



Enhancing Caries Detection in Bitewing Radiographs Using YOLOv7

Wannakamon Panyarak¹ · Kittichai Wantanajittikul³ · Arnon Charuakkra¹ · Sangsom Prapayatasotok¹ · Wattanapong Suttpak²

Received: 2 March 2023 / Revised: 9 June 2023 / Accepted: 9 June 2023 / Published online: 28 August 2023
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2023

Abstract

The study aimed to evaluate the impact of image size, area of detection (IoU) thresholds and confidence thresholds on the performance of the YOLO models in the detection of dental caries in bitewing radiographs. A total of 2575 bitewing radiographs were annotated with seven classes according to the ICCMS™ radiographic scoring system. YOLOv3 and YOLOv7 models were employed with different configurations, and their performances were evaluated based on precision, recall, *F1*-score and mean average precision (mAP). Results showed that YOLOv7 with 640 × 640 pixel images exhibited significantly superior performance compared to YOLOv3 in terms of precision (0.557 vs. 0.268), *F1*-score (0.555 vs. 0.375) and mAP (0.562 vs. 0.458), while the recall was significantly lower (0.552 vs. 0.697). The following experiment found that the overall mAPs did not significantly differ between 640 × 640 pixel and 1280 × 1280 pixel images, for YOLOv7 with an IoU of 50% and a confidence threshold of 0.001 ($p=0.866$). The last experiment revealed that the precision significantly increased from 0.570 to 0.593 for YOLOv7 with an IoU of 75% and a confidence threshold of 0.5, but the mean-recall significantly decreased and led to lower mAPs in both IoUs. In conclusion, YOLOv7 outperformed YOLOv3 in caries detection and increasing the image size did not enhance the model's performance. Elevating the IoU from 50% to 75% and confidence threshold from 0.001 to 0.5 led to a reduction of the model's performance, while simultaneously improving precision and reducing recall (minimizing false positives and negatives) for carious lesion detection in bitewing radiographs.

Keywords Bitewing radiograph · Caries detection · Dental caries · Detection area · Confidence threshold

✉ Wattanapong Suttpak
wattanapong.su@up.ac.th

Wannakamon Panyarak
wannakamon.p@cmu.ac.th

Kittichai Wantanajittikul
kittichai.wan@cmu.ac.th

Arnon Charuakkra
arnon.c@cmu.ac.th

Sangsom Prapayatasotok
sangsom.prapaya@cmu.ac.th

¹ Division of Oral and Maxillofacial Radiology, Department of Oral Biology and Diagnostic Sciences, Faculty of Dentistry, Chiang Mai University, Suthep Road, Suthep, Mueang Chiang Mai District, Chiang Mai 50200, Thailand

² Division of Computer Engineering, School of Information and Communication Technology, University of Phayao, Phahon Yothin Road, Mae Ka, Mueang Phayao District, Phayao 56000, Thailand

³ Department of Radiologic Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Suthep Road, Suthep, Mueang Chiang Mai District, Chiang Mai 50200, Thailand

Introduction

Tooth decay, a prevalent oral disease with significant impact on overall health worldwide, can be effectively managed through early detection during clinical and radiographic examination. This can help reduce the risk of tooth loss and decrease the economic burden associated with the disease [1]. However, caries detection has posed a significant challenge for deep learning developers and researchers in recent years. Enamel lesions have been found to be underestimated more often than dentinal lesions [2]. Additionally, determining the depth of lesions from radiographic examination may exhibit low sensitivity, especially in the case of initial lesions [3]. A study by Cantu et al. demonstrated that dentists exhibited very low sensitivity for detecting initial-stage caries.

In comparison, the neural network (U-net model) exhibited significantly and substantively higher sensitivity than dentists (0.75 versus 0.36; min-max, 0.19–0.65; $p=0.006$). The sensitivity of dentists for initial lesions was low (all dentists except one had a sensitivity below 0.25), while their sensitivity for advanced lesions ranged from 0.40

to 0.75. On the other hand, the U-net model consistently achieved high sensitivities of 0.70 or above for all severity levels of lesions [4]. In a randomized control trial of AI-supported dentists, significant improvements in sensitivity were observed for enamel caries, with a sensitivity of 0.81 and a mean area under the receiver operating characteristics (ROC) curve of 0.89, compared to dentists without AI support (sensitivity of 0.72 and area under the curve of 0.85) [5]. Enhancing sensitivity for early lesions is crucial as it can facilitate caries management by arresting their progression to extensive stages, thereby avoiding the need for restorative interventions. Although increasing sensitivity may lead to an increased rate of false positives, it serves as a reminder for practitioners to provide early preventive interventions and does not result in over-treatment [6].

Various deep learning algorithms have been developed for object detection tasks that can be implemented for caries detection. The You-Only-Look-Once (YOLO) model, introduced by Redmon et al. in 2016, has shown excellent performance in real-time object detection across more than 1000 classes [7].

An updated version in 2018, YOLO version 3 (YOLOv3), outperformed the Faster Region-Based Convolutional Neural Networks (Faster R-CNNs) in automatically detecting initial caries lesions and cavities using intraoral photographs taken with smartphones, achieving an accuracy of 87.4 and 71.4% for YOLOv3 and Faster R-CNN, respectively [8]. Apart from caries detection through photographs, the YOLO algorithm has also been applied to RadioVisioGraphy (RVG) images, yielding an accuracy of 87% when a “confidence threshold” of 0.3 was set to filter out low confidence potential dental caries (false positives) while retaining regions with confidence scores equal to or above 0.3 [9].

While the confidence threshold reduces false positives, it is important to note that some true positives may potentially be missed. To address this, an intersection over union (IoU) threshold is used to determine the degree of overlap between the predicted bounding box and the ground truth bounding box, ensuring true positives are considered. In a study using YOLOv3 with an IoU threshold of 50% (IoU₅₀), good performance was demonstrated in detecting carious lesions in bitewing radiographs, achieving an accuracy of 94.19% for premolars and 94.97% for molars. However, these results did not represent enamel or dentin caries [10]. Based on the International Caries Classification and Management System (ICCMS™), carious depth is categorized into four classes with three subclasses: class 0 (no carious lesion); initial stages (RA) with subclasses RA1, RA2 and RA3, moderate stage (RB); and extensive stages (RC) with subclasses RC5–6 [11]. According to this classification, the YOLOv3 model was able to detect and classify 7 level depths of caries with an acceptable precision ranging from 0.35 to 0.77. However, for initial caries detection (classes RA1, RA2 and RA3), YOLOv3

showed inferior performance with an average precision of 0.10, 0.38 and 0.37, respectively, at IoU₅₀ [12].

Recently, YOLO version 7 (YOLOv7) has been introduced as a real-time object detection algorithm that promises to deliver high speed and accurate results [13]. The new algorithm has reformed its training strategy, known as the “Bag of Freebies” approach. This training technique improves the overall accuracy of the model without modifying the network architectures and incurring no inference cost [14]. YOLOv7 has been designed to be scalable and adaptable to various device sizes. In comparison to other real-time object detection algorithms, YOLOv7-E6E has shown to have the maximum accuracy of 56.8% and to be 120% faster than its competitors [13]. This superior performance could prove to be beneficial for detecting small objects, such as carious lesions in intraoral radiographs.

The purpose of this research is to evaluate and compare the efficacy of YOLOv7 in identifying carious lesions of varying depths within bitewing radiographs. Initially, the performance of YOLOv7 was compared to that of YOLOv3. Subsequently, various parameters, such as input image size, detection and confidence thresholds, were modified and evaluated.

Materials and Methods

The study design received approval from the Institutional Ethical Review Board (approval no. 22/2021) of the Faculty of Dentistry, Chiang Mai University, Thailand. A total of 2575 bitewing radiographs were included in the study, with 2144 images allocated for training, 256 images for validation and 175 images for testing. The radiographs were obtained between January 2016 and December 2021. To perform the annotation procedures, a single researcher exported all bitewing radiographs of permanent dentition from the Picture Archiving and Communication System (PACS) within the Faculty of Dentistry's dental hospital. The researcher saved the radiographs in the Portable Network Graphic (PNG) format while ensuring the anonymization of all images to safeguard patient privacy.

Three oral and maxillofacial radiologists, who have at least 8 years of experience, performed the annotation cooperatively after reaching the consensus. The bounding box covering occluso-cervical and mesio-distal dimensions of each tooth in the radiograph with labeling to seven classes based on the ICCMS™ radiographic scoring system [11] using the LabelImg software (<https://github.com/tzutalin/labelImg>). The seven classes based on radiographic carious depths are as follows: non-carious tooth, class 0; initial stages (RA): RA1, radiolucency in the outer half of the enamel; RA2, radiolucency in the inner half of the enamel and/or enamel-dentin junction (EDJ); RA3, radiolucency limited to the outer one-third of dentin; moderate stages

(RB): RB4, radiolucency reaching the middle one-third of dentin; extensive stages (RC): RC5, radiolucency reaching the inner one-third of dentin and clinically cavitated; and RC6, radiolucency into the pulp and clinically cavitated. Then, each annotated radiograph was exported to the Extensible Markup Language (XML) format for further analysis.

Performance Evaluation of YOLO Models on Caries Detection Tasks

The performance of YOLO models, as well as other object detection algorithms, is often evaluated using a suite of metrics, including precision, recall, *F1*-score as described in Hossin and Sulaiman in 2015 [15].

In evaluating the performance of object detection models, it is important to consider both true positive detections and false positive detections. A true positive is a detection made by the model that matches the ground truth label, while a false positive is a detection made by the model that does not match the ground truth label. A false negative, in contrast, is a part of the ground truth that was not predicted by the model. The true negative, however, cannot be measured in the context of object detection [16]. As a result, metrics such as accuracy and specificity cannot be reported in this experiment.

The mAP is a metric used to evaluate the performance of object detection algorithms, including YOLO. It is a measure of the overall accuracy of an object detector. It is computed as the mean of AP over different intersection over union (IoU) thresholds and different classes [17]. The AP is computed by the area under the curve between the interpolated precision (p_{interp}) and recall level (r), where $p_{\text{interp}}(r)$ represents the highest precision at any recall level ($r' \geq r$) as described in below equation:

$$p_{\text{interp}} = \max_{r' \geq r} P(r)$$

For example, $\text{mAP}_{0.5}$ is the area under the curve between precision and recall when the IoU threshold is set to 0.5 or 50%. The IoU threshold determines the level of overlap between the predicted bounding box and the ground truth bounding box. The $\text{mAP}_{0.5}$ can be calculated as in the equation below.

$$\text{mAP}_{0.5} = \frac{\sum_{c=1}^n \text{AP}_{0.5}^c}{n}$$

In this experiment, $\text{mAP}_{0.5}$ is the average AP for classes 1 to n , where n is equal to 7 in this study. The mAP metric can be represented across multiple ranges, such as $\text{mAP}_{0.5}$ to $\text{mAP}_{0.75}$, which is the average value calculated incrementally by steps of 0.05.

$$\text{mAP}_{0.5:0.75} = \frac{\sum_{\text{iou} \in \{0.5, 0.55, \dots, 0.75\}} \text{mAP}_{\text{iou}}}{10}$$

In simple terms, the AP measures the precision of the object detector in detecting objects and counting them correctly and the mAP calculates the average of these precision scores over different classes and IoU thresholds. A higher mAP score indicates a better performance of the object detector.

Caries Detection Experiments Using YOLO Models

A total of 2400 radiographs were used in the training of YOLOv3 and YOLOv7 models. The parameters were varied following the procedures outlined in Fig. 1. For testing the models, a total of 175 radiographs were utilized. The study aimed to evaluate the effect of the following on the performance of the YOLO models:

1. Input Image Size

In accordance with the default settings, YOLOv3 permits a maximum input of 608×608 pixel images for its operation, while YOLOv7 is limited to a maximum of 1280×1280 pixel images or a smaller size of 640×640 pixel images. This study hypothesized that dental caries details might be more explicit in larger images. To assess the impact of image size, YOLOv3 was established with a maximum image input of 608×608 pixels, while YOLOv7 was established with an input of 640×640 pixel images (YOLOv7_640).

2. Area of Detection (IoU) Thresholds

The IoU represents the area of detection that should be as close as possible to the ground truth obtained by experts. In other words, an IoU close to 100% is considered a perfect detection, which is unlikely in reality. However, as previously reported, YOLOv3 was implemented with an IoU threshold of 50 and 75% for detecting caries in bitewing radiographs [12]. Similar IoU thresholds and settings were applied to the YOLOv7 model.

3. Confidence Thresholds

The confidence threshold represents the cutoff point for the model to present its certainty in the output. The confidence threshold ranged from 0 to 1. When the confidence threshold is set to 0, all detected objects are counted. For instance, if the threshold is set as low as 0.001, the model will keep all objects with a certainty above 0.001 or 0.1% identified confidence. On the other hand, if the threshold is set to 0.5, the model will keep objects of which it is 50% confident in correctly identifying. Given its more sophisticated convolutional layers, YOLOv7 was expected to perform better and thus, a comparison was made within YOLOv7 with confidence thresholds of 0.001 and 0.5.

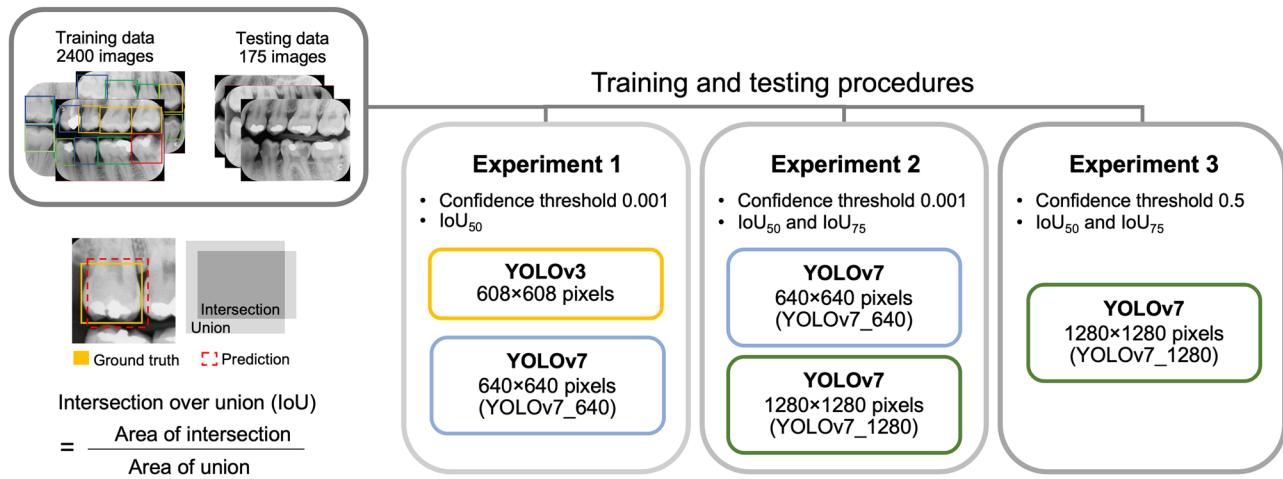


Fig. 1 Schematic diagram illustrates intersection over union (IoU) operations and the experiments conducted in this study

All performances of both models were compared using Wilcoxon's signed rank test and a *p*-value of less than 0.05 was considered statistically significance.

Results

The quantitative data for each category in each dataset is presented in Table 1.

Experiment 1: A Performance Comparison Between YOLOv3 and YOLOv7 Using Comparable Input Image Size

Under a similar setting, with an IoU of 50 and a default confidence threshold of 0.001, the precision, *F*1-score and mAP of the YOLOv3 (0.27; 0.38; 0.46, respectively) were significantly lower than those of the YOLOv7_640 (0.56; 0.56; 0.56, respectively). However, the recall presented significantly lower, as illustrated in the scattered plot in Fig. 2. Based on these findings, further experiments were conducted using YOLOv7.

Experiment 2: YOLOv7 with Various Image Sizes and IoUs

In the second experiment, we assessed the performance of YOLOv7 using two different image sizes: 640×640 pixels (YOLOv7_640) and 1280×1280 pixels (YOLOv7_1280). The IoUs were configured to 50% (IoU₅₀) and 75% (IoU₇₅), with the default confidence threshold set to 0.001. Statistical analysis demonstrated no significant difference in the overall mAPs between the two image sizes (*p*=0.866). Furthermore, the mAP results indicated no significant difference between the two image sizes when employing different IoUs (IoU₅₀ vs. IoU₇₅; YOLOv7_640, *p*=0.799; YOLOv7_1280, *p*=0.735), as illustrated in Fig. 3. The precise outcomes of precision, recall and *F*1-score for each image size and IoU can be found in the Appendix section (Figs. 7 and 8).

Experiment 3: YOLOv7_1280 with a Confidence Threshold of 0.5

After conducting a thorough analysis and finding no statistical differences in both image sizes and IoU, we proceeded with

Table 1 Quantitative data of each category in each dataset

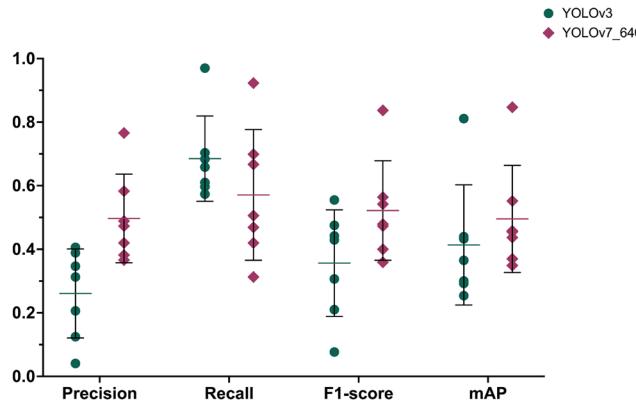
	Images	Data (teeth)					
		No caries	RA1	RA2	RA3	RB4	RB5
Training	2144	10,276	1495	2115	2337	642	463
Validation	256	1169	143	213	231	67	41
Testing	175	668	102	152	170	45	36
Total	2575	12,113	1740	2480	2738	754	540
							408

RA initial stages, RB moderate stage, RC Extensive stages according to the International Caries Classification and Management System (ICCMS™)

Validation and testing results between YOLOv3 and YOLOv7 for 640×640-pixel (YOLOv7_640) input bitewing radiographs

a.

Validation results

**b.**

Testing results

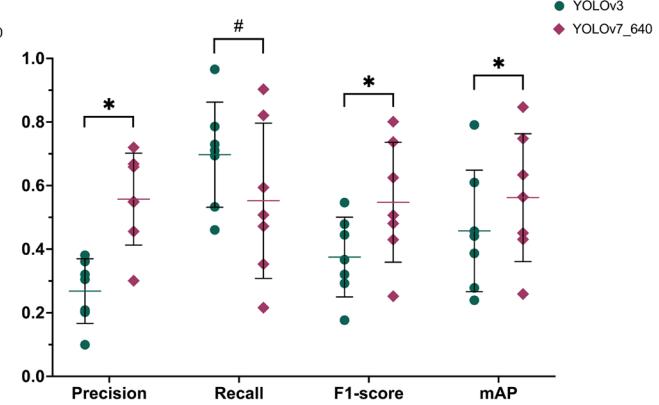


Fig. 2 Depicts the validation and testing outcomes comparing YOLOv3 for 608×608 pixels and YOLOv7 for 640×640 pixels (YOLOv7_640) input bitewing radiographs at an IoU of 50 and a confidence threshold of 0.001. The asterisk (*) denotes a statistically

significant increase in precision, *F1*-score and mean average precision (mAP) achieved by YOLOv7_640. Conversely, the hash symbol (#) denotes a statistically significant decrease in recall for YOLOv7_640

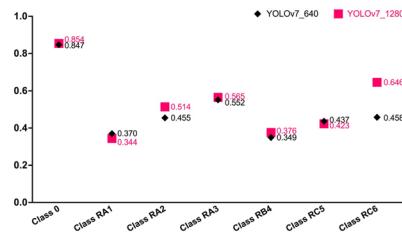
the YOLOv7_1280 model, using IoU_{50} , IoU_{75} and confidence thresholds of 0.001 and 0.5. Our findings indicated that when increasing the confidence threshold to 0.5, the precision of IoU_{75} significantly increased. However, the recall of both IoUs significantly decreased, resulting in a decreasing trend of *F1*-score. Increasing the confidence threshold led to significantly lower mAPs in both IoUs, as shown in Fig. 4a. The precision-recall curves, which represented the mAPs, displayed similar patterns for all classes, except in classes 0 and RC6 (as depicted in the appendix section; Fig. 9). This finding is

illustrated in detail in Fig. 4b–d in which shows that the precision, recall, *F1*-score of both IoUs were similar, except in class 0 and RC6, where IoU_{50} were significantly higher than IoU_{75} when increasing confidence threshold to 0.5. The bounding boxes generated by YOLOv7 under IoU_{50} and a confidence threshold of 0.001, in comparison to the bounding boxes annotated by experts, are depicted in Fig. 5. Furthermore, to gain further understanding of the model's perception and classification of carious teeth, we employed Grad-CAM visualization [18], as shown in Fig. 6.

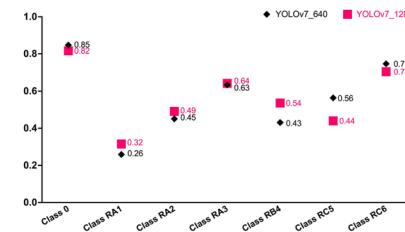
Mean average precision (mAP) of YOLOv7 models with image resolution of 640×640-pixel (YOLOv7_640) and 1280×1280-pixel (YOLOv7_1280) under IoU_{50} and IoU_{75} , and confidence threshold of 0.001

a.

Validation results

**b.**

Testing results of YOLOv7 under IoU_{50}

**c.**

Testing results of YOLOv7 under IoU_{75}

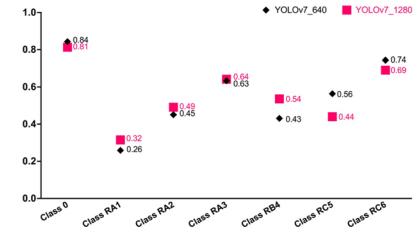


Fig. 3 Displays the mean average precision (mAP) of YOLOv7 models with an Intersection over Union (IoU) of 50 and 75, at a confidence threshold of 0.001. The mAP is measured for bitewing

radiographs with input resolutions of 640×640 (YOLOv7_640) and 1280×1280 pixels (YOLOv7_1280)

Performance of YOLOv7 model with image resolution of 1280×1280 pixels under IoU₅₀ and IoU₇₅

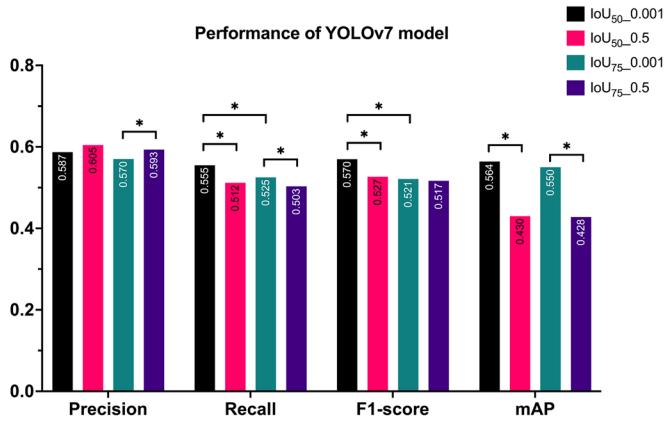
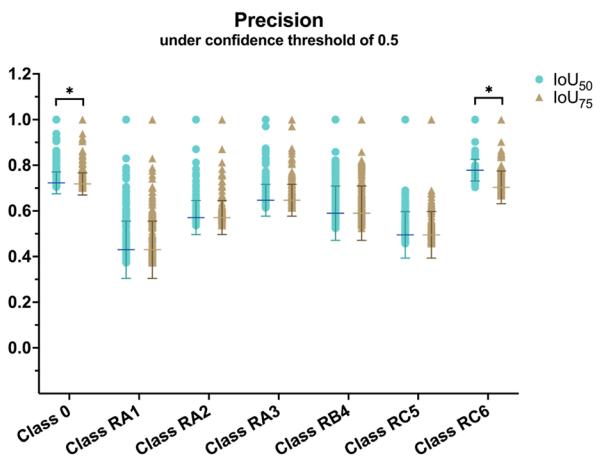
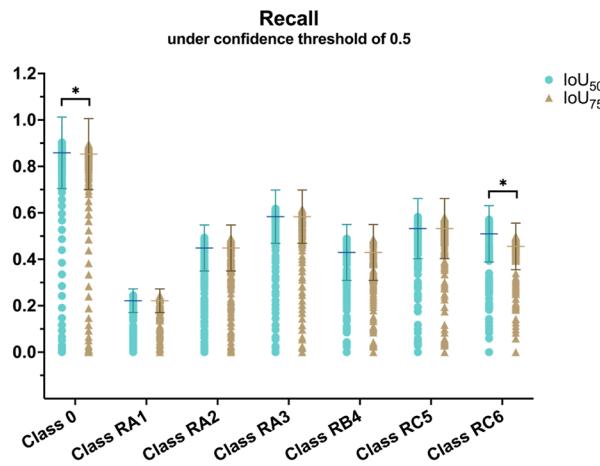
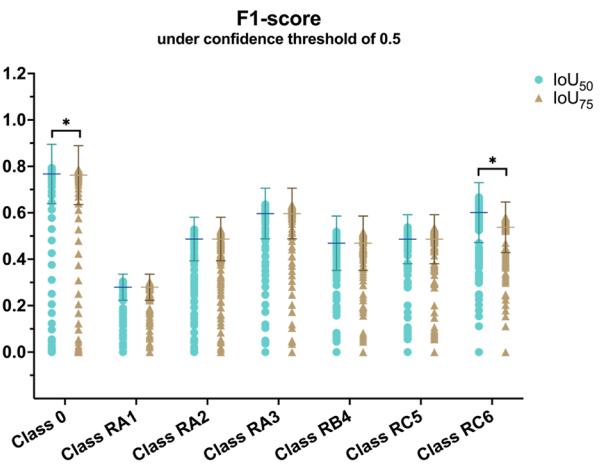
a.**b.****c.****d.**

Fig. 4 Depicts bar graphs illustrating the precision, recall, *F*1-score and mAP of the YOLOv7 model using an image input size of 1280×1280 pixels. The evaluations are performed under IoU₅₀ and IoU₇₅ with confidence thresholds of 0.001 and 0.5 (a). Additionally,

the precision, recall and *F*1-score plots (b, c, d) representing the mAP for caries detection in each ICCMS™ classification are presented for both IoUs. The results reveal significant differences in class 0 and class RC6, denoted by the asterisk (*) notation

Discussion

This study aimed to evaluate the performance of YOLOv7 in various scenarios for caries detection in bitewing radiographs. With enhancements made to improve the detection of small objects, the overall performance of YOLOv7 surpassed that of the previous version, YOLOv3, which was compared in this study. The utilization of deep learning YOLO models for radiograph analysis proves beneficial for the automated detection and analysis of caries-related radiographic features. These models can effectively aid dentists in identifying and localizing dental caries, thereby enhancing the diagnostic process.

In this case, mAP is the metric that takes into account the precision-recall trade-off and evaluates the detector's performance across all classes and different levels of confidence thresholds [19]. Precision and recall are inverse metrics; when the confidence threshold is subtle, such as 0.001, recall tends to be high (close to 1), while precision is low due to false alarms, as presented in Fig. 2. Therefore, there is a trade-off between the precision and recall of YOLOv3 while those of YOLOv7 were in harmony. Our findings revealed that the mAP of YOLOv7 was significantly higher than that of YOLOv3 under the same parameter setting. This is because YOLOv7 has been scaled up the depth of the computational

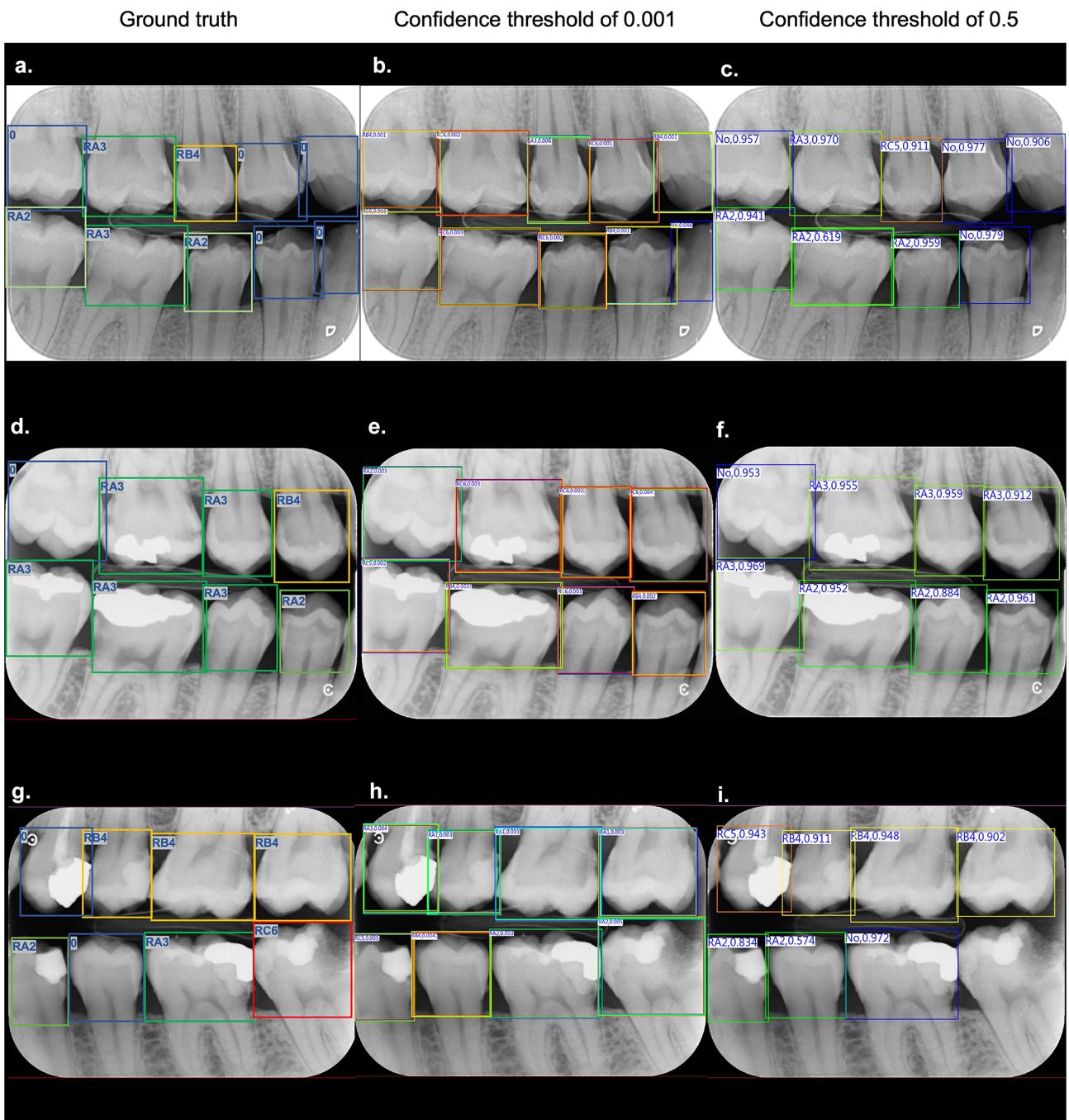


Fig. 5 Displays examples of bitewing radiographs with bounding boxes that were annotated by radiologists based on the ICCMS™ radiographic scoring system (**a, d, g**). The middle column (**b, e, h**) demonstrates the bitewing radiographs with bounding boxes that were created by the YOLOv7 model with IoU_{50} and a confidence threshold of 0.001. This resulted in multiple bounding boxes covering each tooth, indicating that the model kept all the predictions with a confidence score above 0.1%.

block by 1.5 times and the width of the transition block by 1.25 times. The former helps to increase the speed and accuracy of the model, while the latter involves increasing the number of layers or channels in the block, which can help to

improve the performance of the model by increasing its capacity to learn complex features [20]. Therefore, YOLOv7 is able to learn more complicated features, such as the radiographic appearance of carious lesions or any other dental lesions.

dence score above 0.1%. The right column (**c, f, i**) shows the bitewing radiographs with bounding boxes that were created by the model with a confidence threshold of 0.5, resulting in a single bounding box covering each tooth with a confidence score above 50%. It is noteworthy that the model removed the bounding box that covered the left mandibular second molar (**i**) as the confidence score did not exceed 50%

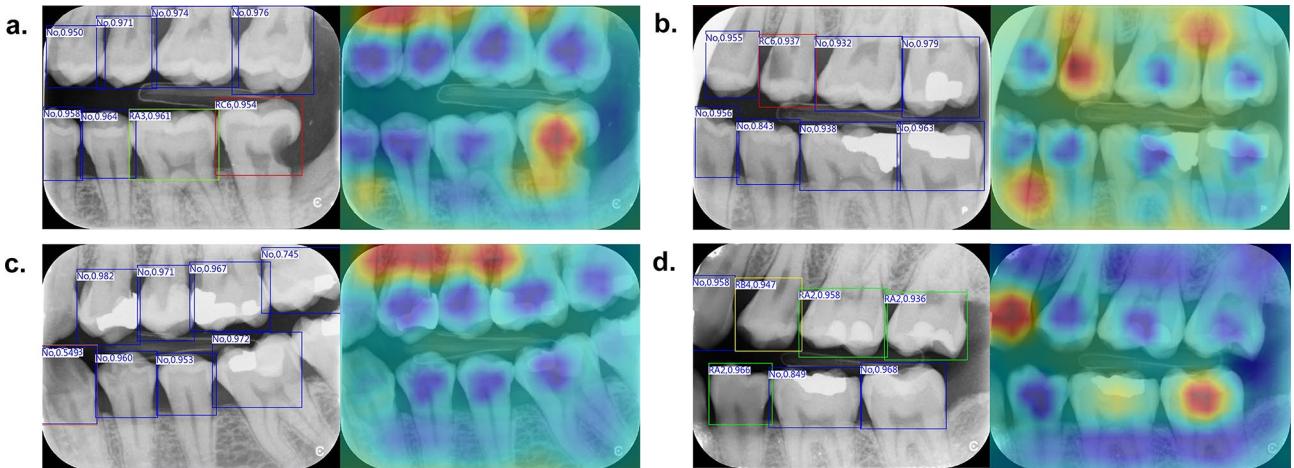


Fig. 6 Demonstrates the utilization of Grad-CAM to represent the focal points of the YOLO model. For instance, accurate positive detections are depicted as highlighted red spots in the carious teeth class RC6, specifically in the left mandibular second molar (**a**) and left maxillary second premolar (**b**). Conversely, false positive results are observed in the form of a double bounding box for class 0 and

RC5, located in the left mandibular canine, where Grad-CAM does not display any highlighted spot (**c**). Additionally, a false negative result is illustrated in the left mandibular second molar, corresponding to an intense highlighted spot in Grad-CAM (**d**). However, this radiograph exhibits a false positive result in the left maxillary second molar as class RA2, attributed to the presence of restoration

Despite using comparable image sizes, the results indicate that YOLOv7 exhibited superior performance for caries detection under similar conditions. This improvement could be attributed to the “label assignment” process, which involves identifying and assigning objects to a particular class, first introduced in the Faster R-CNN object detector algorithm [21]. In this study, the predefined category was the caries depth level. A recent study has demonstrated that YOLOv7 presented 75% fewer parameters, 36% less computation and yields a 1.5% higher AP than YOLOv4 [13].

As YOLOv7 is designed to work with the maximum image size of 1280×1280 pixels, we hypothesized that YOLOv7 would perform better in the larger image sizes with higher pixel numbers for detecting and classifying caries at various depths. As expected, the findings revealed that the larger image size presented superior precision, recall and $F1$ -scores in some classes. Particularly for detecting small lesions such as initial caries (classes RA1, RA2 and RA3), the mAPs of YOLOv7_1280 with IoU_{50} were higher than those of YOLOv7_640 in detecting these initial carious lesions. On the contrary, advanced caries (RC5 and RC6 classes) were better detected when using YOLOv7_1280 with IoU_{75} , as presented by the higher mAPs. Identifying advanced carious lesions are crucial since they are clinically visible and cavitated. Additionally, using full-scale images would be beneficial and could increase the model’s performance without requiring image manipulation. A recent report on real-time dental instrument detection found that the AP for dental instrument detection increased by 17.3% when annotating a specific part of the instrument compared to labeling the entire instrument using an input image of 1280×1280 pixels

[22]. The utilization of high resolution (1280×1280 pixels) allowed for more detailed image analysis, resulting in precise identification of caries regions. This can contribute to more accurate diagnoses and treatment planning.

Based on our previous findings, we proceeded with YOLOv7 using an image input of 1280×1280 pixels. However, we altered the IoUs to 50% and 75% and increased the confidence threshold to 0.5. Our findings revealed that increasing the confidence threshold improved precision in both IoUs, but the improvement was more significant in IoU_{75} . In applications such as medical image analysis, high precision is often critical for ensuring accurate diagnosis and treatment. This is because a false positive detection in medical imaging could lead to unnecessary treatments or interventions, while a false negative detection could result in a missed diagnosis, delayed treatment, or even patient harm. By setting a higher confidence threshold in object detector algorithms, the number of false positive detections can be reduced, thereby improving the precision of the model. However, this can lead to a lower recall or the number of true positive detections, leading to missed detections of essential features in the image. As presented in Fig. 4a, the recall values of both IoUs decreased when altering the confidence threshold to 0.5. In medical imaging, avoiding false positives is often more important than detecting every possible feature in the image. Therefore, these applications often use a higher confidence threshold to prioritize precision over recall. This can help ensure that the model only detects features that are highly likely present in the image while minimizing the risk of false positives.

The confidence threshold is a parameter in object detection algorithms establishes the minimum confidence level necessary for a detection to be considered valid. It represents the minimum probability or score necessary for an object to be deemed a valid detection. Typically set between 0 and 1, a higher threshold indicates a greater confidence requirement for considering a detection valid. For example, if the confidence threshold is set to 0.5, any detection with a probability score below 0.5 will be rejected and not considered valid. This helps to filter out false positives and ensure that only high-confidence detections are reported. The YOLOv4 and YOLOv5 models reached maximum mAP at a threshold of 0.5, so the confidence threshold for YOLOv7 was set at 0.001 and 0.5 [23]. YOLOv7 is designed to work with a confidence threshold of 0.001 by default, meaning that only objects with a confidence score of 0.1% or higher will be reported. Objects with a confidence score below 0.001 will be ignored and unreported as detections. This could result in many false positives, which are detections reported as carious lesions but are not actually so. Therefore, the output of confidence threshold 0.001 showed multiple bounding boxes containing predictions with various confidence scores, as shown in Fig. 5. This result might instead be a burden for general practitioners to differentiate the unnecessary detections. Conversely, when we set the confidence threshold to 0.5, the model considered the carious lesions with a confidence score of 50%. However, this could also increase the false negatives that some carious lesions might be misclassified or undetected due to a low confidence score.

The significant decrease of mAP after increasing IoU and confidence threshold would result from rejecting an ambiguous decision with low confidence scores. Looking closely at the scattered plots (Fig. 4b–d), there were different distributions in class 0 and class RC6 due to the radiographic appearance difference between non-carious teeth and large carious teeth. This led to the model being able to detect the lesion easily even if we increased the area of overlap and the confidence thresholds. These findings were supported by the evidence of the improved lead-guided label assignment strategy of the YOLOv7-E6E model by 0.3–0.6% improvement. To clarify, the lead head is the main detection head that predicts the objects' bounding boxes and class probabilities. It processes the entire image and extracts features at different scales, which are then used to predict the bounding boxes and class probabilities. Additionally, the lead head works with the auxiliary head to improve object detection accuracy in the YOLO model. The auxiliary head is used to improve small object detection by being attached to the pyramid in the middle layer of the intermediate layers of the lead head for training. It is used to detect small objects that the lead head might miss. The auxiliary head predicts small objects' bounding boxes and class probabilities and is designed to operate at a smaller scale [24]. The output of the auxiliary head is combined with the output of the lead head to produce the final detection results. YOLOv7 allows each pyramid of lead head to still get

information from objects with different sizes [13]. By leveraging the capabilities of YOLOv7, it can assist dentists in detecting conditions that may be missed or overlooked. This can aid in early detection and intervention, improving the overall effectiveness of dental care.

The study findings revealed several limitations related to the unbalanced data distribution within each class. Specifically, there were a high number of sound tooth samples compared to a smaller number of enamel caries samples. The model might become biased towards detecting non-carious instances and struggle to accurately identify and distinguish initial carious teeth. It was observed that the YOLOv7 model's effectiveness could potentially be enhanced by incorporating a larger amount of carious data during the training process. However, this approach poses challenges, as it would require a time-consuming annotation procedure to label a large number of additional carious instances accurately. Additionally, acquiring a substantial amount of expert-annotated carious data might require involving more experts, further increasing the resource and time requirements. These limitations highlight the trade-offs and practical considerations in expanding the dataset for training deep learning models. While additional carious data would likely enhance the model's performance, the associated annotation efforts and expert involvement should be carefully considered in terms of feasibility and resource allocation. It is important for future studies to explore strategies to address the class imbalance issue effectively, potentially through data augmentation techniques or specialized algorithms that can handle imbalanced datasets.

Additionally, the presence of restorations introduces complexities that may result in false positive or false negative results. Our primary objective was to compare the performance of the model in detecting radiographic caries across various depths, regardless of the presence of restorations. Although the YOLOv7 model demonstrated decreased detection performance in the presence of restorations (as depicted in Fig. 6), we aimed to evaluate the performance of the YOLO model under diverse conditions, including those involving restorations commonly found in radiographs. Therefore, we deliberately included a substantial number of radiographs, regardless of the absence or presence of dental restorations, to ensure a comprehensive analysis. To overcome this limitation, further enhancements can be made to the YOLOv7 model by exploring specific techniques that address the challenges associated with restorations and secondary caries detection. These techniques may involve refining the model architecture, incorporating additional training data, or employing advanced image pre-processing techniques to mitigate the impact of restorations on the detection process. These efforts will lead to a more robust and accurate caries detection in dental radiographs, enabling improved performance in real clinical scenarios.

Conclusion

Based on the results, the YOLOv7 model is feasible for detecting carious lesions on bitewing radiographs. It outperformed YOLOv3 in terms of precision, $F1$ -score and mAP. It could perform well with either 640×640 or 1280×1280 pixels image input, resulting in a non-significant difference in mAP between the two image sizes or different IoUs. The model was found to have improved performance when the confidence threshold was increased to 0.5, although this led to a higher number of false negatives. Increasing the confidence threshold to 0.5 significantly increased the precision of IoU_{75} , but significantly decreased the recall and led to a significant decrease in mAP when increasing the confidence threshold from 0.001 to 0.5.

for both IoU_{50} and IoU_{75} . Overall, the YOLOv7 model shows promise in detecting carious lesions on bitewing radiographs, providing additional insights and improving the overall diagnostic process. Collaboration between dentists and these automated systems can enhance the accuracy and efficiency of dental care delivery.

Appendix

The results, presented in Fig. 7, showed that the precision of YOLOv7_640 was significantly lower than that of YOLOv7_1280 in classes RA1, RA2 and RA6. Additionally,

YOLOv7 model performance under IoU_{50} and confidence threshold of 0.001

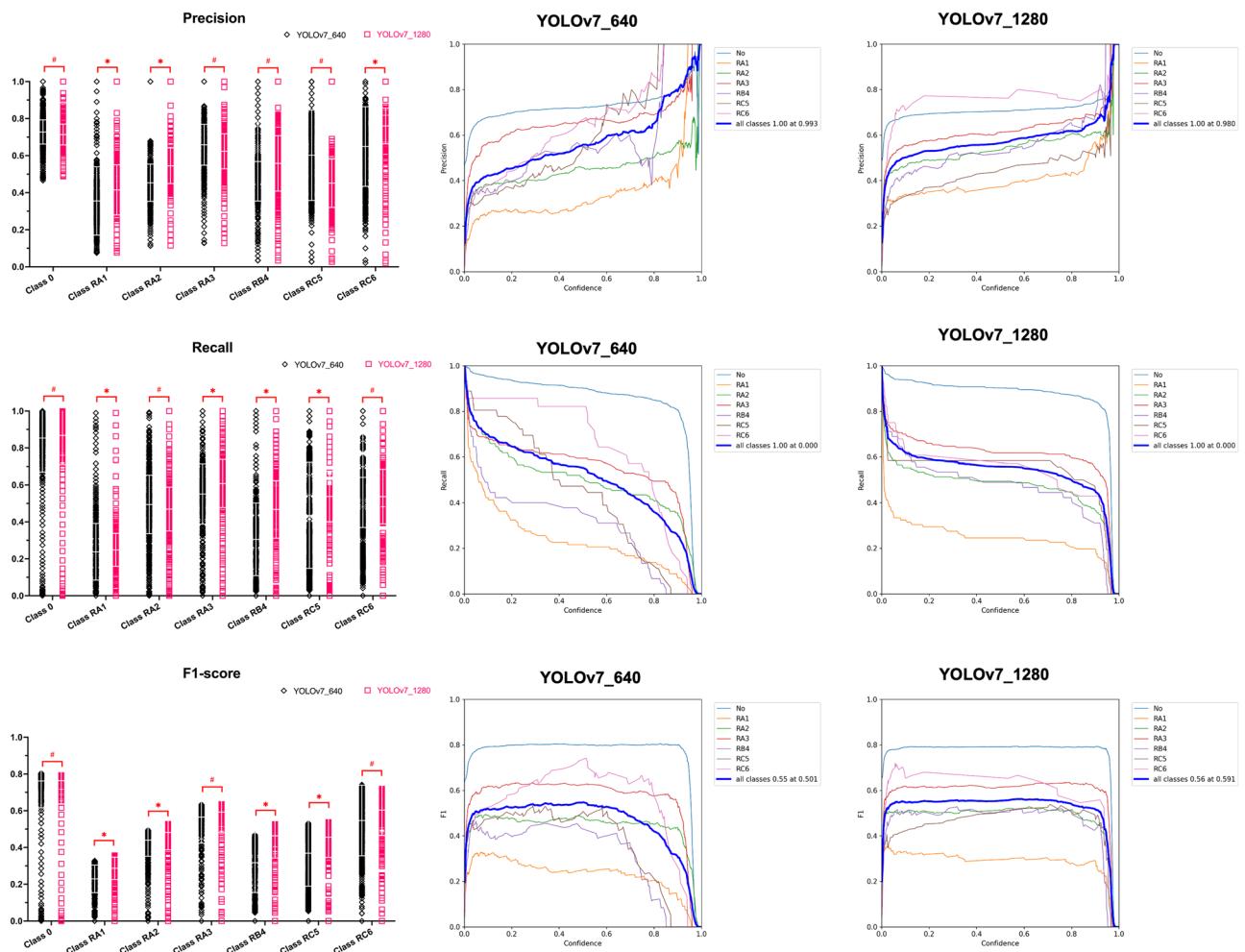


Fig. 7 Presents scattered plots illustrating the precision, recall and $F1$ -score of the YOLOv7 models under the IoU_{50} and confidence threshold of 0.001. Additionally, the performance plots for both YOLOv7_640 and YOLOv7_1280, corresponding to input bitewing radiographs of 640×640 and 1280×1280 pixels, respectively, are shown. These evaluations are specifically conducted for caries

detection, utilizing the ICCMSTM radiographic scoring system. In the plots, significantly higher values of YOLOv7_1280 are denoted by an asterisk (*), while significantly lower values are indicated by the hash symbol (#). The confidence thresholds range from 0 to 1.0, allowing for a comprehensive analysis of the model's performance

its recall was significantly lower in classes RA1, RA3, RB4 and RC5, resulting in a lower *F1*-score in these classes.

When the IoU was increased to 75% (IoU_{75}) and other hyperparameters were unchanged, the results were still significantly different between the two image sizes. The precision of YOLOv7_640 was still lower than that of YOLOv7_1280 in classes RA1, RA2, and RA6, and its recall was lower in most classes except for RA3 and RC6. This led to a lower *F1*-score in all classes, except for class RC6, as shown in Fig. 8.

The precision-recall curves of the YOLOv7 model at different IoU thresholds (50% and 75%) using a confidence threshold of 0.5 are shown in Fig. 9. The results demonstrate significant differences in the performance of class 0 and class RC6. Notably, the precision-recall curves associated with IoU_{75} exhibit significantly lower values compared to IoU_{50} , leading to a decrease in the mAP for these classes.

YOLOv7 model performance under IoU_{75} and confidence threshold of 0.001

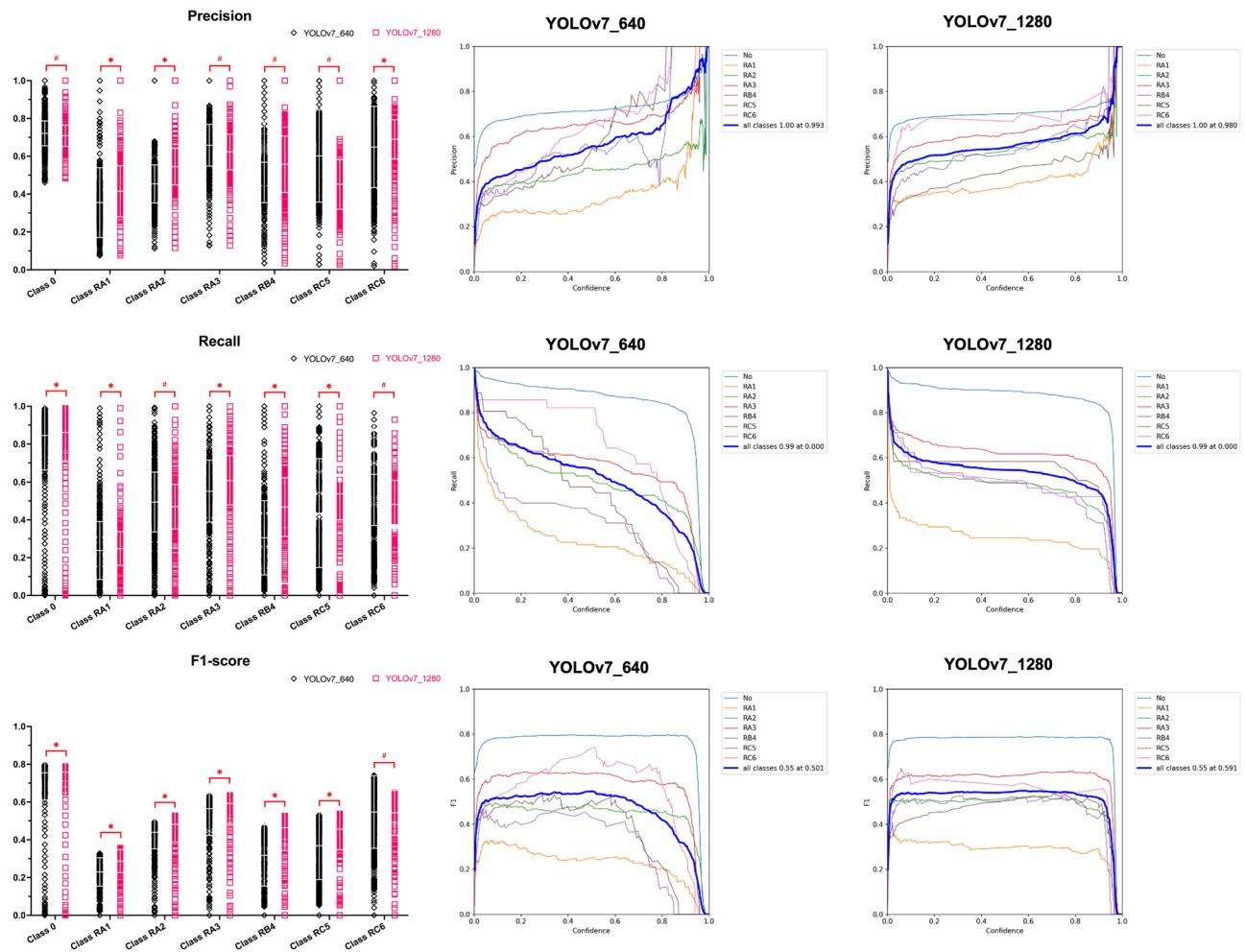


Fig. 8 Illustrates scatter plots showcasing the precision, recall and *F1*-score of YOLOv7 models, considering an IoU threshold of 75% (IoU_{75}) and a confidence threshold of 0.001. Additionally, the performance plots, encompassing confidence thresholds ranging from 0 to 1.0, are presented for both input bitewing radiographs of 640×640 (YOLOv7_640) and 1280×1280 (YOLOv7_1280) pixels, specifically for caries detection using the ICCMS™ radiographic scoring system. In the figure, an asterisk (*) denotes significantly higher values observed for YOLOv7_1280, while a hash symbol (#) indicates significantly lower values observed for YOLOv7_1280

(YOLOv7_640) and 1280×1280 (YOLOv7_1280) pixels, specifically for caries detection using the ICCMS™ radiographic scoring system. In the figure, an asterisk (*) denotes significantly higher values observed for YOLOv7_1280, while a hash symbol (#) indicates significantly lower values observed for YOLOv7_1280

Precision-recall (PR) curves of the YOLOv7 model under IoU_{50} and IoU_{75} and confidence threshold of 0.5

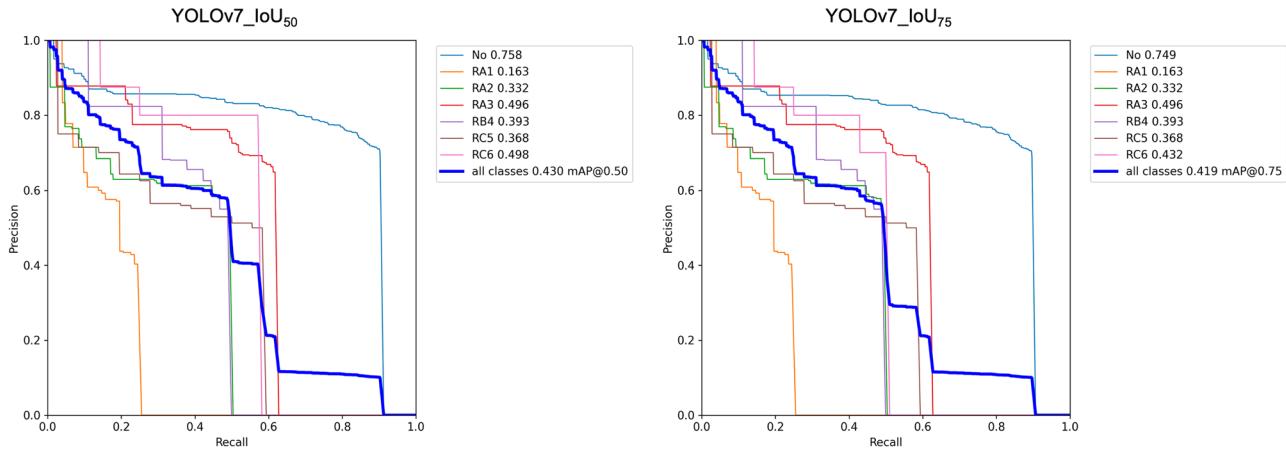


Fig. 9 Illustrates the precision-recall (PR) curves of the YOLOv7 model under two different Intersection over Union (IoU) thresholds, namely 50% (IoU_{50}) and 75% (IoU_{75}), while utilizing a confidence threshold of 0.5. The analysis reveals significant differences between

these two IoUs in the performance of class 0 and class RC6. Specifically, the PR curves associated with IoU_{75} exhibit significantly lower values, leading to a reduced mean average precision (mAP)

Author Contribution All authors contributed to the study conception and design. Material preparation and data collection were performed by Wannakamon Panyarak. The image annotation was performed by Wannakamon Panyarak, Arnon Charuakkra and Sangsom Prapayatasok. Deep learning implementation and analysis were performed by Wattanapong Suttpak and Kittichai Wantanajittikul. The first draft of the manuscript was written by Wannakamon Panyarak and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of Interest The authors declare no competing interests.

References

- Righolt AJ, Jevdjovic M, Marques W, Listl S. Global-, Regional-, and Country-Level Economic Impacts of Dental Diseases in 2015. *J Dent Res*, 97(5):501–507, 2018.
- Nascimento MM, Bader JD, Qvist V, Litaker MS, Williams OD, Rindal DB, et al. Concordance between preoperative and postoperative assessments of primary caries lesion depth: results from the Dental PBRN. *Oper Dent*, 35(4):389–396, 2010.
- Menem R, Barngkgei I, Beirut N, Al Haffar I, Joury E. The diagnostic accuracy of a laser fluorescence device and digital radiography in detecting approximal caries lesions in posterior permanent teeth: an in vivo study. *Lasers Med Sci*, 32:621–628, 2017.
- Cantu AG, Gehring S, Krois J, Chaurasia A, Rossi JG, Gaudin R, et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J Dent*, 100:103425, 2020.
- Mertens S, Krois J, Cantu AG, Arsiwala LT, Schwendicke F. Artificial intelligence for caries detection: Randomized trial. *J Dent*, 115:103849, 2021.
- Lee Y. Diagnosis and prevention strategies for dental caries. *J Lifestyle Med*, 3(2):107, 2013.
- Redmon J, Divvala S, Girshick R, Farhadi A, editors. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 779–788, 2016.
- Thanh MTG, Van Toan N, Ngoc VTN, Tra NT, Giap CN, Nguyen DM. Deep Learning Application in Dental Caries Detection Using Intraoral Photos Taken by Smartphones. *Appl Sci*, 12(11):5504, 2022.
- Sonavane A, Kohar R. Dental Cavity Detection Using YOLO. *Proceedings of Data Analytics and Management*: Springer, 141–152, 2022.
- Bayraktar Y, Ayan E. Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. *Clin Oral Invest*, 26(1):623–632, 2022.
- Pitts NB, Ismail AI, Martignon S, Ekstrand K, Douglas GV, Longbottom C, et al. ICCMS™ guide for practitioners and educators, 2014.
- Panyarak W, Suttpak W, Wantanajittikul K, Charuakkra A, Prapayatasok S. Assessment of YOLOv3 for caries detection in bitewing radiographs based on the ICCMS™ radiographic scoring system. *Clin Oral Invest*, 27(4):1731–1742, 2023.
- Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7464–7475, 2022.
- Zhang Z, He T, Zhang H, Zhang Z, Xie J, Li M. Bag of freebies for training object detection neural networks. *arXiv preprint arXiv:1902.04103*, 2019.
- Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. *Int J Data Min Knowl Manag Process*, 5(2):1, 2015.
- Padilla R, Passos WL, Dias TL, Netto SL, Da Silva EA. A comparative analysis of object detection metrics with a companion open-source toolkit. *Electronics*, 10(3):279, 2021.
- Manning CD. *Introduction to information retrieval*: Syngress Publishing; 2008.

18. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, editors. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision, 618–626, 2017.
19. Padilla R, Netto SL, Da Silva EA, editors. A survey on performance metrics for object-detection algorithms. 2020 international conference on systems, signals and image processing (IWSSIP). IEEE, 237–242, 2020.
20. Wang C-Y, Bochkovskiy A, Liao H-YM, editors. Scaled-yolov4: Scaling cross stage partial network. Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, 13029–13038, 2021.
21. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst*, 9199(10.5555):2969239–2969250, 2015.
22. Oka S, Nozaki K, Hayashi M. An efficient annotation method for image recognition of dental instruments. *Sci Rep*, 13(1):169, 2023.
23. Wenkel S, Alhazmi K, Liiv T, Alrshoud S, Simon M. Confidence score: The forgotten dimension of object detection performance evaluation. *Sensors*, 21(13):4350, 2021.
24. Jin G, Taniguchi R-I, Qu F. Auxiliary detection head for one-stage object detection. *IEEE Access*, 8:85740–85749, 2020.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.