. Zakaria, A. Saboor, A. Maqsood, S. J. Abdulkadir, and N. Mokhtar, “Deep Learning Methods for Tooth Detection and Classification in Various Dental Image Datasets: A Taxonomy and Future Directions,” Int. J. Recent Innov. Trends Comput. Commun., vol. 12, no. 2, pp. 69–77, Apr. 2024.

[4] S. AbuSalim, N. Zakaria, M. R. Islam, G. Kumar, N. Mokhtar, and S. J. Abdulkadir, “Analysis of Deep Learning Techniques for Dental Informatics: A Systematic Literature Review,” Healthcare, vol. 10, no. 10, p. 1892, Oct. 2022, doi: 10.3390/healthcare10101892.

[5] M. Ali, M. Irfan, T. Ali, C. R. Wei, and A. Akilimali, “Artificial intelligence in dental radiology: a narrative review,” Ann. Med. Surg., vol. 87, no. 4, pp. 2212–2217, Mar. 2025, doi: 10.1097/MS9.0000000000003127.

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” Apr. 23, 2020, arXiv: arXiv:2004.10934. doi: 10.48550/arXiv.2004.10934.

[7] Y. E. Almalki et al., “Deep Learning Models for Classification of Dental Diseases Using Orthopantomography X-ray OPG Images,” Sensors, vol. 22, no. 19, p. 7370, Sep. 2022, doi: 10.3390/s22197370.

[8] B. Beser et al., “YOLO-V5 based deep learning approach for tooth detection and segmentation on pediatric panoramic radiographs in mixed dentition,” BMC Med. Imaging, vol. 24, p. 172, Jul. 2024, doi: 10.1186/s12880-024-01338-w.

[9] S. AbuSalim, N. Zakaria, S. A. Mostafa, Y. K. Hooi, N. Mokhtar, and S. J. Abdulkadir, “Multi-Granularity Tooth Analysis via Faster Region-Convolutional Neural Networks for Effective Tooth Detection and Classification,” Int. J. Adv. Comput. Sci. Appl. IJACSA, vol. 14, no. 6, Jun. 2023, doi: 10.14569/IJACSA.2023.0140678.

[10] A. F. Makarim, T. Karlita, R. Sigit, B. S. B. Dewantara, and A. Brahmanta, “Deep Learning Models for Dental Conditions Classification Using Intraoral Images,” JOIV Int. J. Inform. Vis., vol. 8, no. 3, pp. 1099–1108, Sep. 2024, doi: 10.62527/joiv.8.3.1914.

[11] A. R. Ismail and M. S. S. Taseen, “Deep Learning Approach for Dental Anomalies X-ray Imaging using YOLOv8,” Knowl. Eng. Data Sci., vol. 7, no. 2, pp. 164–175, Mar. 2025, doi: 10.17977/um018v7i22024p164-175.

[12] U. Şevik and O. Mutlu, “Detection of Dental Anomalies in Digital Panoramic Images Using YOLO: A Next Generation Approach Based on Single Stage Detection Models,” Diagnostics, vol. 15, no. 15, p. 1961, Jan. 2025, doi: 10.3390/diagnostics15151961.

[13] S. Sukegawa et al., “Evaluation of multi-task learning in deep learning-based positioning classification of mandibular third molars,” Sci. Rep., vol. 12, no. 1, p. 684, Jan. 2022, doi: 10.1038/s41598-021-04603-y.

[14] M. Prados-Privado, J. G. Villalón, C. H. Martínez-Martínez, and C. Ivorra, “Dental Images Recognition Technology and Applications: A Literature Review,” Appl. Sci., vol. 10, no. 8, p. 2856, Jan. 2020, doi: 10.3390/app10082856.

[15] Y. Xiong et al., “Simultaneous detection of dental caries and fissure sealant in intraoral photos by deep learning: a pilot study,” BMC Oral Health, vol. 24, no. 1, p. 553, May 2024, doi: 10.1186/s12903-024-04254-1.

[16] B. V, V. V. Giri, M. Dodia, P. R. Sanjaya, R. T. Arun, and S. Maity, “Machine Learning-Based Predictive Analysis of Dental Treatment Outcomes,” in 2025 Tenth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Nov. 2025, pp. 1–6. doi: 10.1109/ICONSTEM65670.2025.11374450.

[17] N. Pishghadam, R. Esmaeilyfard, and M. Paknahad, “Explainable deep learning for age and gender estimation in dental CBCT scans using attention mechanisms and multi task learning,” Sci. Rep., vol. 15, no. 1, p. 18070, May 2025, doi: 10.1038/s41598-025-03305-z.

[18] J. A. Nava-Pintor et al., “Context-Aware Caries Segmentation in Periapical Radiographs Using a Hybrid Multi-Task Learning Framework with Partial Annotations,” Appl. Sci., vol. 16, no. 1, p. 264, Jan. 2026, doi: 10.3390/app16010264.

[19] Z. Khurshid et al., “Multi-Regional deep learning models for identifying dental restorations and prosthesis in panoramic radiographs,” BMC Oral Health, vol. 25, p. 1750, Nov. 2025, doi: 10.1186/s12903-025-07138-0.

[20] K. Zhang, J. Wu, H. Chen, and P. Lyu, “An effective teeth recognition method using label tree with cascade network structure,” Comput. Med. Imaging Graph., vol. 68, pp. 61–70, Sep. 2018, doi: 10.1016/j.compmedimag.2018.07.001.

[21] “Dental OPG Kennedy Dataset.” Accessed: Oct. 30, 2025. [Online]. Available: https://www.kaggle.com/datasets/orvile/dental-opg-kennedy-dataset

[22] S. Abusalim, S. Mostafa, N. Zakaria, S. Jadid Abdulkadir, and N. Mokhtar, Data Augmentation on Intra-Oral Images Using Image Manipulation Techniques. 2022, p. 120. doi: 10.1109/ICDI57181.2022.10007158.

[23] C.-Y. Wang, H.-Y. M. Liao, I.-H. Yeh, Y.-H. Wu, P.-Y. Chen, and J.-W. Hsieh, “CSPNet: A New Backbone that can Enhance Learning Capability of CNN,” Nov. 27, 2019, arXiv: arXiv:1911.11929. doi: 10.48550/arXiv.1911.11929.

[24] R. Khanam and M. Hussain, “YOLOv11: An Overview of the Key Architectural Enhancements,” Oct. 23, 2024, arXiv: arXiv:2410.17725. doi: 10.48550/arXiv.2410.17725.

[25] H. Tekis, T. Zirek, and M. Tassoker, “Comparative analysis of YOLO variants for dental implant brand identification,” Sci. Rep., vol. 15, no. 1, p. 33304, Sep. 2025, doi: 10.1038/s41598-025-15383-0.

[26] S. Mureșanu et al., “Automating Dental Condition Detection on Panoramic Radiographs: Challenges, Pitfalls, and Opportunities,” Diagnostics, vol. 14, no. 20, p. 2336, Jan. 2024, doi: 10.3390/diagnostics14202336.

[27] R. B. Peker and C. O. Kurtoglu, “Evaluation of the Performance of a YOLOv10-Based Deep Learning Model for Tooth Detection and Numbering on Panoramic Radiographs of Patients in the Mixed Dentition Period,” Diagnostics, vol. 15, no. 4, p. 405, Jan. 2025, doi: 10.3390/diagnostics15040405.

[28] Y. E. Almalki et al., “Deep Learning Models for Classification of Dental Diseases Using Orthopantomography X-ray OPG Images,” Sensors, vol. 22, no. 19, p. 7370, Jan. 2022, doi: 10.3390/s22197370.

[29] A. Adnan, M. Rizwan, H. Attaullah, S. Basheer, and M. Quasim, “Deep Architectural Classification of Dental Pathologies Using Orthopantomogram Imaging,” Comput. Mater. Contin., vol. 85, no. 3, pp. 5073–5091, 2025, doi: 10.32604/cmc.2025.068797.

[30] R. H. Putra et al., “Automated permanent tooth detection and numbering on panoramic radiograph using a deep learning approach,” Oral Surg. Oral Med. Oral Pathol. Oral Radiol., vol. 137, no. 5, pp. 537–544, May 2024, doi: 10.1016/j.oooo.2023.06.003.

[31] S. AbuSalim et al., “Multi-granularity tooth analysis via YOLO-based object detection models for effective tooth detection and classification,” IAES Int. J. Artif. Intell. IJ-AI, vol. 13, no. 2, pp. 2081–2092, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp2081-2092.

[32] S. AbuSalim, N. Zakaria, S. A. Mostafa, A. Maqsood, S. J. Abdulkadir, and N. Mokhtar, “Hierarchical deep learning model for tooth classification in intra-oral dental photographs,” Neural Comput. Appl., vol. 37, no. 32, pp. 26659–26688, Nov. 2025, doi: 10.1007/s00521-025-11639-y.