IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3591~3602
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3591-3602

YOLO-based deep learning for tooth detection, segmentation, and numbering in panoramic radiographs

Sri Oktamuliani[1], Luqyana Mahdiyah[1], Haritsul Haq[1], Nesa Perdana Putri[2], Wulandani Liza Putri[3]
[1]Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Andalas, Padang, Indonesia
[2]Division of Dentistry, Dental Clinic, Dr. Muhammad Zein Regional General Hospital, Painan, Indonesia
[3]Department of Orthodontics, Faculty of Dentistry, Universitas Andalas, Padang, Indonesia

Article Info

Article history:

Received Oct 14, 2025
Revised May 13, 2026
Accepted May 24, 2026

Keywords:

Deep learning
Dental imaging
Panoramic radiograph
Segmentation
Tooth detection
Tooth numbering
You only look once

ABSTRACT

Panoramic dental radiographs are essential for diagnosis and treatment planning, but are difficult to interpret due to overlapping anatomical structures and limited image quality. Automating tooth detection, segmentation, and numbering can enhance clinical efficiency and reduce observer variability. This study developed a you only look once version 8 (YOLOv8)-based deep learning model for automatic detection and segmentation of teeth in panoramic radiographs. A dataset of 302 images containing 9,009 annotated teeth was divided into 70% training, 20% validation, and 10% testing sets. Teeth were labeled using the Federation Dentaire Internationale (FDI) two-digit numbering system with bounding boxes and segmentation masks. Model performance was assessed using precision, recall, F1-score, and mean average precision (mAP) at intersection over union (IoU) thresholds of 0.5 and 0.5–0.95. The model achieved bounding box precision, recall, and F1-score of 0.9075, 0.9352, and 0.9212, respectively, and segmentation scores of 0.9078, 0.9344, and 0.9209. Bounding box mAP@0.5 reached 0.9510 and mask mAP@0.5 was 0.9504, while stricter thresholds lowered performance to 0.7763 and 0.6896. Class-wise analysis showed high accuracy overall but reduced performance on posterior teeth due to anatomical overlap and peripheral image quality. YOLOv8 enables fast, accurate, and robust tooth detection and segmentation, supporting real-time computer-aided diagnosis in orthodontics, prosthodontics, and forensic dentistry.

Corresponding Author:

Sri Oktamuliani
Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Andalas
Limau Manis Campus, Padang 25163, Indonesia
Email: srioktamuliani@sci.unand.ac.id

1. INTRODUCTION

Panoramic radiography is an extraoral imaging technique that captures a comprehensive view of the maxillofacial region, including the mandible, maxilla, and their supporting structures. It is widely used in dentistry because it is simple to perform, provides coverage of all teeth and jaws, and delivers a relatively low radiation dose. Clinically, panoramic radiographs are employed to diagnose jaw fractures, evaluate temporomandibular joint (TMJ) symmetry, examine tooth root morphology, detect cysts and lesions, and assess dentition during growth and development [1]–[3]. Almost all dental treatments require complementary radiographic examinations to support accurate diagnosis and optimal treatment planning [4], [5].

Despite these advantages, panoramic radiographs suffer from inherent limitations related to the physics of imaging. Compared with intraoral radiography, panoramic images have lower spatial resolution, are subject to geometric distortion, and often display overlapping anatomical structures [6]–[8]. These factors make interpretation complex and time-consuming, and they may contribute to variability between clinicians. As a result, there has been growing interest in developing computer-aided diagnostic systems that can automatically interpret panoramic images and support decision-making in dental practice [9]–[11].

Artificial intelligence (AI) technology has brought significant advances in the field of medical image analysis. AI can help accelerate and enhance the image identification and analysis process. Machine learning is a branch of science that studies algorithms and statistical models, enabling computer systems to perform specific tasks without requiring explicit instructions. This technology relies on patterns and conclusions derived from data. To obtain these patterns and conclusions, machine learning algorithms build mathematical models based on sample data known as training data [12]. Machine learning algorithms are divided into three types. First, supervised learning is a machine learning technique in which a model is trained to make predictions based on labeled training data. Second, unsupervised learning studies patterns and relationships in data without labels. Third, reinforcement learning is a machine learning model that makes decisions through trial and error.

Recent advances in AI have led to significant progress in medical image analysis. Machine learning enables computers to learn from data and make predictions without being explicitly programmed [13]–[15]. Machine learning techniques are typically categorized as supervised, unsupervised, or reinforcement learning, depending on the type of feedback provided during training [16]–[18]. Among these approaches, deep learning, based on artificial neural networks has become a dominant paradigm because of its ability to model complex, nonlinear patterns in large datasets [16], [19]. Deep learning methods have achieved state-of-the-art results in tasks such as image classification, segmentation, and object detection, making them particularly suitable for medical imaging applications.

Image segmentation, the process of delineating regions of interest within an image is a key step in dental image analysis. Manual segmentation, while accurate, is labor-intensive, requires expert knowledge, and is prone to inter- and intra-operator variability. Automatic segmentation methods have been developed to overcome these challenges and improve reproducibility. For example, several studies have successfully segmented teeth and jaw structures in panoramic radiographs using traditional image processing techniques such as thresholding and morphological operations [20]–[23]. However, these approaches are often time-consuming, sensitive to variations in image quality, and still require expert supervision, which limits their scalability in clinical practice. This has motivated a shift toward deep learning-based segmentation approaches, which can provide fast, robust, and fully automated results.

With the advent of deep learning, convolutional neural networks (CNNs) have become the preferred approach for medical image analysis, including segmentation and classification. Architectures such as U-Net and its variants have achieved excellent performance in segmenting organs and lesions in radiological images. In dentistry, deep learning has been applied to tasks such as tooth numbering, caries detection, and mandibular canal segmentation. A recent study proposed a deep learning model for automatic tooth detection and numbering in panoramic radiographs, achieving high accuracy and demonstrating the potential of AI-based systems to assist dental practitioners [24]. However, U-Net-based methods often require substantial computational resources and are not optimized for real-time applications.

Object detection models, particularly those based on the you only look once (YOLO) family, have gained attention for their speed and accuracy. YOLO, first introduced in 2025, redefined object detection by framing it as a single regression task, significantly reducing inference time [25]. Recent studies have applied YOLO variants to dental and medical imaging tasks. demonstrated the successful application of YOLO variants in dental and medical imaging tasks. YOLOv8 has been utilized for detecting and segmenting radiolucent mandibular lesions on panoramic radiographs, achieving mean detection accuracies of 0.918 without augmentation and 0.952 with augmentation, outperforming previous models trained on smaller datasets [26]. These findings highlight YOLOv8's capability to manage challenging cases involving low contrast and overlapping anatomical structures.

Outside dentistry, YOLOv8 has shown superior performance in various segmentation tasks. A comparative study between YOLOv5 and YOLOv8 for corrosion segmentation on metal surfaces and found that YOLOv8 produced more accurate results, particularly on complex and irregular patterns [27]. Similarly, YOLOv8 applied to plant leaf disease detection using 2,850 annotated images across 19 classes achieved 98% precision and a 97% F1-score, further confirming its versatility and high performance across different domains [28].

Despite these promising results, relatively few studies have implemented YOLO for combined tooth detection, segmentation, and numbering in panoramic radiographs. Most prior work has focused on detection or landmark identification alone, and challenges remain in segmenting posterior teeth due to reduced contrast and anatomical overlap. Furthermore, the integration of standardized numbering systems, such as the Federation Dentaire Internationale (FDI) two-digit system, into deep learning pipelines remains limited, yet is essential for clinical adoption. To address these gaps, this study proposes a YOLOv8m-seg model for the automated detection, segmentation, and numbering of teeth in panoramic dental radiographs. A dataset of 302 radiographs comprising 9,009 annotated teeth was used, with annotations based on FDI two-digit system.

2. METHOD

2.1. Dataset preparation

This study utilized panoramic dental radiographs collected from the Dental and Oral Hospital Universitas Andalas (RSGM Unand). A total of 302 images were included after screening for quality and completeness. Imaging parameters were standardized at 80 kV tube voltage, 9.0 mA tube current, and 14.0 s exposure time, with a measured dose of 157.2 mGy/cm^2. Images with severe motion artifacts, missing anatomical structures, or inadequate exposure were excluded to ensure a consistent dataset. All images were initially stored in DICOM format, which contains both raw pixel data and acquisition metadata such as exposure parameters. For model training, the DICOM files were exported into bitmap image file (BMP) format and resized to 640×640 pixels to meet YOLOv8 input requirements. Histogram equalization and intensity normalization were applied to improve image contrast and enhance the visibility of anatomical structures. Patient information was anonymized before processing.

2.2. Annotation and labeling

Prior to model training, all panoramic radiographs underwent a structured region of interest (ROI) annotation process using the Roboflow platform. Each tooth was manually labeled according to the FDI two-digit dental nomenclature system, where the first digit indicates the quadrant and the second digit indicates the tooth number within that quadrant as shown in Figure 1. For each tooth, both bounding boxes and polygonal segmentation masks were generated, enabling the model to perform instance segmentation rather than simple detection.

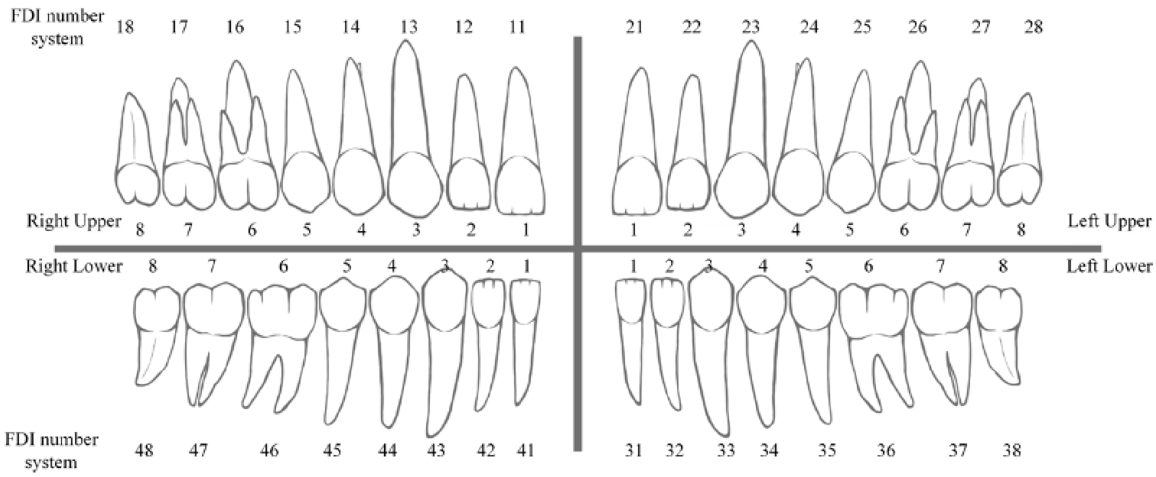

Figure 1. FDI two-digit tooth numbering system

All annotations were independently reviewed by one experienced dentist to ensure labeling accuracy and consistency. Any discrepancies between annotators were resolved through consensus discussions to minimize inter-operator variability. The final dataset comprised 9,009 annotated teeth across 302 panoramic radiographs, with a balanced representation of maxillary and mandibular teeth. This ensured that the model was exposed to sufficient anatomical variability across quadrants. An example of the annotation workflow and ROI labeling is illustrated in Figure 2. Each tooth was manually delineated with a polygonal segmentation mask and labeled according to the FDI two-digit system.

2.3. YOLOv8 work architecture

YOLOv8, the latest iteration of the YOLO object detection family, was selected for this study due to its high speed, accuracy, and robustness in handling small objects commonly present in dental panoramic radiographs. Unlike earlier YOLO versions, YOLOv8 employs anchor-free detection, a decoupled classification/regression head. It also uses a lightweight CSPDarknet-inspired backbone, resulting in improved generalization and reduced computational overhead.

The YOLOv8 architecture is divided into three main components: backbone, neck, and head as shown in Figure 3. The model consists of a backbone for feature extraction, a neck for multi-scale feature fusion, and a head for anchor-free detection and instance segmentation.

i) Backbone: responsible for extracting image features such as edges, textures, and shapes from the input image. It consists of convolutional layers (Conv), spatial pyramid dilated convolution (SPD-Conv), cross stage partial connections with fusion (C2f), and spatial pyramid pooling-fast (SPPF). These modules enhance the model's ability to capture multi-scale features, which is critical for detecting small and overlapping teeth.
ii) Neck: aggregates and refines features from multiple scales of the backbone using upsampling, concatenation, convolutional layers, and the global attention mechanism (GAM) to preserve spatial information and improve detection robustness.
iii) Head: produces the final outputs, including bounding boxes, object classification, and segmentation masks. YOLOv8 uses anchor-free head, which improves flexibility and localization accuracy [26], [29].

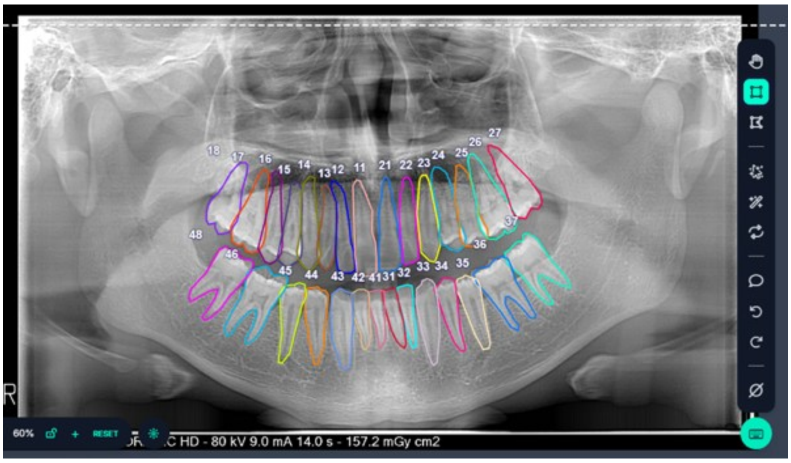

Figure 2. Example of ROI annotation process performed in Roboflow

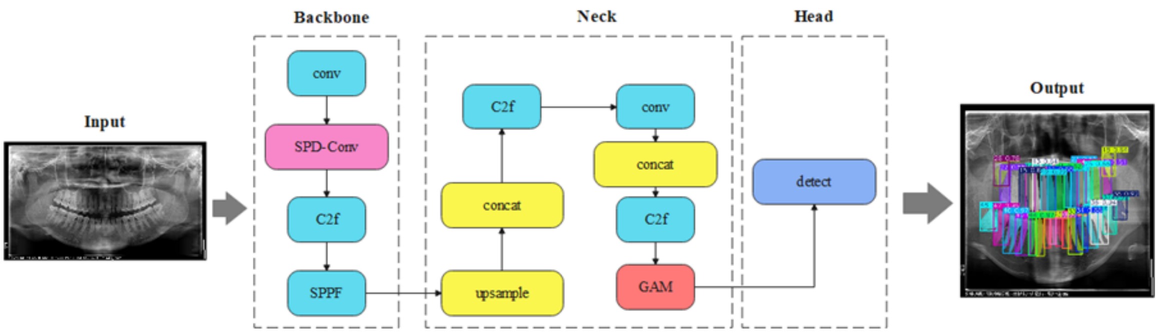

Figure 3. YOLOv8 work architecture

This study employed the YOLOv8-seg variant, which performs instance segmentation by predicting both bounding boxes and segmentation masks for each tooth. Among the available model sizes (nano, small, medium, large, and x-large), YOLOv8m-seg was chosen to balance inference speed with accuracy, making it suitable for potential real-time applications in clinical settings. Model training was conducted on Google Colab, leveraging graphics processing unit (GPU) acceleration for efficient computation. The training duration varied depending on the dataset size, the number of output classes, and the batch configuration. Hyperparameters, such as the learning rate, optimizer settings, and maximum number of epochs, were tuned iteratively to achieve optimal model performance.

2.4. Training setup

Model training was performed on a system equipped with an NVIDIA SMI GPU (driver version: 550.54.15, CUDA version: 12.4), 64 GB of RAM, running macOS Monterey, and additionally supported by Google Colab with GPU acceleration for enhanced computational efficiency. The dataset was randomly divided into 70% training (211 images), 20% validation (61 images), and 10% testing (30 images), ensuring a balanced distribution of tooth classes across all subsets. This split minimized sampling bias and maintained class-level representation in each partition. Notably, data augmentation techniques were not applied in this study to preserve the natural distribution of pixel intensities and anatomical variations inherent to panoramic radiographs. This decision was made to avoid introducing artificial artifacts that could distort tooth boundaries, which are already affected by imaging physics (e.g., scatter and geometric distortion).

Future work will evaluate augmentation strategies such as rotation, scaling, and contrast adjustments to further enhance model generalization. The model was trained for 300 epochs with a batch size of 16, using the default YOLOv8 composite loss function, which combines classification, bounding box regression, and mask losses to jointly optimize detection and segmentation performance. Early stopping was implemented based on validation mAP to prevent overfitting and reduce unnecessary computation.

2.5. Evaluation metrics

Model performance was assessed on the independent test set using standard object detection and segmentation metrics [30], [31]. The primary metric was mean average precision (mAP), calculated at an intersection over union (IoU) threshold of 0.5 (mAP@0.5) and averaged across thresholds from 0.5 to 0.95 in increments of 0.05 (mAP@0.5:0.95), providing both lenient and strict measures of detection accuracy. In addition, precision, defined as the proportion of correctly predicted teeth among all predictions, was used to evaluate the model's ability to avoid false positives, while recall, defined as the proportion of correctly detected teeth among all ground-truth annotations, reflected its ability to minimize false negatives. To capture the trade-off between precision and recall, the F1-score was calculated as the harmonic mean of the two, representing a balanced measure of overall model performance. Performance was reported both at the per-class level (for individual teeth) and as an average across all classes, enabling a comprehensive assessment of detection and segmentation quality at both granular and holistic levels.

3. RESULTS AND DISCUSSION

All experimental settings and implementation details are summarized in Table 1. These include model configuration, training hyperparameters, dataset splits, software versions, and hardware specifications. Table 1 is intended to support reproducibility without changing the meaning of the original statement.

Table 1. Experiment configuration and training details

Category	Configuration
Model architecture	YOLOv8 (ultralytics) – instance segmentation
YOLOv8 variant	YOLOv8m-seg
YOLOv8 version/package	Ultralytics YOLOv8 (v8.x, pip package ultralytics)
Framework	PyTorch
Python version	Python 3.12
Optimizer	AdamW (lr =0.00027, momentum =0.9)
Initial learning rate	0.01
Weight decay	0.0005
Warm-up strategy	Linear warm-up (YOLOv8 default)
Batch size	16
Input image size	640×640 pixels
Number of epochs	300
Early stopping	Enabled
Early stopping patience	50 epochs (based on validation mAP)
Loss function	YOLOv8 composite loss (classification + bounding box regression + mask loss)
Checkpointing	Best model saved based on highest validation mAP
Dataset split	70% training (211 images), 20% validation (61 images), 10% testing (30 images)
Class balancing	Balanced distribution of tooth classes across all subsets
Data augmentation	Not applied
Random seed	Fixed seed used for dataset splitting and training to ensure reproducibility
Annotation format	YOLO segmentation format
GPU	NVIDIA GPU (driver 550.54.15, CUDA 12.4)
System RAM	64 GB
Operating system	macOS Monterey
Cloud environment	Google Colab with GPU acceleration
Code and data availability	Training/validation/test splits, annotation export scripts, and configuration files are available upon reasonable request

3.1. Training and validation performance

The YOLOv8m-seg model was trained for 300 epochs using Google Colab with early stopping to prevent overfitting. Figure 4 shows the training and validation curves for box loss, classification loss, mask loss, precision, recall, and mAP metrics. Both training and validation losses decreased consistently, converging toward the final epochs, indicating that the model successfully minimized errors. Precision, recall, and mAP steadily increased during training, reaching a plateau near the final epochs. This demonstrates that the model achieved stable generalization without overfitting.

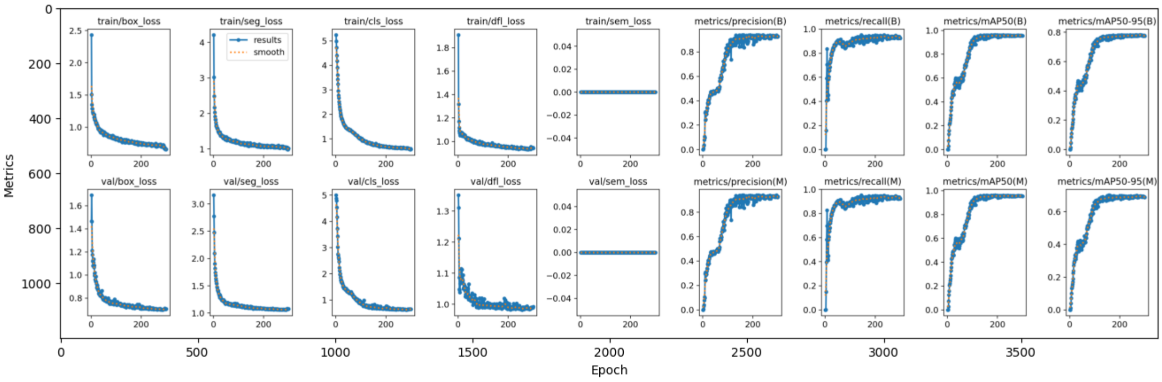

Figure 4. Training and validation loss functions and performance metrics of the YOLO-based segmentation model across 300 epochs

Figure 4 illustrates the training and validation behavior of the YOLO-based segmentation model over 300 epochs. The loss functions (box, segmentation, classification, and distribution focal loss) consistently decreased throughout the training process, with a rapid reduction during the first 50–80 epochs followed by gradual convergence. Importantly, the training and validation curves remain closely aligned, indicating good generalization without overfitting. The performance metrics show that bounding box precision and recall stabilized at approximately 0.93–0.95, while segmentation precision and recall reached similar levels with only minor fluctuations. The mAP curves further demonstrate strong model performance, with bounding box mAP@0.5 approaching 0.96 and mAP@0.5–0.95 stabilizing near 0.79. In contrast, mask mAP@0.5 reached ~0.96 but dropped to ~0.70 at stricter IoU thresholds (mAP@0.5–0.95). This gap highlights that while object localization is highly accurate, capturing precise segmentation boundaries is more challenging. Overall, the curves confirm stable convergence, efficient optimization, and excellent detection performance, with segmentation accuracy limited primarily by the precision of fine boundaries. The training and validation loss curves demonstrate a consistent and gradual decline across all components, including bounding box regression, segmentation, classification, and distribution focal loss. A rapid reduction in loss was observed during the initial 50–80 epochs, followed by a stable convergence phase around epoch 150. Importantly, the training and validation curves remained closely aligned throughout the process, indicating that the model achieved stable learning without signs of overfitting. This convergence behavior suggests that the training procedure was efficient and that the model reached an optimal balance between fitting the training data and generalizing to unseen cases.

The training behavior observed in this study is consistent with reports from prior works in dental imaging using deep learning models. Karaoglu *et al.* [24] applied a CNN-based model for automatic tooth detection and numbering on panoramic radiographs, reporting stable convergence after approximately 200 epochs, with precision and recall values of 0.90–0.92. Similarly, Rašić *et al.* [26] reported that YOLOv8 achieved a rapid reduction in training loss during the early epochs, with final mAP@0.5 values around 0.95 for mandibular lesion detection, aligning closely with the present bounding box results (mAP@0.5 ≈ 0.96). However, as in our study, they observed a performance drop when evaluated at stricter IoU thresholds (mAP@0.5–0.95, ≈0.75–0.80), highlighting the challenge of fine boundary delineation in complex radiographic images. Compared to earlier YOLOv8 studies on panoramic tooth landmark detection [32], which typically reported mAP@0.5 values between 0.90 and 0.93, the YOLOv8m-seg model used in this study demonstrates improved detection accuracy and more stable convergence. Nevertheless, the reduced segmentation performance at high IoU thresholds (~0.70 for mAP@0.5–0.95) remains a limitation, as also noted by Lee *et al.* [22] in U-Net-based segmentation of panoramic radiographs, where boundary ambiguity in posterior teeth limited fine-grained accuracy.

3.2. Quantitative performance metrics

The overall evaluation metrics confirmed the robustness of the YOLO-based segmentation model as shown in Table 2. Bounding box precision and recall achieved values of 0.9075 and 0.9352, respectively, while segmentation precision and recall reached 0.9078 and 0.9344. The corresponding F1-scores were 0.9212 (bounding box) and 0.9209 (mask), reflecting a strong balance between precision and recall for both detection and segmentation tasks.

The mAP provided further insights into model performance. At an IoU threshold of 0.5, the bounding box mAP was 0.9510, while the mask mAP reached 0.9504, confirming excellent overall detection and segmentation capability. When evaluated across stricter thresholds (mAP@0.5:0.95), performance declined to 0.7763 for bounding boxes and 0.6896 for masks, indicating that precise boundary alignment remains more challenging for segmentation compared to object detection.

Table 2. Comparison of deep learning approaches for dental image analysis

Study	Method	Dataset (images)	Task	Reported Performance	Remarks
Lee *et al.* [22]	CNN (FCN-based)	1,000 panoramics	Tooth segmentation	mIoU ≈ 0.82	Effective but computationally heavy, limited real-time use
Kim *et al.* [11]	CNN + heuristic	500 panoramics	Tooth detection and numbering	Accuracy = 0.90	Relied partly on heuristic rules, less robust to image variation
Karaoglu *et al.* [24]	U-Net + heuristic	650 panoramics	Tooth detection and numbering	Accuracy ≈ 0.94	High performance but requires complex preprocessing
Rašić *et al.* [26]	YOLOv8	2,000 panoramics	Lesion detection	Precision = 0.918 (no aug), 0.952 (aug)	Demonstrated YOLOv8 strength in dental tasks
Tafala *et al.* [32]	YOLOv8	400 cephalograms	Landmark detection	mAP@0.5 = 0.95	Confirmed high accuracy in dental landmark tasks
This study	YOLOv8m-seg	302 panoramics (9,009 annotated teeth)	Tooth detection, segmentation, numbering	Bounding box: precision 0.9075, recall 0.9352, F1-score 0.9212, mAP@0.5=0.9510, mAP@0.5:0.95=0.7763; mask: precision 0.9078, Recall 0.9344, F1 0.9209, mAP@0.5=0.9504, mAP@0.5:0.95=0.6896	Outperforms most prior works in combined detection + segmentation; segmentation mAP@0.5:0.95 lower (0.6896), especially in posterior teeth

Class-wise analysis further highlighted that certain tooth, such as 45, 48, and 38, consistently achieved higher confidence and accuracy, whereas posterior teeth, including 18, 24, 26, and 28, exhibited slightly lower performance. This variation may be attributed to reduced visibility at the periphery of panoramic images, anatomical overlap, or limited sample representation of these teeth in the training dataset. The bounding box mAP@0.5 (0.9510) is comparable to the performance reported [32], who achieved an mAP of 0.95 using YOLOv8 for cephalometric landmark detection. However, the segmentation mAP@0.5:0.95 (0.6896) is somewhat lower than the results in general object segmentation tasks. This discrepancy highlights the increased complexity of dental segmentation, where fine anatomical boundaries and overlapping structures introduce additional challenges compared to general object detection.

3.3. Confusion matrix analysis

The confusion matrix in Figure 5 provides further insights into the distribution of classification errors across all tooth classes. The strong diagonal dominance indicates that the majority of predictions were correctly classified. However, a small number of off-diagonal elements reflect misclassifications, particularly among teeth in the posterior region (classes 24–28). These findings are consistent with the lower precision and recall observed for these categories in the class-wise analysis. Additionally, a minor degree of confusion was observed with the background class, particularly for teeth with less distinct boundaries. Nonetheless, the overall structure of the matrix confirms high classification accuracy and minimal cross-class errors.

These findings are consistent with previous studies that applied deep learning to panoramic radiographs and reported higher misclassification rates in posterior teeth due to anatomical complexity and overlapping shadows [22]. Similarly, YOLO-based approaches have also demonstrated a performance gap between anterior and posterior teeth, with frequent confusion occurring in maxillary premolars and molars where root overlap is common [26], [32]. The results corroborate these challenges, reinforcing that posterior teeth remain the most difficult to classify accurately.

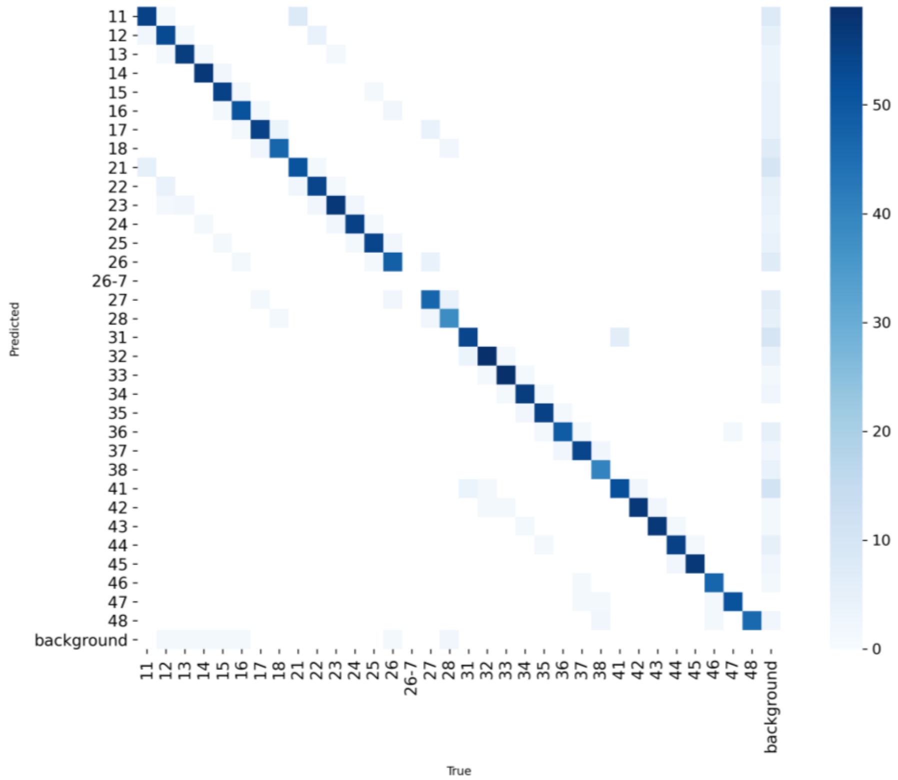

Figure 5. Confusion matrix of the YOLO segmentation model

3.4. Analysis of detection and segmentation results

Figure 6 presents the output of the YOLO segmentation model applied to a standard panoramic dental radiograph. This figure present class labels, confidence scores, and colored segmentation masks. High performance was achieved across most teeth, with slightly lower confidence in posterior regions. Each tooth is correctly identified and labeled with its corresponding class number, accompanied by confidence scores ranging from 0.87–0.97. The bounding boxes (rectangular outlines) and segmentation masks (colored overlays) demonstrate precise localization of individual teeth across both the maxillary and mandibular arches.

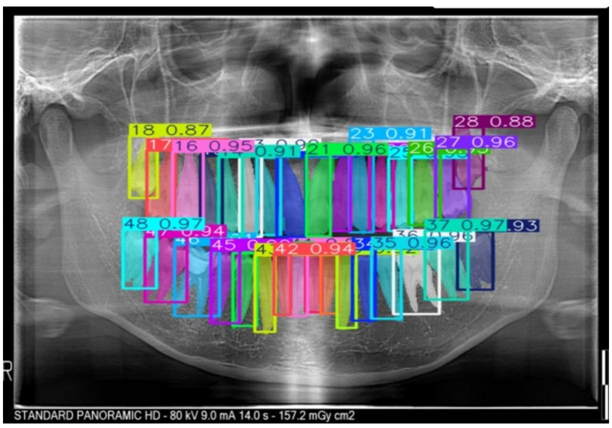

Figure 6. Example of YOLO-based tooth detection and segmentation on a panoramic dental radiograph

Most detections exhibit high confidence (>0.90), reflecting strong model reliability in tooth recognition. For example, teeth 42, 45, 48, and 37 achieved confidence scores between 0.94 and 0.97, indicating excellent agreement between predicted and true boundaries. Lower confidence values, such as tooth 18 (0.87) and tooth 28 (0.88), correspond to regions where visibility is reduced due to anatomical overlap, positioning, or image contrast, highlighting areas where the model faces greater challenges. Despite the dense arrangement of teeth and overlapping bounding boxes in the central region, the model successfully distinguishes individual structures without significant misclassification. This suggests robust capability in handling complex anatomical presentations typical of panoramic imaging. However, occasional overlaps between neighboring teeth (e.g., teeth 21–26 and 33–36) may introduce ambiguity and should be further refined through improved boundary-aware segmentation methods.

Figure 6 is consistent with previous studies applying deep learning models to panoramic dental imaging. Earlier research using a U-Net-based segmentation approach demonstrated that posterior regions often exhibited boundary ambiguities, leading to under- or over-segmentation [22]. Likewise, YOLOv8-based studies reported high-confidence detections across most teeth, although posterior landmarks generally showed slightly lower precision due to anatomical complexity [32]. Another study on YOLOv8 for mandibular lesion detection highlighted that irregular or low-contrast regions remained susceptible to reduced confidence scores despite the model's overall strong performance [26]. Overall, Figure 6 illustrates the model's capacity for high-accuracy dental landmark identification and segmentation, confirming its potential for automated tooth numbering, clinical assessment, and decision support in dental imaging.

3.5. Clinical and practical implications

The results show the potential of YOLO-based segmentation for automated tooth detection and numbering in panoramic dental radiographs. High accuracy across most classes suggests that the model can serve as a reliable tool for clinical applications, including orthodontic assessment, treatment planning, and forensic identification. Importantly, the model also demonstrates high computational efficiency. Training was completed in 0.941 hours for 300 epochs, reflecting rapid convergence compared to many CNN-based frameworks. During inference, the model achieved an average processing time of 0.2 ms for preprocessing, 3.0 ms for inference, and 4.9 ms for post-processing per image, confirming its capability for near real-time performance in clinical settings. Such efficiency is particularly valuable in busy dental practices, where fast turnaround of radiographic analysis can improve workflow and decision-making.

Previous clinical evaluations of AI in dental imaging have already shown the importance of robust automated systems for diagnosis and treatment planning [11], [22], [24], [26], [32]. The results extend these findings by demonstrating that YOLO-based segmentation achieves comparable accuracy to earlier CNN-based methods, while offering faster inference and potential for real-time applications. However, like prior studies, we confirm that image quality and anatomical complexity at the periphery remain significant limitations that need to be addressed in future research. Despite these promising results, this study has several methodological limitations that should be considered when interpreting the findings. In particular, model performance was evaluated using a single fixed train–validation–test split rather than image-level k-fold cross-validation or repeated multi-seed experiments. While this strategy was adopted to maintain an independent test set and align with common practice in YOLO-based segmentation frameworks, it may limit the robustness of performance estimation for relatively small datasets. Future studies should incorporate k-fold cross-validation or repeated random splits to quantify performance variability across different data partitions and provide more statistically robust estimates of clinical reliability.

4. CONCLUSION

This study demonstrated the effectiveness of a YOLOv8-based deep learning model for automated tooth detection, segmentation, and numbering in panoramic dental radiographs. The model achieved high accuracy, with bounding box metrics of precision 0.9075, recall 0.9352, F1-score 0.9212, and mAP@0.5 of 0.9510, while segmentation performance was comparable (precision 0.9078, recall 0.9344, F1-score 0.9209, and mAP@0.5 of 0.9504). Although performance declined at stricter thresholds (mAP@0.5:0.95=0.7763 for bounding boxes and 0.6896 for masks), the overall results confirm robust localization and segmentation capabilities. Class-wise and qualitative analyses further confirmed the model's reliability in central and anterior teeth, while highlighting challenges in posterior teeth (18, 24, 26, and 28) due to overlapping anatomy and reduced image quality at the periphery. Despite these limitations, the model consistently achieved high confidence predictions (>0.90) across most classes and maintained consistent numbering aligned with the ground truth. The findings demonstrate the potential of YOLOv8 as a fast, accurate, and scalable tool for clinical dentistry, supporting applications such as orthodontic assessment, prosthodontic planning, and forensic identification while improving diagnostic efficiency. However, this study was limited by the use of a single train–validation–test split without cross-validation, which may affect the robustness of performance evaluation. Future studies should incorporate image-level cross-validation, expand datasets, apply targeted augmentation and boundary refinement techniques, and validate the model on larger multi-institutional datasets to improve segmentation accuracy and ensure clinical generalizability.

ACKNOWLEDGMENTS

The author sincerely thanks drg. Harfindo Nismal, Sp.BM, Director of RSGM Unand, along with Mr. Adnan, drg. Gunawan, Sp.RKG, Mr. Idhil Ihsan, Amd.Kes.Rad, and the Radiology Department staff of RSGM Unand for their valuable support and assistance throughout this study. The author also gratefully acknowledges Mr. Rico Adrial, M.Si, Dr. Afdhal Muttaqin, M.Si, and Ms. Mutya Vonnisa, M.Sc. for their constructive feedback and insightful suggestions.

FUNDING INFORMATION

This research was funded by the Faculty of Mathematics and Natural Sciences, Universitas Andalas, under the RKAT scheme through research contract no. 11/UN.16.03.D/PP/FMIPA/2023, fiscal year 2023.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sri Oktamuliani	✓	✓	✓	✓	✓	✓				✓		✓	✓	✓
Luqyana Mahdiyah		✓			✓	✓		✓	✓		✓		✓	
Haritsul Haq		✓	✓				✓		✓		✓			
Nesa Perdana Putri	✓			✓	✓		✓		✓					
Wulandani Liza Putri	✓	✓		✓	✓	✓		✓		✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

Ethical approval was obtained from the Faculty of Medicine, Universitas Andalas (No: 513/UN.16.2/KEP-FK/2024). The study was conducted in accordance with the Declaration of Helsinki, guaranteeing ethical use of human imaging data for research purposes.

DATA AVAILABILITY

The data supporting the findings of this study are available upon reasonable request from the corresponding author, [SO]. Due to patient privacy considerations and institutional data governance policies, the full-resolution panoramic radiographs cannot be made publicly available. A subset of the dataset structure, annotation format, and project configuration used in this study is accessible via the Roboflow platform (https://universe.roboflow.com/tugas-akhir-rlyxr/segmentasi-gigi) to facilitate transparency and reproducibility. Access to the complete dataset, including training, validation, and test splits, is restricted and may be granted upon request for research purposes, subject to approval of ethics and data use agreements.

REFERENCES

[1] H. C. Crow, E. Parks, J. H. Campbell, D. S. Stucki, and J. Daggy, "The utility of panoramic radiography in temporomandibular joint assessment," *Dentomaxillofacial Radiology*, vol. 34, no. 2, pp. 91–95, 2005, doi: 10.1259/dmfr/24863557.

[2] S. Vaseemuddin, "Incidental findings on panoramic radiograph: a clinical study," *Journal of Advanced Medical and Dental Sciences Research*, vol. 4, no. 6, 2016, doi: 10.21276/jamdsr.2016.4.6.54.
[3] Â. G. D. Schroder *et al.*, "Diagnosis of TMJ degenerative diseases by panoramic radiography: is it possible? a systematic review and meta-analysis," *Clinical Oral Investigations*, vol. 27, no. 11, pp. 6395–6412, Oct. 2023, doi: 10.1007/s00784-023-05293-8.
[4] S. Narain, "The role of 3D imaging in endodontic diagnosis and treatment planning: a systematic review," *Journal of Dental Care*, vol. 1, no. 2, pp. 22–31, 2024.
[5] A. A. de Magalhães and A. T. Santos, "Advancements in diagnostic methods and imaging technologies in dentistry: a literature review of emerging approaches," *Journal of Clinical Medicine*, vol. 14, no. 4, p. 1277, 2025, doi: 10.3390/jcm14041277.
[6] C. Angelopoulos, S. Thomas, S. Hechler, N. Parissis, and M. Hlavacek, "Comparison between digital panoramic radiography and cone-beam computed tomography for the identification of the mandibular canal as part of presurgical dental implant assessment," *Journal of Oral and Maxillofacial Surgery*, vol. 66, no. 10, pp. 2130–2135, 2008, doi: 10.1016/j.joms.2008.06.021.
[7] H. G.-Araujo, N. O.-Santos, A. X. M. Mancini, M. L. Oliveira, and C. O.-Santos, "Retrospective assessment of dental implant-related perforations of relevant anatomical structures and inadequate spacing between implants/teeth using cone-beam computed tomography," *Clinical Oral Investigations*, vol. 24, pp. 3281–3288, 2020, doi: 10.1007/s00784-020-03205-8.
[8] Ö. Özalp *et al.*, "Comparing the precision of panoramic radiography and cone-beam computed tomography in avoiding anatomical structures critical to dental implant surgery: a retrospective study," *Imaging Science in Dentistry*, vol. 48, no. 4, 2018, doi: 10.5624/isd.2018.48.4.269.
[9] K. Zaborowicz, M. Zaborowicz, K. Cieślińska, A. D.-Micker, M. Firlej, and B. Biedziak, "Artificial intelligence methods in the detection of oral diseases on pantomographic images—a systematic narrative review," *Journal of Clinical Medicine*, vol. 14, no. 9, May 2025, doi: 10.3390/jcm14093262.
[10] R. Jagtap *et al.*, "Automatic feature segmentation in dental panoramic radiographs," *Science Progress*, vol. 107, no. 4, Oct. 2024, doi: 10.1177/00368504241286659.
[11] C. Kim, D. Kim, H. Jeong, S.-J. Yoon, and S. Youm, "Automatic tooth detection and numbering using a combination of a CNN and heuristic algorithm," *Applied Sciences*, vol. 10, no. 16, Aug. 2020, doi: 10.3390/app10165624.
[12] M. L. Parsanna, "A review on various machine learning algorithms," *International Journal of Scientific Research in Engineering and Management*, vol. 9, no. 3, pp. 1–9, Mar. 2025, doi: 10.55041/IJSREM42692.
[13] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom),* New Delhi, India, 2016, pp. 1310-1315.
[14] S. Ray, "A quick review of machine learning algorithms," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Feb. 2019, pp. 35–39, doi: 10.1109/COMITCon.2019.8862451.
[15] P. P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, Aug. 2018, pp. 1–6, doi: 10.1109/ICCUBEA.2018.8697857.
[16] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, "An unsupervised machine learning algorithms: comprehensive review," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 911–921, Apr. 2023, doi: 10.12785/ijcds/130172.
[17] E. F. Morales and H. J. Escalante, "A brief introduction to supervised, unsupervised, and reinforcement learning," in *Biosignal Processing and Classification Using Computational Learning and Intelligence*, Academic Press, 2022, pp. 111–129, doi: 10.1016/B978-0-12-820125-1.00017-8.
[18] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, Nov. 2006, doi: 10.1007/s10462-007-9052-3.
[19] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, Nov. 2021, doi: 10.1007/s42979-021-00815-1.
[20] M. P. Muresan, A. R. Barbura, and S. Nedevschi, "Teeth detection and dental problem classification in panoramic X-ray images using deep learning and image processing techniques," in *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, Sep. 2020, pp. 457–463, doi: 10.1109/ICCP51029.2020.9266244.
[21] J.-Y. Cha, H.-I. Yoon, I.-S. Yeo, K.-H. Huh, and J.-S. Han, "Panoptic segmentation on panoramic radiographs: deep learning-based segmentation of various structures including maxillary sinus and mandibular canal," *Journal of Clinical Medicine*, vol. 10, no. 12, Jun. 2021, doi: 10.3390/jcm10122577.
[22] J.-H. Lee, S.-S. Han, Y. H. Kim, C. Lee, and I. Kim, "Application of a fully deep convolutional neural network to the automation of tooth segmentation on panoramic radiographs," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 129, no. 6, pp. 635–642, Jun. 2020, doi: 10.1016/j.oooo.2019.11.007.
[23] A. H. Abdi, S. Kasaei, and M. Mehdizadeh, "Automatic segmentation of mandible in panoramic X-ray," *Journal of Medical Imaging*, vol. 2, no. 4, Nov. 2015, doi: 10.1117/1.JMI.2.4.044003.
[24] A. Karaoglu, C. Ozcan, A. Pekince, and Y. Yasa, "Numbering teeth in panoramic images: a novel method based on deep learning and heuristic algorithm," *Engineering Science and Technology, an International Journal*, vol. 37, Jan. 2023, doi: 10.1016/j.jestch.2022.101316.
[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
[26] M. Rašić, M. Tropčić, P. Karlović, D. Gabrić, M. Subašić, and P. Knežević, "Detection and segmentation of radiolucent lesions in the lower jaw on panoramic radiographs using deep neural networks," *Medicina*, vol. 59, no. 12, Dec. 2023, doi: 10.3390/medicina59122138.
[27] E. Casas, L. Ramos, C. Romero, and F. R.-Echeverría, "A comparative study of YOLOv5 and YOLOv8 for corrosion segmentation tasks in metal surfaces," *Array*, vol. 22, Jul. 2024, doi: 10.1016/j.array.2024.100351.
[28] M. Z. A. Shahriar, B. Jahan, A. Al Mamun, M. J. Hossen, and S. H. Mazumder, "Bangladeshi crops leaf disease detection using YOLOv8," *Heliyon*, vol. 10, no. 18, Sep. 2024, doi: 10.1016/j.heliyon.2024.e36694.
[29] M. G. Ragab *et al.*, "A comprehensive systematic review of YOLO for medical object detection (2018 to 2023)," *IEEE Access*, vol. 12, pp. 57815–57836, 2024, doi: 10.1109/ACCESS.2024.3386826.
[30] T. Y. Lin *et al.*, "Microsoft COCO: common objects in context," in *Computer Vision – ECCV 2014*, Cham, Springer: Springer, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.
[31] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.
[32] I. Tafala, F.-E. B.-Bouazza, A. Edder, O. Manchadi, M. E.-Taoussi, and B. Jioudi, "Cephalometric landmarks identification through an object detection-based deep learning model," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 2, pp. 859–867, 2024, doi: 10.14569/IJACSA.2024.0150286.

BIOGRAPHIES OF AUTHORS

Sri Oktamuliani received her Doctor of Biomedical Engineering degree from the Graduate School of Biomedical Engineering, Tohoku University, Japan in 2019. She obtained her B.Sc. in Physics from Universitas Andalas in 2011 and her M.Sc. in Physics from Institut Teknologi Bandung in 2014. She is currently a lecturer at the Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Andalas, Indonesia. Her research interests are in medical physics, with a particular focus on medical imaging, radiation protection, and dosimetry applications. She has published over 40 papers in international journals and conferences. She can be contacted at email: srioktamuliani@sci.unand.ac.id.

Luqyana Mahdiyah received her Bachelor of Science (B.Sc.) degree in Physics from Universitas Andalas in 2024. During her undergraduate studies, she developed an interest in medical physics, particularly in medical imaging and image analysis. Her undergraduate thesis focused on applying deep learning for the automatic segmentation of panoramic dental images, achieving high accuracy and highlighting the potential of AI in supporting dental diagnosis. She can be contacted at email: luqyanamahdiyah@gmail.com.

Haritsul Haq received a Bachelor of Science in Physics degree from Universitas Andalas in 2024. He previously worked as a laboratory analyst at the product certification agency LSPro Ceprindo. He is currently a medical physicist at RSUD Dr. Muhammad Zein Painan in Pesisir Selatan Regency, West Sumatra Province, Indonesia. His research interests include medical physics, specifically medical imaging, radiation protection, and dosimetry applications. He can be contacted at email: haqharitsul@gmail.com.

Nesa Perdana Putri received her general practitioner degree in dentistry from the Faculty of Dentistry, Universitas Andalas, in 2015. She is currently practicing as a dentist at RSUD Dr. Muhammad Zein Painan and also runs a private dental clinic, Ganesha Dental Care Painan. In addition to her clinical practice, she is actively involved in continuing professional development (CPD) programs accredited by the Ministry of Health of the Republic of Indonesia and has participated in the continuing internship program for dentists, further enhancing her professional skills and competencies while practicing. She can be contacted at email: nesadentist@gmail.com.

Wulandani Liza Putri received her orthodontic specialist degree from the Universitas Indonesia in 2021. She earned her doctor of dental surgery (DDS) degree from the Faculty of Dentistry, Universitas Andalas, in 2014. She is currently a lecturer at the Faculty of Dentistry, Universitas Andalas, Indonesia. Her research interests focus on digital imaging in dentistry. She can be contacted at email: wulandanilizaputri@dent.unand.ac.id.