Robust Inspired Detection Scheme Based on Learning Model for Multi-Class Identification in Energy Radiographic Images

Gorrepati Rajani Reddy
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India.
Email: rajinireddy@kluniversity.in

Misbah Bibi
Department of Computer Engineering, Jeju National University, Jeju, 63243, Republic of Korea.
Email: misbahbibi@stu.jejunu.ac.kr

Do-Hyeun Kim
Department of Computer Engineering, Jeju National University, Jeju 63243, Republic of Korea
Email: kimdh@jejunu.ac.kr,

Abstract **- House object Accurate localization and classification of dental structures in the radiographic images are crucial for both automatic dental diagnosis and treatment planning. This paper proposes a deep learning-based tooth detection framework that integrates a multi-stage architecture, comprising a backbone for hierarchical feature extraction, a neck module for multi-scale feature refinement, and a detection head for bounding box regression and class prediction. The backbone employs a series of convolution-batch normalization-activation blocks to capture low- and mid-level anatomical patterns from input dental images. The neck module fuses multi-scale features using upsampling and concatenation operations, enhancing contextual representation. Then, the final predictions, including tooth class probabilities, objectness scores, and optimized bounding boxes, are generated in the detection head. Experimental evaluation on multiple tooth categories reveals powerful detection performance, with an overall precision of 0.786, an overall recall of 0.848, mAP@50 of 0.881, and mAP@50–95 of 0.656. Class-wise results display excellent performances for Central Incisors, Canines, and Lateral Incisors, having mAP@50–95 values of 0.845, 0.808, and 0.823, respectively, while comparatively low scores have been observed for some classes like the 2nd Molar. In general, the proposed model effectively balances accuracy and robustness across diverse tooth types.**

Index Terms - Dental image analysis, Multi-scale feature fusion, Radiographic imaging, Bounding box regression, Automated dental diagnosis.

I. INTRODUCTION

The integration of artificial intelligence and deep learning into dental imaging has revolutionized traditional diagnostic workflows, which are today faster, more consistent, and highly accurate in clinical decision-making. Initial dental radiograph analysis studies focused on classical image processing and neural network-based techniques as a starting point for automated diagnosis systems. For example, Silva et al. (2018) presented one of the earliest comprehensive studies on automatic tooth segmentation in X-ray images, pointing to the main trends, benchmarking strategies, and the challenge of variability in datasets and anatomical complexity [1].

This paper laid down the foundation for the importance of standard datasets and robust feature extraction methods for reliable dental image interpretation. Advancement of machine learning has greatly increased the precision and efficiency of dental diagnostics. Geetha et al. (2020) demonstrated the application of back-propagation neural networks in detecting dental caries from digital radiographs, proving that early-stage lesions can be effectively detected using a supervised learning approach [4]. Similarly, optimization-based deep learning techniques have also been used to improve the accuracy of complex dental image recognition. Mahdi et al. (2020) proposed an optimizationenhanced deep learning method for recognizing teeth in panoramic radiographs; achieving better performance, the refinement of network parameters for diverse anatomical variations significantly contributed to such results [5].

Recent breakthroughs in architectures for deep learning have further enhanced the capabilities of AI dental diagnostics. Chen et al. (2019) introduced a framework using object detection for automatic teeth detection and numbering in periapical radiographs, with a demonstration of how CNNs can identify and classify individual teeth within varying orientations, occlusions, and image qualities [3]. These advances point to the importance of deep learning in establishing automated, reproducible, and clinician-assistive dental diagnostic systems. Over the past years, there has been a fast pace toward incorporating AI into a wide range of dental applications, from simple image analysis to complete diagnosis of diseases. A comprehensive review, starting from the basics and finishing with the most recent developments in the field, has been performed by Ossowska et al. (2022) [2]. All the reviewed material indicates an increased trend in using machine learning for caries detection, orthodontic assessment,

periodontal evaluation, and implant planning. It is mentioned that AI increases not only diagnostic precision but also aids in workflow optimization and reduces the workload for clinicians, particularly in radiographic interpretation.

II. Literature Survey

Deep learning has become an important tool in dental image analysis for automatic diagnosis, segmentation, and classification across a wide range of imaging modalities. Early efforts in the subject primarily focused on using CNNs with challenging radiographic datasets. Lee et al., in 2020, showed one of the first applications of deep learning in the diagnosis of cystic lesions in panoramic and CBCT images [6]. Their work manifested the great capability of neural networks to outperform traditional diagnostic methods, especially when the pathological region had ambiguous borders. In the continuation of deep learning development, the focus shifted to the automatic feature extraction and pathology detection in periapical radiographs. Khan et al., in 2021, proposed a CNN-based framework for the automatic detection of dental features and pathologies, which improved the accuracy and manual intervention in interpreting the periapical radiograph [7]. Their model efficiently detected structural dental features like pulp chambers and periodontal ligaments, and set the base for complex diagnostic tasks.

Building on object detection methodology, Cha et al. (2021) proposed an R-CNN to quantify peri-implant bone loss based on periapical radiographs [8]. This approach allowed for accurate localization of the peri-implant regions and automated bone level measurement, providing a useful tool to aid in implant maintenance protocols. Kim et al. (2021) employed deep learning to predict paresthesia after third molar extraction [9]. Their study demonstrated that neural networks may interpret the radiographic patterns indicative of nerve proximity and extraction complexity. These studies together highlighted the potential of deep learning to provide patient-specific risk estimates and to aid in clinical decision-making. Studies were conducted regarding the classification and detection tasks using panoramic radiographs. Muramatsu et al. (2021) enhanced the detection and classification of teeth by including multisized input data to overcome the problems caused by variation in scale and overlapping of anatomical structures [10]. The results indicated that their automatic dental charting was an effective way of combining panoramic imaging with deep learning for routine clinical documentation.

Recent works in 2024 and beyond mark the shift toward dedicated and large-scale models for tooth classification and segmentation. Yilmaz et al. (2024) developed a deep learning-based method that is dedicated to tooth classification in panoramic radiographs, yielding high accuracy for a wide range of dental patterns [11]. For pediatric dental image analysis, Beser et al. (2024) presented a YOLOv5-based tooth detection and segmentation method in mixed dentition images that can provide clinicians with a fast and accurate diagnosis in a clinical environment with a child population [12]. Furthermore, Hu et al. (2024) expanded the AI application area into 3D dental imaging by presenting a fully automatic system for the segmentation and fine classification of mixed dentition from CBCT images with superior performance in volumetric analyses compared to conventional 2D approaches [13].Very recently, Balel et al. (2025) investigated AI applications in implant dentistry through the development of an AI-based framework for detecting and numbering dental implants in panoramic radiographs [14]. Their approach yielded highly accurate results both in detecting implant positions and annotations, thus providing clinical support in postoperative assessment and long-term implant monitoring.

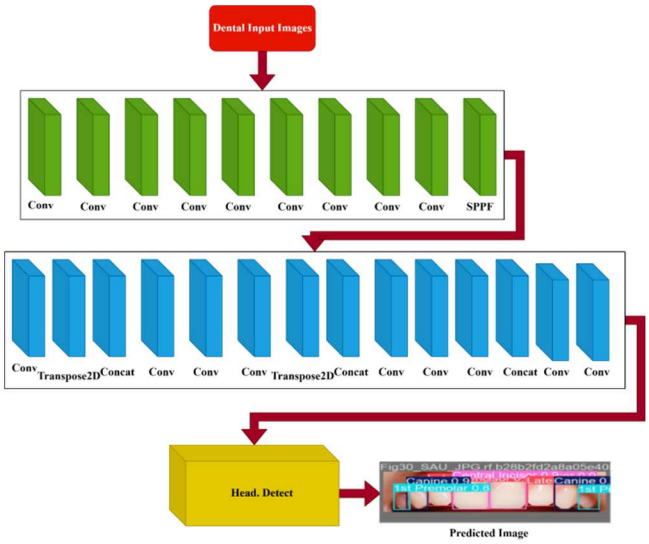

Fig.1 Proposed model architecture

III. Proposed Model

In Fig.1., the input stage begins with dental X-ray images represented as a tensor H×W×3, where H, W, and 3 denote the height, width, and color channels of the image, respectively. Before feeding the image into the network, it is typically normalized to stabilize training and improve feature extraction. This normalization is performed by scaling pixel intensities to the range [0,1], ensuring consistent input distribution for the subsequent processing stages.

The backbone stage acts as the basic feature extractor in dental image detection, transforming the raw pixel-level information

into high-quality hierarchical representations suitable for deeper processing. Each block typically consists of a sequence of Convolution, Batch Normalization, Activation, which are known for optimizing gradient flow and reducing computational redundancy. The convolution operation represents the crucial component for learning spatial features such as tooth contours, enamel boundaries, root structures, and texture variations within dental radiographs. A convolution layer applies a set of learnable kernels K across the input feature map X to produce an output feature Y, computed as

$$Y_{i,j,k} = \sum_{m,n,p} X_{i+m,j+n,p} K_{m,n,p,k} \quad (1)$$

which aggregates information within a local neighborhood around each spatial location. However, simply applying convolution is insufficient for stable and efficient training, especially when processing large dental datasets. Therefore, the output of each convolution is normalized using Batch Normalization, which standardizes the activations by subtracting the mini-batch mean and dividing by the batch variance, expressed as

$$\hat{x} = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \quad (2)$$

$$y = \gamma \hat{x} + \beta \quad (3)$$

This step reduces internal covariate shift, accelerates convergence, and allows the model to use higher learning rates without diverging. Following normalization, the activations pass through the SiLU (Sigmoid Linear Unit) non-linear function, defined as

$$Silu(x) = \frac{x}{1+e^{-x}} \quad (4)$$

Which preserves small gradients, avoids harsh saturation, and enhances the model's ability to extract subtle details such as narrow gaps between teeth or fine anatomical structures. Together, these Conv–BN–SiLU blocks form a powerful backbone capable of capturing both low-level (edges, intensity gradients) and mid-level (tooth shapes, root outlines) features, ultimately producing a rich and discriminative feature map that feeds into the subsequent neck and detection head for accurate dental structure localization and classification.

The Neck stage is critical for refinement and fusing the features extracted by the backbone: while the backbone focuses on progressively learning low-level and mid-level patterns, the neck needs to refine, fuse, and strengthen multi-scale feature representations, which are, in turn, expected to be of paramount importance for detecting variously sized objects in dental images-indeed, incisors, canines, premolars, and molars vary greatly in size and shape. The main principle here is to enable the model to combine deep semantic features-from later layers-with fine spatial featuresfrom early layers-so the detector can accurately localize small or partially visible dental structures. The process begins with Feature Pyramid Fusion, where feature maps from different scales, typically denoted as f_i (higher-resolution feature) and f_{i+1} (lower-resolution but semantically richer), are fused. Since f_{i+1} has smaller spatial dimensions, it must be upsampled before fusion. This is mathematically expressed as

$$F_{fused} = Concat\big(f_i, Up(f_{i+1})\big) \quad (5)$$

where Up denotes an interpolation operation, such as bilinear interpolation, used to expand the spatial resolution of the deeper feature map without learning additional parameters. Bilinear upsampling is defined as

$$Up(X) = Bilinear(X) \quad (6)$$

Which computes each new pixel value as a weighted average of the nearest four pixels, ensuring smooth and artifact-free resizing. After the feature maps are concatenated, the combined representation contains detailed spatial patterns from shallow layers and strong semantic information from deep layers, making it more suitable for identifying dental structures that may vary in orientation, scale, or contrast. This fused representation is then passed through a sequence of convolutional refinement layers represented as

$$F_{out} = \emptyset\big(F_{fused}\big) \quad (7)$$

this neck module ensures that the detector receives balanced, multi-scale, information-rich features, enabling it to precisely localize small dental structures, differentiate between closely positioned teeth, and maintain robustness across varying imaging conditions. The careful blending of coarse high-level features and fine low-level features makes the neck an essential bridge between the backbone and the prediction head in any modern deep learning based dental image detection system. The detection head is responsible for generating the final predictions, which include class probabilities, the bounding box coordinates (x,y), width w, height h, and the objectness score that indicates whether a tooth exists in the predicted region.

IV. Experiments

A.Dataset

The distribution of dental tooth classes based on the number of annotated instances available for each category is shown in Fig.2. It includes seven major tooth types: 1st Molar, 1st Premolar, 2nd Molar, 2nd Premolar, Canine, Central Incisor, and Lateral Incisor. The frequencies show noticeable variation across classes, indicating class imbalance within the dataset. The highest number of instances appears in the 1st Premolar, Canine, Central Incisor, and Lateral Incisor categories, each with close to or above 1800 samples, suggesting strong

representation of these teeth. The 1st Molar and 2nd Premolar classes also have substantial counts, with around 1500–1600 instances. However, the 2nd Molar class has significantly fewer samplesroughly 750making it the least represented category. This imbalance highlights the need for careful model training strategies, such as weighted loss functions or data augmentation, to ensure that underrepresented classes like the 2nd Molar are learned effectively by deep learning models.

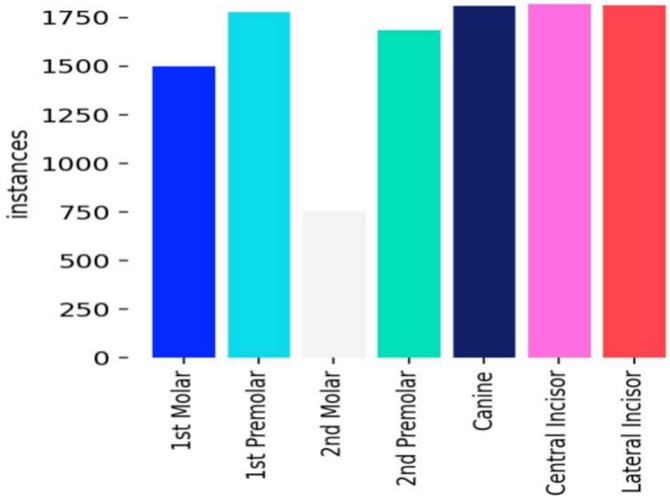

Fig.2. Dataset

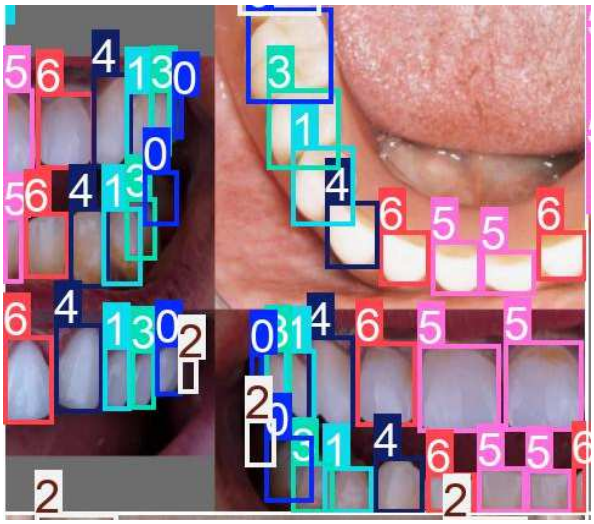

Fig.3. Mosaic Augmented image

In mosaic augmentation, four different images are combined into one composite image, thereby allowing the model to learn varied contexts, lighting conditions, and spatial arrangements in one training instance. In this case, the quadrants each show different teeth with bounding box annotations of several classes of teeth, differentiated by color and class IDs is shown in Fig.3. The created scene is highly complicated, with objects varying in scale, orientation, and position. This augmentation increases the complexity significantly since the model is exposed to more difficult configurations, enhancing its generalization capability to real-world dental radiographs or intraoral images. Mosaic augmentation increases the number of bounding boxes available per image and helps the model in learning good feature interaction, specifically for small, partially occluded teeth.

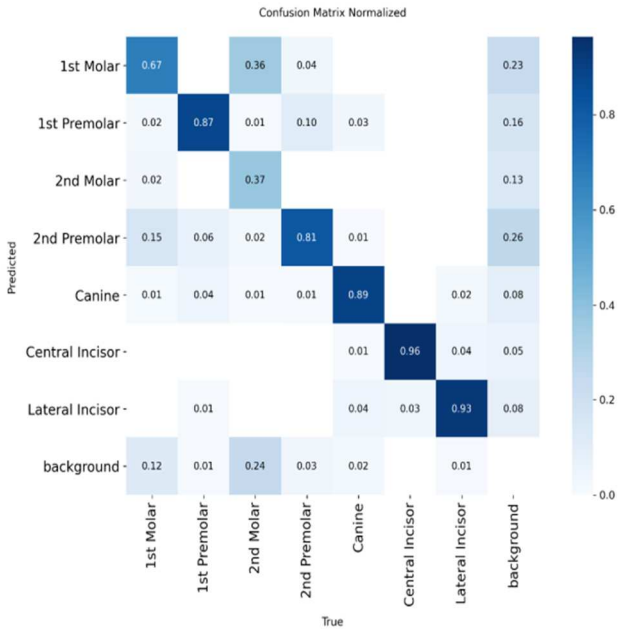

Fig.4. Confusion Matrix

The normalized confusion matrix represents the performance of the dental tooth classification model for eight classes, including the background category is shown in Fig.4. The values on the diagonal indicate correct predictions and are particularly strong for several classes, such as Central Incisor (0.96), Lateral Incisor (0.93), Canine (0.89), and 1st Premolar (0.87), indicating that the model steadily identifies these types of teeth. For such classes as 1st Molar (0.67) and 2nd Premolar (0.81), moderate performance is observed with noticeable misclassifications, especially for similar classes. The 2nd Molar class has the lowest accuracy, with only 0.37 correctly classified and a substantial confusion with background, 1st Premolar, and 1st Molar, which reflects class imbalance and visual similarity issues present in the dataset. Misclassifications into the background category also emerge for several classes, especially for 2nd Molar and 1st Molar, which indicates detection challenges for partially visible or small teeth. On the whole, this matrix points to great performance for most categories of teeth but shows that rare or visually subtle classes need further refinement.

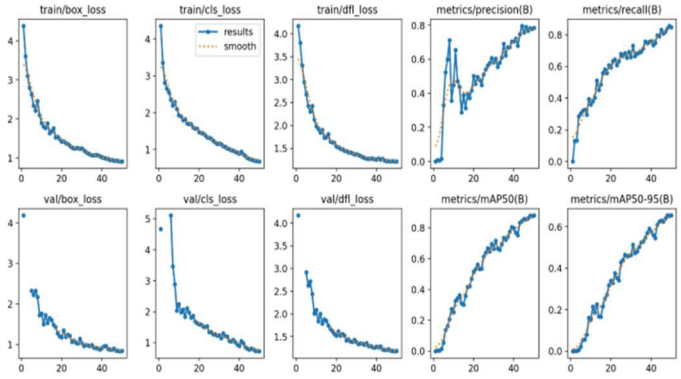

Fig.5. Performance metrics

The complete training and validation performance curves of the tooth classification and detection model across 50 epochs, showing how the model learns to differentiate between different tooth categories is shown in Fig.5. The losses in training, which are box loss, classification loss, and distribution focal loss, consistently decrease over time. This goes to show that the model is getting better at bounding-box localization and class label prediction for various teeth, such as molars, premolars, canines, and incisors. Similarly, there is a downwards trend for these losses in validation, indicating good generalization and low overfitting. On the right, metrics such as precision, recall, mAP50, and mAP50–95 steadily increase throughout training. Precision and recall rising together reflect improved correct object Identification and reduced false detections of tooth classes, even those with similar morphology. The mAP curves show robust improvement, demonstrating that the model becomes increasingly able to detect teeth more accurately across different IoU thresholds. In summary, these curves confirm that the model learns effectively from the dataset and becomes progressively more reliable at distinguishing between different tooth types in intraoral images.

TABLE 1 Model parameters

Layers	Parameters	GFLOPs	Preprocess Time	Inference Time	Postprocess Time
120	172,986,693	608.3	0.1 ms	3.5 ms	1.0 ms

Table 1 summarizes the computational characteristics and efficiency of the YOLOv6x model. The overall architecture comprises 120 layers with 172,986,693 trainable parameters, which illustrates a deep and highly expressive network capable of learning complex dental features. This model requires 608.3 GFLOPs, reflecting its high computational demand during forward propagation. Despite this big size, the processing times per image remain efficient; for example, pre-processing takes 0.1 ms, inference requires 3.5 ms to generate predictions, and post-processing, which includes operations such as non-maximum suppression, takes 1.0 ms.

TABLE 2 Performance Metrics

Class	Precision (P)	Recall (R)	mAP@50	mAP@50–95
All Classes	0.786	0.848	0.881	0.656
1st Molar	0.68	0.771	0.782	0.484
1st Premolar	0.803	0.941	0.958	0.717
2nd Molar	0.644	0.461	0.606	0.324
2nd Premolar	0.684	0.899	0.89	0.588
Canine	0.883	0.922	0.97	0.808
Central Incisor	0.921	0.978	0.986	0.845
Lateral Incisor	0.885	0.963	0.977	0.823

Table 2 presents the detection performance of the model across different tooth classes using key evaluation metrics such as Precision, Recall, mAP@50, and mAP@50–95. Overall performance for all classes is strong, with a Precision of 0.786, Recall of 0.848, and an mAP@50 of 0.881, indicating reliable detection capability. Individual class results show variations depending on tooth type. Central Incisors and Canines achieve the highest performance, with Precision values above 0.88 and mAP@50–95 scores of 0.845 and 0.808, respectively, demonstrating excellent localization and classification accuracy. Similarly, Lateral Incisors and 1st Premolars exhibit strong results, reflecting effective detection of these tooth structures. In contrast, performance for the 2nd Molar class is comparatively lower, with a Recall of 0.461 and mAP@50–95 of 0.324, indicating difficulty in detecting these teeth likely due to occlusion, anatomical variations, or lower visibility in radiographs.

Fig.6. Validated Energy Constraint Radiographic image

The predicted image provides an output of the tooth-classification model, where numerous teeth have been detected and labeled along with their respective classes and confidence scores is shown in Fig.6. Each colored bounding box in the image represents a type of tooth prediction, such as 2nd Premolar, 2nd Molar, Central Incisor, or Lateral Incisor, with its numeric value showing the model's confidence in that respective prediction. The overlapping boxes, therefore, indicate that the model has identified several possible areas of teeth but seems to be doing poorly in areas where visual similarities among adjacent teeth are prominent, hence having multiple predictions at one location. Some lower confidence scores, like 0.30 for 2nd Molar and 0.38 for 2nd Premolar, are indicative of uncertainty, mostly caused by the subtleness of the anatomical difference or the small number of training samples for the classes concerned. On the other hand, higher confidence predictions (e.g., 0.9+) show strong model reliability for well-represented or visually distinctive categories.

V. Conclusion

The proposed multi-stage dental detection framework demonstrates that combining hierarchical feature extraction, multi-scale feature refinement, and an optimized detection head can effectively identify and classify diverse tooth structures in radiographic images. The experimental results confirm strong overall performance, particularly for clearly distinguishable classes such as Central Incisors, Lateral Incisors, and Canines, while also highlighting areas requiring further improvement, such as the detection of 2nd Molars. Despite these challenges, the model consistently achieves high precision, recall, and mAP values, validating its reliability for automated dental analysis. Future research can focus on enhancing performance for low-visibility teeth and integrating additional contextual learning techniques to further improve robustness.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. RS-2024-00405278), and this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT)(No. RS-2024-00346238), Any correspondence related to this paper should be addressed to DoHyeun Kim

References

[1] Silva G., Oliveira L., Pithon M. Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. Expert Syst. Appl. 2018; 107:15–31. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Ossowska A., Kusiak A., Świetlik D. Artificial Intelligence in DentistryNarrative Review. Int. J. Environ. Res. Public Health. 2022;19:3449.

[3] Bibi, M., Faseeh, M., Khan, A. N., Rizwan, A., Khan, Q. W., Ahmad, R., & Kim, D. H. (2025). A Unified Approach for Object Detection and Depth Map based Distance Estimation in Security and Surveillance Systems. *IEEE Access*.

[4] Geetha V., Aprameya K.S., Hinduja D.M. Dental caries diagnosis in digital radiographs using back-propagation neural network. Health Inf. Sci. Syst. 2020;8:8.

[5] Mahdi, F.P., Motoki, K. & Kobashi, S. Optimization technique combined with deep learning method for teeth recognition in dental panoramic radiographs. Scientific Reports, 10(1), 19261, 2020.

[6] Lee, J.H., Kim, D.H. & Jeong, S.N. Diagnosis of cystic lesions using panoramic and cone beam computed tomographic images based on deep learning neural network. Oral Diseases, 26(1), 152–158, 2020.

[7] Khan H.A., Haider M.A., Ansari H.A., Ishaq H., Kiyani A., Sohail K., Muhammad M., Khurram S.A. Automated feature detection in dental periapical radiographs by using deep learning. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. 2021;131:711–720.

[8] Cha J.-Y., Yoon H.-I., Yeo I.-S., Huh K.-H., Han J.-S. Peri-Implant Bone Loss Measurement Using a Region-Based Convolutional Neural Network on Dental Periapical Radiographs. J. Clin. Med. 2021;10:1009.

[9] Bibi, M., Khan, A. N., Faseeh, M., Khan, Q. W., & Ahmad, R. (2024). A synthetic data generation approach with dynamic camera poses for long-range object detection in AI applications. *IEEE Access*, *12*, 194505-194520.

[10] Muramatsu C., Morishita T., Takahashi R., Hayashi T., Nishiyama W., Ariji Y., Zhou X., Hara T., Katsumata A., Ariji E., et al. Tooth detection and classification on panoramic radiographs for automatic dental chart filing: Improved classification by multi-sized input data. Oral Radiol. 2021;37:13–19.

[11] Yilmaz, S., Tasyurek, M., Amuk, M., Celik, M., Canger, E.M. Developing deep learning methods for classification of teeth in dental panoramic radiography. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 138(1), 118–127, 2024.

[12] Beser, B., Reis, T., Berber, M.N., Topaloglu, E., Gungor, E., Kılıc, M.C., Duman, S., Çelik, Ö., Kuran, A. & Bayrakdar, I.S. YOLO-V5 based deep learning approach for tooth detection and segmentation on pediatric panoramic radiographs in mixed dentition. BMC Medical Imaging, 24(1), 172, 2024.

[13] Hu, Y., Liu, C., Liu, W., Xiong, Y., Zeng, W., Chen, J., Li, X., Guo, J. & Tang, W. Fully automated method for three-dimensional segmentation and fine classification of mixed dentition in cone-beam CT using deep learning. Journal of Dentistry, 151, 105398, 2024.

[14] Balel, Y., Sağtaş, K., Teke, F. & Kurt, M.A. Artificial Intelligence-Based Detection and Numbering of Dental Implants on Panoramic Radiographs. Clinical Implant Dentistry & Related Research, 27(1), e70000, 2025.