Journal of Associated Medical Sciences 2026; 59(2): 316-326

JAMS
Journal of Associated Medical Sciences

Scopus Indexed & Thai-Journal Citation Index Centre (TCI)

Journal of Associated Medical Sciences

Journal homepage: https://www.tci-thaijo.org/index.php/bulletinAMS/index

Development of a YOLOv8-based deep learning framework for detection and classification of periapical radiolucencies in dental radiographs using the periapical index

Natdanai Hirata[1], Wannakamon Panyarak[2], Kittipit Klanliang[3], Arnon Charuakkra[2], Kittichai Wantanajittikul[1*]

[1]Department of Radiologic Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai Province, Thailand.
[2]Division of Oral and Maxillofacial Radiology, Department of Oral Biology and Diagnostic Sciences, Faculty of Dentistry, Chiang Mai University, Chiang Mai Province, Thailand.
[3]Division of Endodontics, Department of Restorative Dentistry and Periodontology, Faculty of Dentistry, Chiang Mai University, Chiang Mai Province, Thailand.

ARTICLE INFO

Article history:
Received 3 February 2025
Accepted as revised 18 March 2026
Available online 23 March 2026

Keywords:
Apical periodontitis, periapical radiograph, periapical index, deep learning, YOLO.

ABSTRACT

Background: Apical periodontitis (AP) is one of the most common oral diseases worldwide. Radiographs are widely used for diagnosing AP. However, their interpretation is subjective and time-consuming, contributing to increased clinical workload. Although deep learning has been introduced to assist diagnosis, most existing models are trained on limited sets of teeth, restricting their generalizability.

Objective: To investigate the potential of the YOLOv8 model for detecting and classifying periapical radiolucencies in periapical radiographs using the PAI criteria.

Materials and methods: A total of 1,191 periapical radiographs from both mandibular and maxillary jaws were collected and scored. Experts annotated 5,498 bounding boxes covering periapical root areas (PRAs). YOLOv8 models were developed to detect PRAs and classify PAI scores, which were grouped into three classes: PAI 1, PAI 2, and PAI 3-5. A custom post-processing method was implemented to improve inference by filtering overlapping detections. Filtered predictions from the test dataset were evaluated for both post-processing methods using precision, recall, F1-score, and mean average precision at an IoU threshold of 0.5 (mAP50). Confusion matrices were used to analyze misclassification patterns, while average IoU assessed overall detection performance.

Results: The best YOLOv8 model achieved a mean average precision at an intersection over union (IoU) threshold of 0.5 (mAP50) of 0.59 with a global average IoU of 0.62. A binary classification ablation model, grouping PAI 1-2 as “healthy” and PAI 3–5 as “diseased,” achieved an improved mAP50 of 0.66. The custom post-processing algorithm successfully eliminated all overlapping boxes and enhanced model performance.

Conclusions: Despite moderate performance, the YOLOv8-based models demonstrate potential as adjunct tools to assist dental practitioners in radiographic evaluation, particularly for less experienced clinicians. Further development is required to improve clinical applicability and robustness.

* **Corresponding contributor.**
Author’s Address: *Department of Radiologic Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai Province, Thailand.*
E-mail address: kittichai.wan@cmu.ac.th
doi: 10.66285/JAMS.2026.063
E-ISSN: 2539-6056

Introduction

Untreated dental health problems can lead to irreversible damage and negatively affect the quality of life of individuals.[1,2] Apical periodontitis (AP) is an infection of the root canals and surrounding tissues which progresses from caries.[3,4] A periapical radiograph (Pa) is an intraoral radiograph that provides two-dimensional high-resolution detail of tooth roots and surrounding structures. It is commonly used to assess the periapical health of one or several specific teeth. Evaluating or interpreting Pa usually takes less time compared to the other examinations. Additionally, it is non-invasive and widely available.[5] Nonetheless, clinical evaluation is essential for diagnosing periapical lesions, as relying solely on conventional radiography may lead to unnecessary surgical procedures.[6,7] Currently, there is increasing interest in developing radiograph-based diagnostic methods to streamline workflows and reduce dependence on human interpretation. Decades ago, the Periapical Index (PAI) scoring system was introduced as an early attempt to standardize radiographic assessment of periapical conditions.[8] It is one of the most widely used tools for assessing AP on radiographs. It utilizes a 5-point scale, ranging from normal (score 1) to severe AP (score 5). The system was developed based on correlations between radiographic features and histological findings, which are considered gold standard. While the PAI provided a valuable framework for consistent interpretation, it still heavily relied on expert visual assessment. Today, advances in artificial intelligence (AI) offer the potential to automate and enhance such radiographic evaluations,[9,10] building on the foundational goals established by tools like PAI.

Deep learning has been applied in dental imaging to overcome challenges in radiography. The models have shown strong performance in detection, classification, and segmentation, with potential to support dentists in diagnosis.[11-14] Sadr *et al*.[15] conducted a meta-analysis and reported high accuracy of deep learning models for detecting periapical lesions in dental radiographs despite some risk of bias across studies. Berne *et al*. used You Only Look Once (YOLO) version 3 to develop an automated lesion detection system on panoramic images with reliable clinical performance.[16] Moidu *et al*. developed a system that detects and classifies periapical lesions based on PAI scoring in mandibular Pa also using YOLO version 3.[17] The model achieved good performance in classifying all PAI scores. However, it struggled to differentiate the score with early-stage changes from the adjacent scores. Furthermore, their dataset consisted only of mandibular PAs, as they stated that excluding maxillary Pas reduced the complexity caused by the superimposition of maxillary anatomical structures.

YOLO version 8 (YOLOv8) was the latest and most stable version at the time of model development.[18] YOLOv8 provides improved detection accuracy and fast inference speed, making it suitable for real-time clinical applications.[19,20] In addition, its anchor-free detection architecture and enhanced feature extraction enable effective detection of small objects,[21] which is beneficial for identifying lesions in dental radiographs. Although newer versions have since been released, YOLOv8 remains a strong benchmark and has been successfully applied in dental imaging with promising results. Previously, we investigated the performance of deep learning models in classifying cropped tooth root images.[22] The models showed potential in distinguishing normal roots from diseased roots. However, to fully aid dental practitioners with the screening and diagnosing processes, automatic detection of the roots should be provided.

Therefore, the aim of this study was to develop an automatic detection and classification system for periapical root areas (PRAs) in Pa using YOLOv8, with the potential to assist dental practitioners in making diagnoses or evaluations based on radiographs alone, reducing reliance on clinical examinations or histological results, and ultimately enhancing the efficiency of the radiograph interpretation process.

Materials and methods

Data acquisition and inclusion criteria

Ethical approval for this retrospective study was obtained from the Institutional Ethical Review Board, Faculty of Dentistry, Chiang Mai University, Chiang Mai, Thailand. Digital Pas were obtained from the faculty's image database. Radiographic data were acquired from Heliodent Plus intraoral X-ray units (Dentsply Sirona, Charlotte, NC, USA) utilizing either Carestream 7600 or VistaScan Imaging Plates (DÜRR DENTAL, Bietigheim-Bissingen, Germany). The periapical radiographs were fully anonymized and subsequently viewed on a medical-grade display, cropped, and exported portable network graphic (PNG) files, with selected images having a minimum resolution of 800 by 1000 pixels. According to the criteria outlined by Bachani *et al*.,[23] radiographs were included if they demonstrated ideal imaging characteristics and contained at least one tooth showing evidence of pulp infection, a large filling, or a carious lesion approaching the pulp. Images with inadequate quality or artifacts that hindered assessment of the lamina dura were excluded. A total of 1,191 Pas were included from 649 patients (57.00% female, 43.00% male, aged 13-87). Both mandibular and maxillary Pas were used, with mandibular images comprising 76.00% of the total. An overview of the study workflow is presented in Figure 1.

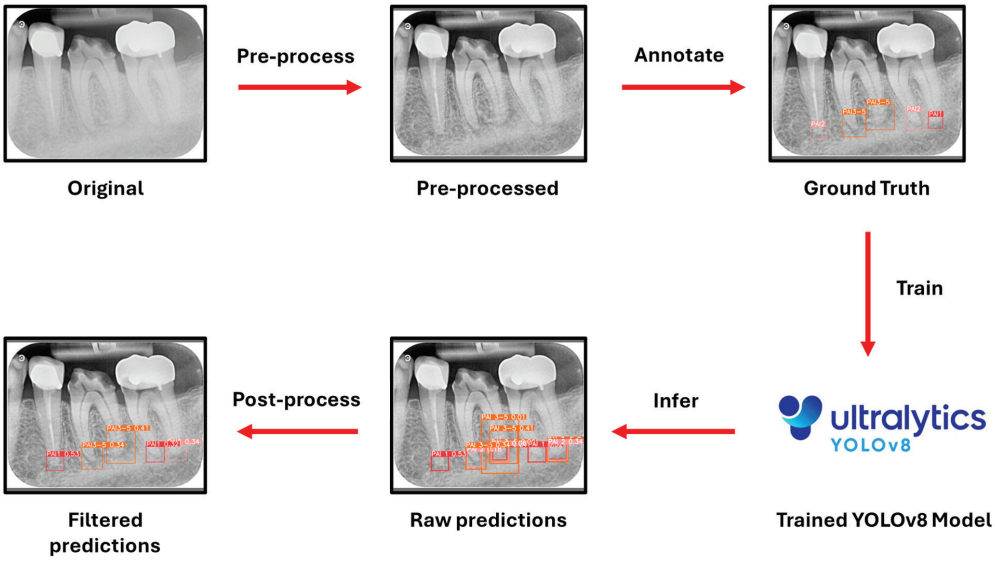

Figure 1. *A schematic illustration of the overall workflow used in this study.*

Image annotation and dataset splitting

The annotation process was conducted jointly by three experts-two oral and maxillofacial radiologists and one endodontist-with each having more than 10 years of clinical experience. The Computer Vision Annotation Tool (CVAT; available at www.cvat.ai) was employed to perform image labeling. A total of 5,498 bounding boxes were created, each covering a PRA. Every PRA was labeled with a corresponding class. PAI scores were divided into three classes for analysis: PAI 1, PAI 2, and PAI 3-5. Examples of PRAs from each class, along with the classification criteria, are shown in Figure 2. The test dataset consisted of 146 Pas, representing the most recent images collected from August through November 2024. The remaining images, 1045 Pas, taken between December 2011 and July 2024, were split into training and validation datasets using an 80:20 ratio (training=836 images, validation=209 images) prior to model training to ensure that no images were shared between the datasets. This ratio was selected as it is commonly used in machine learning studies to provide sufficient data for model training while maintaining a validation set for performance monitoring and hyperparameter tuning, while also ensuring that minority classes retain enough samples for meaningful evaluation. A stratified splitting strategy was used to divide the datasets while preserving the original class distribution to mitigate class imbalance. Class imbalance may affect model performance, as classes with more samples provide more training examples, while underrepresented classes may result in lower detection accuracy and recall. An overview of the class distribution across the training, validation, and test datasets is provided in Table 1, offering insights into the balance and representation of each class used in this study.

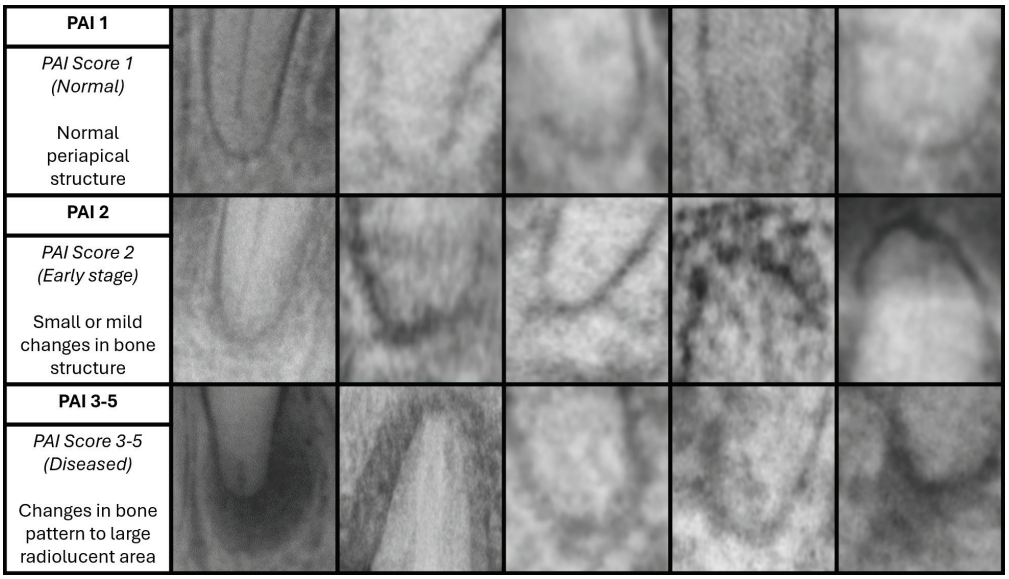

Figure 2. *The classification scheme used in this study, including the PAI scoring criteria for each class and representative example PRAs.*

Table 1. Distribution of dataset instances across classes.

Categories	Training set	Validation set	Test set	Total instances
PAI 1	2,740	716	452	3,908
PAI 2	456	108	84	648
PAI 3-5	622	161	159	942

Image processing and augmentation techniques

Contrast-limited adaptive histogram equalization (CLAHE) and bilateral filtering (BF) techniques were used on all images to normalize images acquired using two different scanning systems.[24-29] Prior to training, YOLOv8 also preprocesses input images by first resizing and padding them to a square of the designated size. The pixel values are then normalized to a range of 0-1, followed by the application of the image augmentation techniques. Image augmentation was applied to the training dataset to increase its variability. The augmentation methods included scaling, vertical flipping, horizontal flipping, and rotation, with respective parameters of 0.30, 0.50, 0.50, and 30°. Additionally, YOLOv8's mosaic augmentation was employed with a probability of 0.50, which combines four images into a single image to improve the model's ability to detect and classify infrequent objects. These techniques also help improve the model's exposure to variations of the minority class.[30]

Model development

Kaggle (available at www.kaggle.com) and Google Colab (available at www.colab.reseach.google.com) online coding platforms were used for model training and development with Python programming environments. We chose the YOLOv8 model architecture with no pre-trained weights. Hyperparameters were chosen using hyperparameter sweeping algorithms. The following training hyperparameters and options were used: a YOLOv8-m model, a square input image size of 1000 pixels, a batch size of 16, a dropout rate of 0.2, the AdamW optimizer with a learning rate of 0.001429, a momentum of 0.9, and a weight decay of 0.0005. The model was trained to provide a box for each PRA detected with its class and confidence score, which represents the model's estimated probability that an object is present within a predicted bounding box.

Post-processing algorithms

YOLOv8 uses Non-Maximum Suppression (NMS) to remove low-confidence and overlapping predictions based on confidence and intersection-over-union (IoU) thresholds. However, some overlapping predictions may remain. To address this, we developed a custom post-processing algorithm that analyzes bounding box coordinates to detect containment and prioritizes the more severe class when overlapping boxes have different labels. The conditions of the custom post-processing algorithm are illustrated in Figure 3.

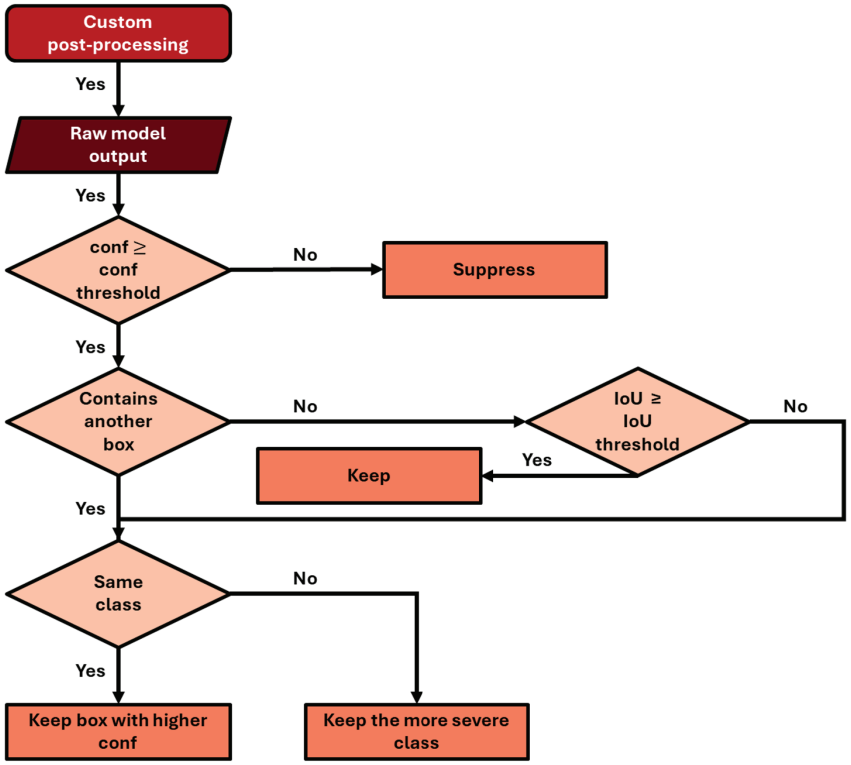

Figure 3. *A flow diagram of the custom post-processing algorithm. conf: confidence.*

Data analysis

Filtered predictions of the test dataset of both post-processing methods were evaluated and compared. Precision, recall, F1-score, and mean average precision at IoU threshold of 0.5 (mAP50) were used to describe and compare the models' performances. Additionally, generated confusion matrices were used to show detailed misclassification patterns, and average IoU was used to represent the models' overall ability in the detection task.

Results

The model achieved the best result at mAP50 of 0.56, recall of 0.57, precision of 0.58, and F1-score of 0.57. On test dataset, the application of both post-processing methods increased the model's mAP50 to a minimum of 0.57, with average IoU scores of 0.62 at 0.3 confidence and 0.59 at 0.5 confidence. The model incorporating the custom post-processing algorithm with a confidence threshold of 0.3 was selected as it demonstrated the best performance with mAP50 of 0.59, recall of 0.60, precision of 0.67, and F1-score of 0.63. The classification patterns are demonstrated in the confusion matrix generated by the model, as shown in Figure 4. In object detection, background class is added. From the confusion matrix, the background row represents missed detections, or when the model could not find the object. While the background column represents false detections, or when the model predicted where there is no object or the background.

Ablation study

An ablation study investigating the detection and classification of PRAs using YOLOv8 model on binary-class dataset was applied to reduce the complexity of the tasks and investigate the models' performances. The scores were dichotomized by grouping scores 1 and 2 as Healthy, and scores 3 to 5 as Diseased. Training parameters were also selected using hyperparameter sweeping process. The parameters were: a YOLOv8-m model, a square input image size of 800 pixels, a batch size of 32, a dropout rate of 0.2, the AdamW optimizer with a learning rate of 0.001, a momentum of 0.937, and a weight decay of 0.0005. The distribution of instances of each class in the dataset are provided in Table 2. As for the post-processing algorithm, the csutom algorithm was modified by removing only the special class condition parts.

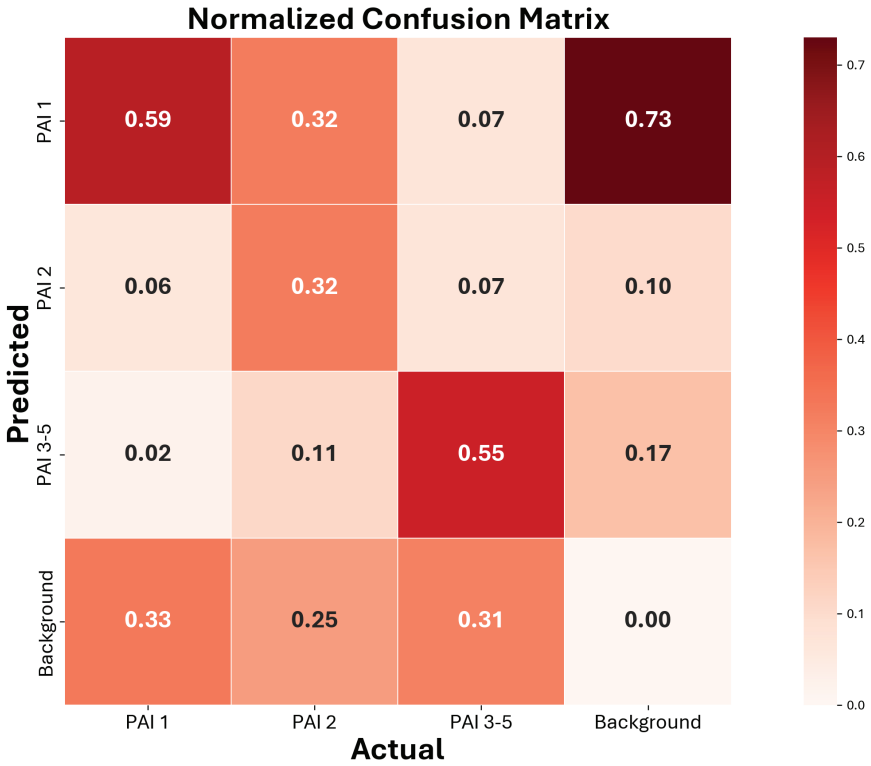

Figure 4. *Normalized confusion matrix of the model employing custom post-processing algorithms at a confidence threshold of 0.3.*

Table 2. Distribution of dataset instances across classes (binary classification).

Categories	Training set	Validation set	Test set	Total instances
Healthy	3247	773	536	4556
Diseased	622	161	159	942

The best-performing model trained on the binary class dataset achieved a mAP50 of 0.78, recall of 0.79, precision of 0.74, and F1 score of 0.76 on the validation dataset. On the test dataset, the raw model output yielded a mAP50 of 0.67. During inference, one overlapping case was observed at a confidence threshold of 0.3, while eight overlapping cases occurred at a threshold of 0.5. For the detection tasks, the model identified PRAs with an average IoU of 0.59 at both thresholds. The binary model employing the custom post-processing method with a confidence threshold of 0.5 was selected for further analysis with mAP50 of 0.66, recall of 0.64, precision of 0.76, and F1-score of 0.69. The corresponding confusion matrix is shown in Figure 4. Examples of common prediction errors after post-processing are shown in Figure 6-8.

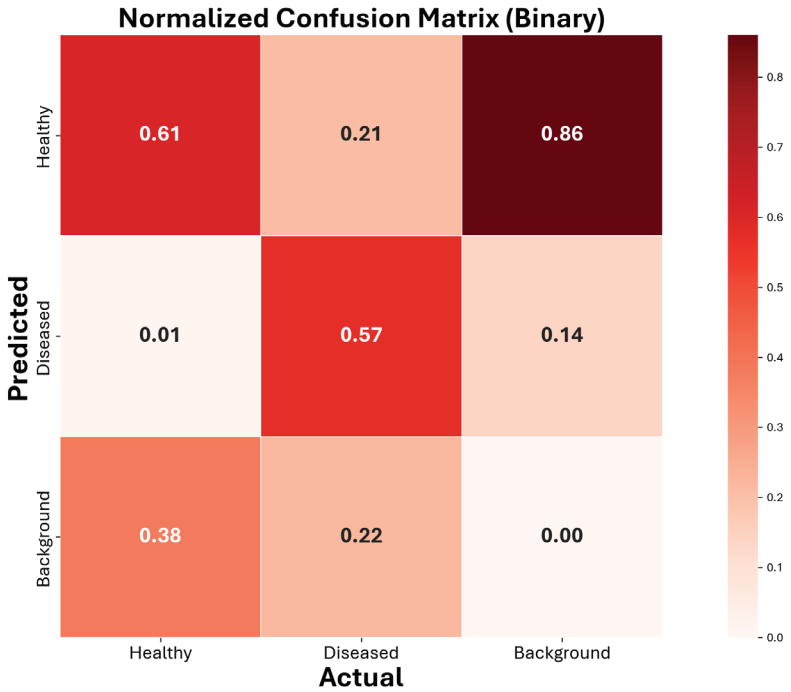

Figure 5. *Normalized confusion matrix of the binary model employing custom post-processing algorithms at a confidence threshold of 0.5.*

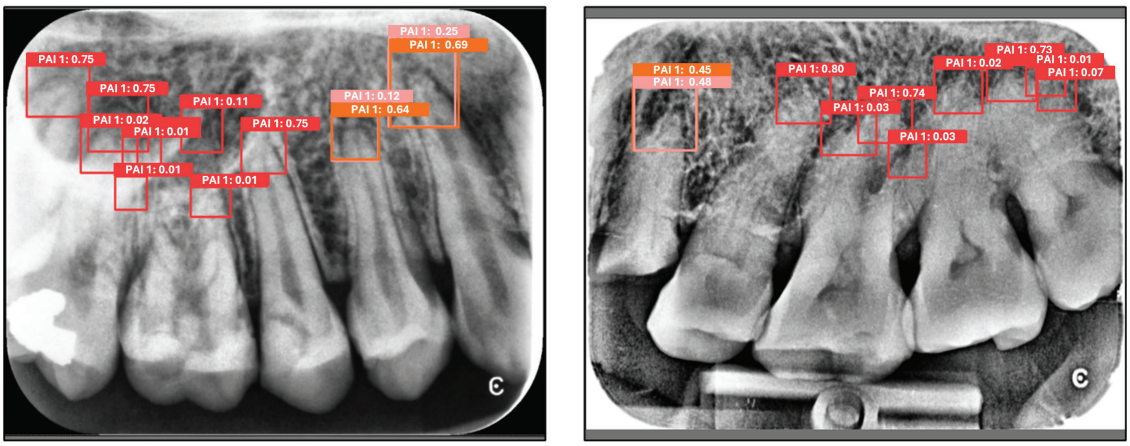

Figure 6. *Predictions on upper jaw Pas showing complete object coverage, but boxes with low confidence are suppressed.*

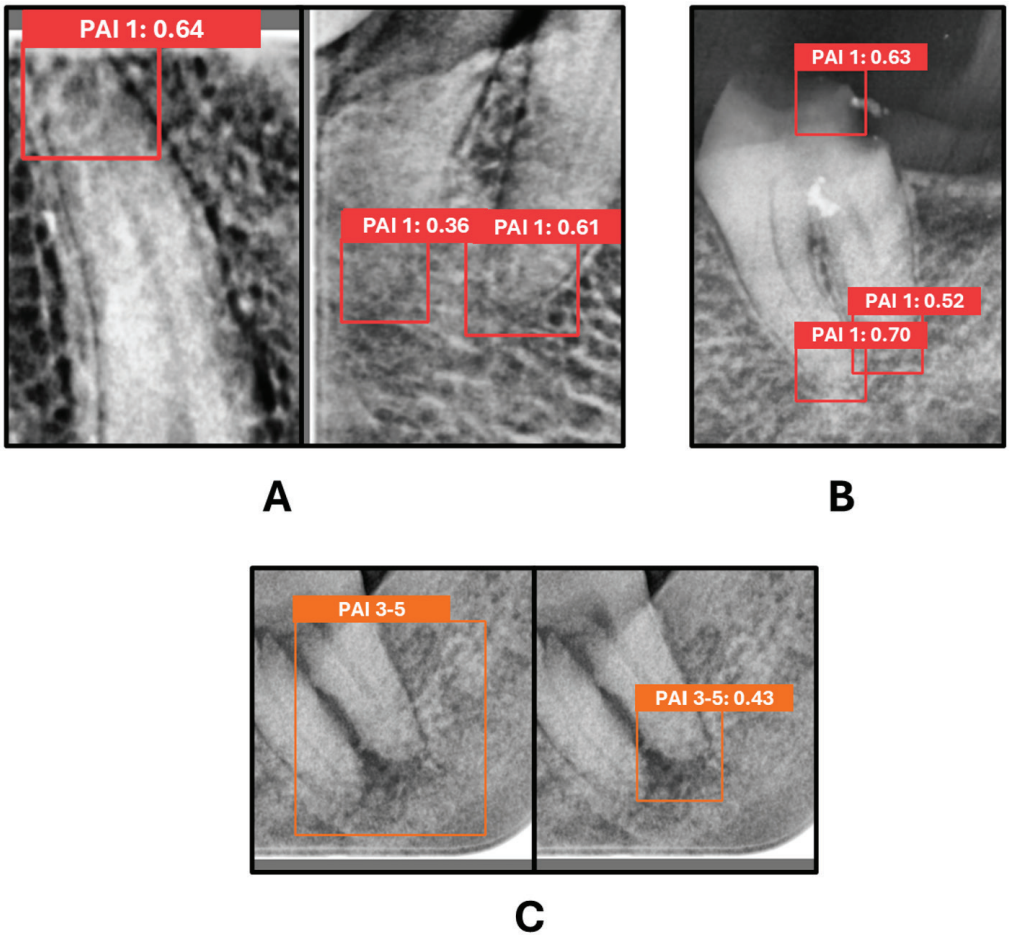

Figure 7. *False detection examples from the model: A: predicted boxes on partially visible roots (confidence 0.64 and 0.36), B: prediction on the tooth crown (confidence 0.63), C: smaller predicted box compared to left ground truth (IoU=0.26).*

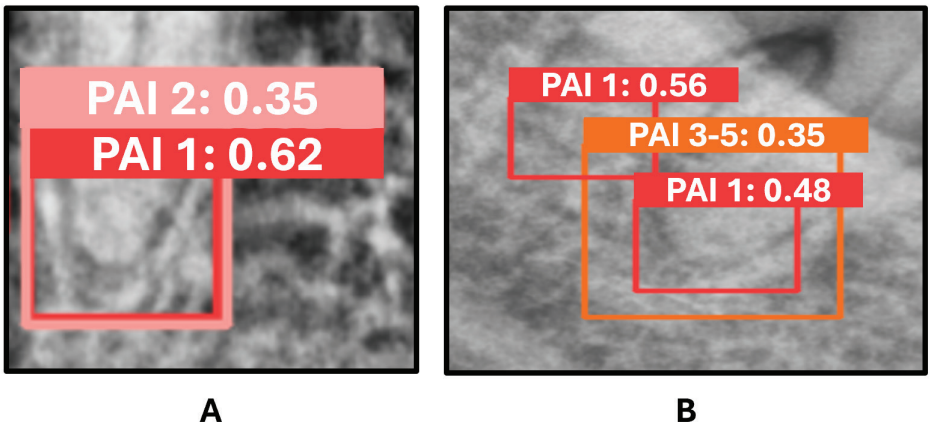

Figure 8. *Examples of bounding boxes not eliminated by NMS:*
A: near-complete overlap between PAI 1 and PAI 2 (IoU 0.99), B: PAI 1 entirely inside PAI 3-5 (IoU 0.40).

Discussion

Radiographs have long been a widely used method for screening and evaluating dental health conditions. PAI scoring system has been used to help assess periapical health in addition to radiograph features. However, interpreting and scoring dental radiographs manually are time-consuming and labor-intensive, placing a significant workload on dental practitioners, especially in high-volume clinical settings. Moreover, the procedures depend heavily on the expertise of trained dentists and radiologists. To address these limitations, deep learning has been introduced to streamline the evaluation process and support dental practitioners. Unlike previous studies that focused primarily on mandibular Pa, this study included images from both the maxillary and mandibular regions. This introduces greater anatomical variability and superimposition, reflecting a more realistic clinical scenario for automated PRA detection. The model achieved moderate accuracy in detecting and classifying PRAs with the highest mAP50 of 0.57.

As shown in Figure 4, the model had difficulty accurately classifying PAI 2, with a correct classification rate of only 0.32, which reduced the overall model accuracy. The moderate performance was influenced by several factors. Firstly, the model showed difficulty distinguishing between PAI 1 and PAI 2 due to subtle differences in radiographic features defined in the PAI scoring criteria, which led to frequent confusion between these classes. A similar pattern has also been reported in previous studies.[17,22,31] Secondly, a considerable number of raw predictions had low confidence scores and were removed during post-processing, resulting in increased missed detections. More than half of the raw predictions (56.00%) had confidence scores below 0.3. These errors were observed mostly in periapical radiographs taken from the upper jaw, as shown in the examples in Figure 6. Detection was more challenging in maxillary radiographs due to anatomical superimposition, such as the sinus floor and adjacent bones, which contributed to lower confidence predictions. In contrast, the mandibular region presents fewer sources of interference, with the mandibular canal being the primary structure that may complicate predictions. Finally, some background misclassifications occurred when normal roots were incorrectly detected as PAI 1 (71.00%), suggesting that the model occasionally confused normal roots with background regions. Figure 7 illustrates examples of these false detections that reduced model accuracy. Among them, size mismatch between the object and the predicted boxes-shown in Figure 7c-was the major contributor to background misclassification: the predicted box was counted as a false detection, while the unmatched object box was recorded as a missed detection.

In the binary classification study, the model predicted with higher confidence due to the lower task complexity. However, its correct classification rates remained moderate. Specifically, 21.00% of Diseased roots were misclassified as Healthy. Upon closer inspection, a single PAI 1 box was misclassified as PAI 3-5 (diseased), suggesting that PAI 2 exhibits characteristics like PAI 3-5, consistent with the PAI scoring system. Nonetheless, most errors in the binary model also stemmed from missed detections rather than misclassifications. Approximately 15.00% of the total predictions were suppressed by the post-processing algorithm at a confidence threshold of 0.3, with an additional 4.00% suppressed at 0.5. Thus, although the binary model generated fewer low-confidence predictions and achieved the highest recall, it still failed to detect some roots in the raw results and the missed detection rate remained high. While most of these predictions were ultimately suppressed, the three-class model was able to detect most complex PRAs. In comparison, the binary model showed stronger performance in detecting complex PRAs on mandibular periapical radiographs. Among false detections, 86% were assigned as healthy, though they comprised only 18.00% of all predictions. The model frequently produced high-confidence predictions for roots that were not fully visible, mostly around the edges. Like the previous model, size mismatch between the object and the predicted box remained the major contributor to background misclassifications. Overall, the model's performance was good, provided that early-stage lesion classification was not the focus.

The custom post-processing algorithm performed well in filtering fully contained boxes that the NMS method could not suppress. Compared with NMS, the model using the custom algorithm achieved lower recall because some overlapping cases were still left unsuppressed. However, this resulted in a slight improvement in precision, indicating that the algorithm and its special class conditions handled overlapping cases more effectively. Since only a small number of overlapping cases remained after confidence thresholding, the differences between NMS and custom post-processing models were minimal. At a confidence threshold of 0.5, no fully contained boxes were present after filtration, and both methods produced identical inference results. Some bias was introduced by forcing the model to prioritize the more severe class, which sometimes led to incorrect predictions. For example, in Figure 8a, the algorithm suppressed a contained PAI 1 box (with higher confidence) and retained the PAI 2 box, which was the correct score. In contrast, Figure 8b shows a case where the model was forced to keep a PAI 3-5 box and discard a PAI 1 box, resulting in an error. Despite this limitation, implementing class conditions increased the model's sensitivity to diseased roots, which is valuable for screening in clinical practice.[17] Nevertheless, having practitioners review the model's results can help prevent errors caused by such bias.

Moidu *et al.*[17] developed YOLOv3 models for PAI scoring in periapical radiographs. One was a five-class model covering all PAI scores, and the other was a binary model. Although a direct comparison is not appropriate due to differences in study settings, the results are discussed solely to highlight similarities in classification patterns. Their model reported PAI 1-5 accuracies of 90.90%, 30.00%, 60.00%, 71.00%, and 30.00%, respectively, while our model achieved 59.00% (PAI 1), 32.00% (PAI 2), and 55.00% (PAI 3-5). In binary classification, accuracies for healthy and diseased cases were 76.60% and 92.00% for their model, compared with 61.00% and 57.00% for ours. For both the multi-class and binary models, their results outperformed ours in all categories except PAI 2, indicating that classifying the early stage of the lesion was challenging for both studies. Their advantages were: 1) the data were acquired using a single scanning system, 2) a balanced dataset was used, and 3) most importantly, only mandibular periapical radiographs were included. Considering these advantages, their dataset was less complex and less biased, resulting in more robust models. They also noted that the model tended to overscore by one class. Based on this observation, we implemented a post-processing algorithm to correct predictions following this pattern, thereby enhancing model performance.

Limitations

Although the findings are encouraging, a few limitations should be acknowledged. Firstly, the dataset included a relatively low proportion of maxillary periapical radiographs, accounting for only 25%. To address the increased complexity caused by superimposed maxillary structures, a higher proportion of maxillary images would be necessary. A larger and more diverse dataset would also allow better class distribution.[32] Pa typically includes multiple teeth per image, which makes balancing the dataset practically challenging. In addition, external validation using datasets from other institutions was not performed in this study, and further evaluation on multi-center datasets is necessary to assess the generalizability of the proposed model. Second, our computing resources were limited. Hyperparameter sweeping and tuning were computationally expensive, which restricted the number of combinations we could explore. This reflects a trade-off between computational cost and model performance, as improving detection accuracy typically requires larger models, larger datasets, and more extensive hyperparameter optimization, all of which increase computational requirements. Nevertheless, training with larger models, a bigger dataset, and more hyperparameter combinations could potentially improve performance. Additionally, comparison with other architectures, including transformer-based models and approaches leveraging large-scale pre-training, would be valuable and could be explored in future studies. Finally, we used only CLAHE and BF for image preprocessing, as these methods are commonly employed for normalization in dental imaging. Although image preprocessing is often omitted for object detection models,[33,34] we applied these techniques to help manage dataset variability, since our comparatively smaller model-limited by computational resources-might not have been able to handle it effectively on its own.

While our models have not yet reached the performance of state-of-the-art methods, the limitations identified in this study offer valuable insights and highlight clear avenues for improvement. By addressing these challenges and implementing the strategies discussed, future models could achieve higher accuracy, improved computational efficiency, and greater adaptability to the complexities of dental imaging in clinical practice. Additionally, future work should include a comparison between the proposed model and the diagnostic performance of dentists to further evaluate its clinical applicability in detecting periapical lesions.

Conclusion

In this study, PRAs detection and classification models were developed using the YOLOv8 architecture. The multi-class model achieved moderate performance, with an mAP50 of 0.59, while the binary model reached an mAP50 of 0.66. Although not yet ready for clinical use, these models demonstrate potential as adjunct tools to support dental practitioners in radiographic evaluation, particularly for less experienced users. In a potential clinical workflow, such systems could automatically highlight suspicious regions in periapical radiographs and provide preliminary classification results, allowing clinicians to review the findings more efficiently while maintaining final diagnostic responsibility. However, further research, including larger datasets and improved model performance, is needed to develop a more robust and clinically applicable system.

Ethical approval

Ethical approval for this retrospective study was obtained from the Institutional Ethical Review Board, Faculty of Dentistry, Chiang Mai University, Chiang Mai, Thailand (Approval No. 26/2024) on July 5th, 2024. Due to the deidentification of data of the retrospective study, the informed consent was not required.

Funding

This research received no external funding.

Conflict of interest

The authors declare no conflict of interest.

CRediT authorship contribution statement

Natdanai Hirata: conceptualization, methodology, data curation, formal analysis, software, writing: original

draft preparation, reviewing and editing; **Wannakamon Panyarak**: conceptualization, investigation, resources, writing: reviewing and editing; **Arnon Charuakkra**: investigation, resources, writing: reviewing and editing; **Kittipit Klanliang**: investigation, resources, writing: reviewing and editing; **Kittichai Wantanajittikul**: conceptualization, software, methodology, writing: original draft preparation, reviewing and editing.

Acknowledgements

The authors gratefully acknowledge the support of the master's degree program in Medical Radiation Sciences, Faculty of Associated Medical Sciences, Chiang Mai University, under the CMU Presidential Scholarship.

References

[1] Arias Z, Nizami MZI, Chen X, Xu B, Kuang C, Omori K, et al. Recent advances in apical periodontitis treatment: a narrative review. Bioengineering. 2023; 10(4):488. doi: 10.3390/bioengineering10040488.

[2] Tibúrcio-Machado CS, Michelon C, Zanatta FB, Gomes MS, Marin JA, Bier CA. The global prevalence of apical periodontitis: a systematic review and meta-analysis. Int Endod J. 2021; 54(5): 712-35. doi: 10.1111/iej.13467.

[3] Buonavoglia A, Zamparini F, Lanave G, Pellegrini F, Diakoudi G, Spinelli A, et al. Endodontic microbial communities in apical periodontitis. J Endod. 2023; 49(2): 178-89. doi: 10.1016/j.joen.2022.11.015

[4] Loesche WJ. Microbiology of Dental Decay and Periodontal Disease. In: Baron S, Editor. Medical Microbiology. 4th Edition. Galveston, University of Texas Medical Branch at Galveston, Chapter 99.

[5] Lo Giudice R, Nicita F, Puleio F, Alibrandi A, Cervino G, Lizio AS, et al. Accuracy of periapical radiography and CBCT in endodontic evaluation. Int J Dent. 2018; 2018:2514243. doi: 10.1155/2018/2514243.

[6] Tanomaru-Filho M, Jorge EG, Duarte MA, Gonçalves M, Guerreiro-Tanomaru JM. Comparative radiographic and histological analyses of periapical lesion development. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2009; 107(3): 442-7. doi: 10.1016/j.tripleo.2008.12.009.

[7] Karamifar K, Tondari A, Saghiri MA. Endodontic periapical lesion: an overview on the etiology, diagnosis and current treatment modalities. Eur Endod J. 2020; 5(2): 54-67. doi: 10.14744/eej.2020.42714.

[8] Ørstavik D, Kerekes K, Eriksen HM. The periapical index: a scoring system for radiographic assessment of apical periodontitis. Dent Traumatol. 1986; 2(1): 20-34. doi: 10.1111/j.1600-9657.1986.tb00119.x

[9] Rajmohamed RF, Chapala S, Shazahan MA, Wali P, Botchu R. Evaluating the accuracy and efficiency of AI-generated radiology reports based on positive findings-a qualitative assessment of AI in radiology. Acad Radiol. 2025; 35(12): 7035-40. doi: 10.1016/j.acra.2025.09.012.

[10] Kazimierczak N, Sultani N, Chwarścianek N, Krzykowski S, Janiszewska-Olszowska J, et al. Detection accuracy of an AI platform for dental treatment features on panoramic radiographs tooth- and patient-level analyses. Sci Rep, 2026; 16: 2436. doi: 10.1038/s41598-025-32226-0

[11] Huang C, Wang J, Wang S, Zhang Y. A review of deep learning in dentistry. Neurocomputing. 2023; 554: 126629. doi: 10.1016/j.neucom.2023.126629.

[12] Sohrabniya F, Hassanzadeh-Samani S, Ourang SA, Jafari B, Farzinnia G, Gorjinejad F, et al. Exploring a decade of deep learning in dentistry: a comprehensive mapping review. Clin Oral Invest. 2025; 29: 143. doi: 10.1007/s00784-025-06216-5.

[13] Khan HA, Haider MA, Ansari HA, Ishaq H, Kiyani A, Sohail K, et al. Automated feature detection in dental periapical radiographs by using deep learning. Oral Surg Oral Med Oral Pathol Oral Radiol. 2021; 131(6): 711-20. doi: 10.1016/j.oooo.2020.08.024

[14] Du M, Wu X, Ye Y, Fang S, Zhang H, Chen M. A combined approach for accurate and accelerated teeth detection on cone beam CT images. Diagnostics (Basel). 2022; 12(7): 1679. doi: 10.3390/diagnostics12071679.

[15] Sadr S, Mohammad-Rahimi H, Motamedian SR, Zahedrozegar S, Motie P, Vinayahalingam S, et al. Deep learning for detection of periapical radiolucent lesions: a systematic review and meta-analysis of diagnostic test accuracy. J Endod. 2023; 49(3): 248-61.e3. doi: 10.1016/j.joen.2022.12.007.

[16] Ver Berne J, Saadi SB, Politis C, Jacobs R. A deep learning approach for radiological detection and classification of radicular cysts and periapical granulomas. J Dent. 2023; 135: 104581. doi: 10.1016/j.jdent.2023.104581.

[17] Moidu NP, Sharma S, Chawla A, Kumar V, Logani A. Deep learning for categorization of endodontic lesion based on radiographic periapical index scoring system. Clin Oral Investig. 2021; 26(1): 651-8. doi: 10.1007/s00784-021-04043-y.

[18] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016: 779-88. doi: 10.1109/CVPR.2016.91.

[19] Sohan M, Sai Ram T, Rami Reddy ChV. A Review on YOLOv8 and its advancements. In: Jacob IJ, Piramuthu S, Falkowski-Gilski P, Editor. Data Intelligence and Cognitive Informatics. Singapore: Springer Nature; 2024: pp 529-45.

[20] Singh K, Kumar R, Singh S. Advanced YOLOv8 architecture for multi-class brain tumor detection. J Assoc Med Sci [internet]. 2026 Feb. 25 [cited 2026 Mar. 20];59(2):250-8. available from: https://he01.tci-thaijo.org/index.php/bulletinAMS/article/

view/281957.
[21] Yaseen M. What is YOLOv8: an in-depth exploration of the internal features of the next-generation object detector. ArXiv Preprint posted online August 28, 2024. doi: 10.48550/arXiv.2408.15857.
[22] Hirata N, Pudhieng P, Sena S, Torn-asa S, Panyarak W, Klanliang K, et al. Development of periapical index score classification system in periapical radiographs using deep learning. J Digit Imaging Inform Med. 2024; 38(5): 3125-33. doi: 10.1007/s10278-024-01360-y.
[23] Bachani L, Singh M, Anshul, Lingappa A. Ideal radiographs: an insight. IP Int J Maxillofac Imaging. 2020; 6(3): 56-64. doi: 10.18231/j.ijmi.2020.017.
[24] Choi JW, Han WJ, Kim EK. Image enhancement of digital periapical radiographs according to diagnostic tasks. Imaging Sci Dent. 2014; 44(1): 31-5. doi: 10.5624/isd.2014.44.1.31.
[25] Ari T, Sağlam H, Öksüzoğlu H, Kazan O, Bayrakdar İŞ, Duman SB, et al. Automatic feature segmentation in dental periapical radiographs. Diagnostics. 2022; 12(12): 3081. doi: 10.3390/diagnostics12123081.
[26] Chen IDS, Yang CM, Chen MJ, Chen MC, Weng RM, Yeh CH. Deep learning-based recognition of periodontitis and dental caries in dental X-ray images. Bioengineering (Basel). 2023; 10(8): 911. doi: 10.3390/bioengineering10080911.
[27] Çelik B, Savaştaer EF, Kaya HI, Çelik ME. The role of deep learning for periapical lesion detection on panoramic radiographs. Dentomaxillofacial Radiol. 2023; 52(8): 20230118. doi: 10.1259/dmfr.20230118.s.
[28] Taş İÇ. Application of panoramic dental x-Ray images denoising. Int J Innov Eng Appl. 2023; 7(1): 13-20. doi: 10.46460/ijiea.1134105.
[29] Hana FM, Maulida ID. Analysis of contrast limited adaptive histogram equalization (CLAHE) parameters on finger knuckle print identification. J Phys Conf Ser. 2021; 1764(1): 012049. doi: 10.1088/1742-6596/1764/1/012049.
[30] Ahmed S, Yoshiura N, Hassan S, Khan A. Towards better accuracy on imbalance image datasets using image augmentation and convolutional neural networks. Lecture Notes in Computer Science in International Conference on the Dynamics of Information Systems (LNCS-DIS 2024); 2025; 14661: 189-212. doi: 10.1007/978-3-031-81010-7_13.
[31] Maia Filho EM, Calisto AM, De Jesus Tavarez RR, De Castri Rizzi C, Bezerra Segato RA, Bezerra da Silva LA. Correlation between the periapical index and lesion volume in cone-beam computed tomography images. Iran Endod J. 2018; 13(2): 155-8. doi: 10.22037/iej.v13i2.15040.
[32] Mooijman P, Catal C, Tekinerdogan B, Lommen A, Blokland M. The effects of data balancing approaches: a case study. Appl Soft Comput. 2023; 132: 109853. doi: 10.1016/j.asoc.2022.109853.
[33] Kawaguchi K, Kaelbling LP, Bengio Y. Generalization in deep learning. ArXiv Preprint posted online October 16, 2017. doi: 10.48550/arXiv.1710.05468.
[34] Vakanski A, Xian M. Evaluation of complexity measures for deep learning generalization in medical image analysis. Proceedings of the 2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP); 2021: 1-6. doi: 10.1109/mlsp52302.2021.9596501.