scientific reports

Check for updates

OPEN

Advanced AI-driven detection of interproximal caries in bitewing radiographs using YOLOv8

Mahsa Bayati[1], Behrouz Alizadeh Savareh[2], Hojjat Ahmadinejad[3] & Farzaneh Mosavat[4]✉

Dental caries is a very common chronic disease that may lead to pain, infection, and tooth loss if its diagnosis at an early stage remains undetected. Traditional methods of tactile-visual examination and bitewing radiography, are subject to intrinsic variability due to factors such as examiner experience and image quality. This variability can result in inconsistent diagnoses. Thus, the present study aimed to develop a deep learning-based AI model using the YOLOv8 algorithm for improving interproximal caries detection in bitewing radiographs. In this retrospective study on 552 radiographs, a total of 1,506 images annotated at Tehran University of Medical Science were processed. The YOLOv8 model was trained and the results were evaluated in terms of precision, recall, and the F1 score, whereby it resulted in a precision of 96.03% for enamel caries and 80.06% for dentin caries, thus showing an overall precision of 84.83%, a recall of 79.77%, and an F1 score of 82.22%. This proves its reliability in reducing false negatives and improving diagnostic accuracy. YOLOv8 enhances interproximal caries detection, offering a reliable tool for dental professionals to improve diagnostic accuracy and clinical outcomes.

Keywords Artificial intelligence, Convolutional neural network, Deep learning, Machine learning, Digital bitewing radiography, Dental caries

Dental caries is a common chronic infectious disease process. If left untreated, it can lead to progressive health risks—including pain, infection, and tooth loss[1]. In addition to being troublesome to the individual, it places a burden biologically, socially, and financially on communities. While not immediately life-threatening[2], the active disease process involves the undermining of dental structures through demineralization, which may result in cavitation if not effectively managed[3]. Early detection of dental caries would have resulted in instant suspension of the process or even its reversal, hence avoiding aggressive treatments and reducing healthcare costs. However, early posterior proximal caries are challenging to diagnose clinically [4]. The current standard practices involve visual-tactile examination and radiographic imaging[3], with radiography playing a vital role in diagnosing caries in hard-to-visualize proximal surfaces; thus, caries on these surfaces usually go undiagnosed without radiography (1).

Radiography plays a crucial role in diagnosing dental lesions that are hidden or inaccessible. Panoramic, periapical, and bitewing radiographs are commonly used in clinical diagnosis[5]. Bitewing radiography, in particular, is effective in detecting proximal and occlusal caries that may not be identified during a clinical examination. It has high sensitivity and specificity for assessing the depth of caries[5]. Despite presenting a smaller area than panoramic radiography does, bitewing radiography is more accurate for evaluating lesions[6] and is considered the gold standard for diagnosing demineralized proximal caries. The combination of visual examination with bitewing radiography is considered a standard method for the detection of proximal caries[4]. Although studies have shown that the accuracy of these methods can vary widely, with reliability largely depending on the dentist's clinical experience. Sensitivity for detecting proximal caries has been reported to range between 39 and 94%[7].

Although there has been considerable development regarding maintenance methods and restorative techniques for dental caries over the last few decades, such progress has yet to be replicated in diagnostic methods because of the variability of teeth' anatomical morphology and the shape of restorations[8]. Recent developments in caries detection techniques, such as ultrasound, laser fluorescence (LF), digital fiberoptic transillumination imaging (FOTI), quantitative light-induced fluorescence (QLF), digital subtraction radiography (DSR), computed

[1]Post Graduate Student, Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Tehran University of Medical Sciences, Tehran, Iran. [2]PhD in Medical Informatic, Research and Development Manager, Department of Artificial Intelligence, Naaptech Co, Tehran, Iran. [3]MSC in IT Engineering, CEO, Naaptech Co, Tehran, Iran. [4]Associate Professor, Department of Oral & Maxillofacial Radiology, Faculty of Dentistry, Tehran University of Medical Sciences, Tehran, Iran. ✉email: F-mosavat@sina.tums.ac.ir

tomography with tuned aperture computed tomography (TACT), and electrical conductance measurement (ECM), are oriented toward overcoming these failures of clinical examination methods[8]. However, they are either limited or even impracticable for posterior primary proximal caries diagnosis, and occasionally, there are additional costs for devices[4]. Therefore, bitewing radiography remains the most reliable and widely adopted method for estimating caries activity in usual clinical practice[4].

Recent research has reported that although radiographs increase the sensitivity of visual examinations, they have lower specificity, resulting in many false positives and excessive, inappropriate treatment[9]. The diagnosis of tooth decay by radiography is subjective, with considerable interobserver variation in caries lesion detection[4]. The clinical experience of the dentist is a significant factor in the reliability of radiographs and their ability to make a diagnosis. Moreover, phenomena such as cervical burnout may further complicate the diagnosis of caries by bitewing radiography and, hence, lead to difficulty in differentiating such artifacts from actual caries. Thus, complementary tools using computational techniques increase accuracy in diagnosis and provide more robust evaluations[6].

One of the most interesting and promising developments in caries detection today involves AI systems. For two decades, machine learning techniques have been applied in scientific research for the diagnosis of dental entities. Traditionally, caries detection has been performed through manual and subjective operator interpretation, but this method has proven to be time-consuming and prone to human error in analyzing large image datasets. Recent efforts have successfully implemented deep convolutional neural networks (CNNs) based on deep learning models for learning models for this purpose[5].

The ultimate objective of AI is to make it self-learn, with no interference from a human brain, so that the systems may learn to predict the results in the future from the initial data fed into them (10). Deep learning techniques do not use features designed by humans and are highly generalizable. They have also achieved remarkable accuracy and sensitivity. These technologies have been applied in many fields—from the interpretation of medical imaging and diagnostic support systems, such as infection by *Helicobacter pylori* detection in gastrointestinal endoscopy[10], skin cancer screening[11], and the diagnosis of COVID-19 in tomographic images[12].

Dentistry is particularly suited for AI techniques because of the extensive use of medical imaging in dental treatment. Patients are routinely examined by a variety of imaging modalities—from simple intraoral periapical imaging to computed tomography—to look into the health and condition of the teeth and mouth for adverse effects[4]. Despite the dramatic reduction in the prevalence of caries during the last few decades, its diagnosis has remained challenging. AI, therefore, can automatically recognize dental caries, bone pathologies in dental radiographs, and jaw and facial pathologies with quick results and accuracy not influenced by the direct intervention of a dentist, allowing easier and more effective diagnoses[13]. During these last years, many studies investigated the application of AI to various dental disciplines, with one prominent area being the detection and diagnosis of dental caries. Indeed, previous studies of AI-based caries detection have included accuracy values ranging from 0.73 to 0.98[14]. This variability can be attributed to several factors, including different imaging modalities, varieties of AI models, and kinds of caries lesions considered, including enamel and dentin lesions. It is seen that dataset quality, the particular imaging technique to which the studies belonged- working with bitewing, periapical, or panoramic radiographs, and type of AI architecture implemented were contributing factors to the wide variation in diagnostic accuracy.

The current study belongs to pioneering works applying YOLOv8 for dental caries detection, demonstrating more improved performance metrics than earlier studies using previous versions of YOLO. Advanced features of YOLOv8, such as anchor-free detection mechanisms, the C2f. module, and mosaic augmentation, make it more accurate and faster at detection. Detailed performance metrics regarding different categories of caries provide a thorough assessment of the diagnostic capabilities of the model. Accordingly, the careful annotation process—done with expert radiologists and utilizing Roboflow for data management—takes much effort to return quality annotations, which are critical in model training and testing. This study was conducted to develop a deep learning-based AI model that can detect interproximal caries from bitewing radiographs via YOLOv8.

Materials and methods

Objective

The main goal of this study was to create and assess a deep learning AI model based on the YOLOv8 algorithm for detecting interproximal caries in bitewing radiographs. By focusing on improving diagnostic accuracy and reducing false negatives, we hope to provide dental professionals with a reliable tool that can significantly enhance early caries detection. This advancement is particularly important, as timely diagnosis can lead to better treatment outcomes and help patients avoid more invasive procedures down the line.

Dataset and preprocessing

This study was approved by the Tehran University of Medical Sciences Faculty of Dentistry (IR.TUMS.DENTISTRY.REC.1402.047) and all research was carried out in accordance with relevant guidelines and regulations. This study was a retrospective study, for which the requirement for informed consent was waived by the Institutional Review Board of.

Tehran University of Medical Sciences due to its data source and methods. A team of expert radiologists undertook the random collection of a dataset comprising 552 bitewing radiographs from the archives of an oral and maxillofacial radiology department. These radiographs were obtained via a Dentsply Rinn film-holding device, an Owandy dental X-ray machine, photostimulable phosphor (PSP) digital plates, and a Sordex Digora scanner. The manufacturer of the digital plate recommended a voltage of 60 KVP, a current of 6 MA, and an exposure time of 0.32 s. The images obtained were stored digitally in JPEG image format, with six to eight teeth included in the mean image. The database only includes permanent teeth and has no patient information,

including sex and age, as well as details of the medical condition. Radiographs that would safely be seen in a dentist's daily practice were selected. The images excluded are those that have poor quality, high-quality distortion, and overlapping proximal surfaces due to anatomic arrangements that maintain diagnostic accuracy. All radiographs were carefully reviewed by a well-trained oral and maxillofacial radiology assistant and a radiologist who has spent more than ten years in the clinic, thus being at the highest level of expertise in his or her review process. All the images are uploaded and assigned to the Roboflow app.

Annotation and data augmentation

In this study, we carefully annotated all visible interproximal dental caries seen on bitewing radiographs, using both polygon and smart polygon tools to accurately define the boundaries of each lesion, regardless of its severity. The polygon tool was particularly useful for outlining irregularly shaped lesions, while the smart polygon tool allowed us to create detailed and precise annotations.

We specifically chose to focus on segmenting enamel caries because early detection at this stage is crucial for effective intervention. Identifying caries when they are still confined to the enamel enables timely preventive measures, which can help avoid their progression into the dentin and pulp. Such advancement not only complicates treatment but can also lead to increased patient discomfort and higher costs. By concentrating our efforts on enamel detection, we aimed to enhance diagnostic accuracy and support better management strategies for caries.

In our annotation process, we concentrated on identifying carious lesions along with the enamel regions. Furthermore, during our evaluation in Python, we took the opportunity to analyze the images for distinguishing between caries that were limited to the enamel and those that had progressed into the dentin. This detailed approach allowed us to assess caries at various stages of progression, ensuring a comprehensive evaluation of the lesions.

To facilitate annotating, a universal tool for computer vision datasets, Roboflow, was used. Roboflow supports a wide range of formats for annotations, thereby allowing a researcher to focus on experiments rather than data conversion. It has comprehensive features in annotation, labeling, extensive preprocessing options, and augmentation techniques. These features improve the effectiveness and quality of dataset preparation, making it a useful tool for further computer vision research. First, there were disagreements regarding the identification of caries, or rather, whether caries was present or absent, and the size of the carious lesions; such disagreements were resolved by consensus between the assistant and the radiologist. In cases of persisting discrepancies, the radiographs were re-evaluated by a third independent observer. To increase dataset augmentation techniques, including rotation, scaling, and brightness changes were used. With these augmentation techniques, the dataset increased to three times its original size. Thus, the total set of images reached 1506 images. This dataset was then divided with allocations to train, validate, and test sets at percentages of 80%, 10%, and 10%, respectively. The following are the statistics according to the dataset: total number of caries regions: 6299, total number of enamel regions: 5015, number of caries regions confined within enamel areas: 2461, number of caries regions extending beyond enamel areas to the dentin region: 3838.

Model selection and utilization of YOLOv8

Choosing a robust model architecture is a critical factor in achieving effective image segmentation. This model ensures precise boundary delineation, reducing false positives and negatives while operating efficiently within the application's constraints. These benefits are crucial for many tasks, making model architecture a significant determinant of the feasibility, performance, and reliability of the segmentation results[15]. In this study, YOLOv8, the latest iteration in Ultralytics's You Only Look Once (YOLO) series, was selected because of its exceptional attributes. First, YOLOv8 demonstrates remarkable speed, significantly surpassing other state-of-the-art models. Second, it provides high accuracy, with a mean average precision (mAP) exceeding many competitors. Its advanced architecture includes a new backbone network and anchor-free detection head. Furthermore, YOLOv8's flexibility in supporting multiple tasks such as object detection, instance segmentation, and image classification within a unified framework renders it highly versatile[16]. YOLOv8 represents a substantial advancement in object detection, image classification, and instance segmentation tasks. Building on the foundations of YOLOv5, YOLOv8 incorporates numerous architectural enhancements and improvements in developer experience, thereby increasing its efficiency and accuracy for a wide range of computer vision applications. The model is continually being developed, ensuring long-term support and enhancements driven by community feedback.

YOLOv8 architecture

YOLOv8's architecture incorporates several enhancements over previous versions. It features an anchor-free detection mechanism, which predicts the center of objects directly rather than relying on predefined anchor boxes. This approach reduces the complexity of nonmaximum suppression (NMS) and speeds up the postprocessing of detections. Additionally, YOLOv8 replaces the traditional convolutional layers with a more efficient C2f. module, which concatenates all outputs from the bottleneck layers, enhances feature extraction and reduces the parameter count. This streamlined architecture improves accuracy and maintains a manageable model size, making it suitable for deployment on edge devices and in cloud environments. The architectural features of YOLOv8 include several vital enhancements. It introduces a revised stem with a smaller convolutional kernel, transitioning from a size of 6×6 to a size of 3×3, and a new backbone structure utilizing the C2f. module instead of the C3 module found in YOLOv5. The C2f. module, which is composed of multiple convolutions with residual connections, improves feature representation. Additionally, YOLOv8 employs mosaic augmentation during training, which combines four images to enhance object recognition under varying conditions. This augmentation is selectively applied and turned off during the final training epochs to optimize performance.

These architectural innovations collectively contribute to YOLOv8's superior performance in terms of accuracy and efficiency[17].

Data augmentation and training configuration

This approach enhances the robustness and generalization ability of the model by incorporating a variety of augmentation techniques in training [18]. These augmentations help prevent overfitting and improve the model's performance on unseen data. The augmentations used include random horizontal flipping, where each image has a 50% chance of being flipped horizontally to help the model learn invariance to the left–right orientation, which is particularly useful for symmetrical structures such as teeth; random rotation, where images are randomly rotated by ± 10 degrees to aid the model in handling slight rotational variations during image acquisition; and color jittering, where the brightness, contrast, saturation, and hue of the images are randomly adjusted within a range of ± 0.1, making the model robust to variations in lighting conditions and image quality. The YOLOv8 model was trained for 200 epochs to ensure adequate learning of features relevant to enamel and caries detection. The batch size was set to 16 to balance the memory constraints of the GPU with stable gradient updates. The initial learning rate was set to 0.005; a cosine annealing learning rate scheduler was used to find the best learning rate during training that would change the learning rate, according to a cosine, and eventually reduce it to zero, helping the model fine-tune toward the end of training. For the training and testing of the YOLOv8 model, a high-performance setup equipped with an NVIDIA GeForce RTX 2080 Ti GPU featuring 11 GB of dedicated memory, a potent processor in the form of a 24-thread Core i7 CPU to support high parallelization without losing much time on overhead, and 128 GB of memory, was used.

Evaluation metrics

The performance of the trained model was evaluated against standard metrics for segmentation tasks, including precision, recall, F1 score, false negative rate, and finally, the Dice coefficient. These metrics provide an all-rounded assessment of how well the model performs.

Our approach exploits the advanced capabilities of YOLOv8 segmentation, state-of-the-art efficient data augmentation techniques, and an optimized training configuration to ensure high accuracy and reliability in the automatic detection and segmentation of enamel and caries from bitewing radiographic images.

Result

Diagnostic performance

Table 1 presents the diagnostic performance of the trained network for caries detection, highlighting its high precision and recall. The model's performance metrics for overall caries detection, including precision, recall, and F1 score, are notably high, with a low false negative rate. When focused on caries confined to the enamel region, the model demonstrates even higher precision and recall, with a far lower false negative rate. The performance metrics for dentin caries detection, such as precision, recall, the F1 score, and the false negative rate, are also detailed in the table, highlighting the model's effectiveness across different detection tasks.

Evaluation metrics and curves

The performance of the trained network in enamel-decay segmentation is shown in terms of evaluation metrics and their diagrams.

- Precision–confidence curve: Fig. 1 displays the precision-confidence curve, plotting model precision as a function of the segmentation task confidence threshold. High precision means few false positives. The graph indicates how well this model works at different confidence levels, helping an optimal threshold selection for making predictions.
- Precision-recall curve: The precision-recall curve, also presented in Fig. 1, is a crucial evaluation metric for imbalanced datasets. This curve illustrates the trade-off between precision and recall for different threshold settings. A higher area under the curve indicates higher recall and precision, reassuring the audience that the model excels in detecting relevant instances without triggering many false alarms.
- Recall–confidence curve: Fig. 2 showcases the recall-confidence curve, plotting model recall as a function of the confidence threshold for segmentation tasks. High Recall means that the false negative rate will then be low. This curve informatively depicts how well the model can detect all relevant instances at different confidence levels.
- F1-confidence curve: Also shown in Fig. 2, the F1-confidence curve plots the harmonic mean of precision and recall, known as the F1 score, against different confidence thresholds for bounding box predictions. This stakeholder indicates the model's central tendency, which is weighted in terms of precision and recall, giving an excellent overall indicator of model performance.

Performance of the trained model	precision	recall	F1-score	False negative rate
All caries	84.83%	79.77%	82.22%	20.22%
Caries limited to enamel	96.03%	96.03%	96,03%	3.96%
Caries extended beyond enamel	80.06%	73.46%	76.62%	26.53%

Table 1. Performance Metrics of the Trained Model for Caries Detection.

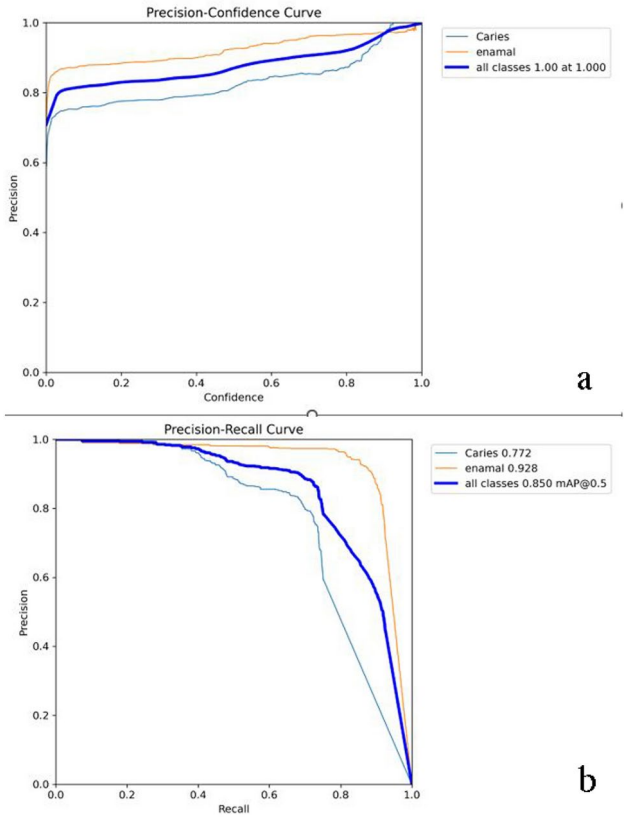

Fig. 1. a Precision–confidence curve for mask predictions. The curve demonstrates the model's ability to maintain high precision at various confidence levels. b: Precision–recall curves for the mask predictions. This figure highlights the model's effectiveness in balancing precision and recall.

- Precision–confidence curve: Precision–confidence curve for box predictions (Fig. 3) illustrates the precision as a function of the confidence threshold in detecting bounding boxes. This curve could help decide on the most appropriate setting for the applied threshold to minimize the false-positive rate.
- Precision–recall plot: Further a precision-recall plot presented in Fig. 3, which indicates how precision and recall change over varying confidence thresholds. This curve is especially relevant for tasks where balance is critical between false positives and false negatives.
- Recall–confidence curve: The recall–confidence curve presented in Fig. 4, displays the model's recall as a function of the confidence threshold for segmentation tasks. High recall indicates a low false negative rate. This metric provides insight into how good our model is at picking up all the relevant bounding boxes for a dataset (Fig. 4).
- F1-Confidence Curve for Mask Predictions: The F1-confidence curve for mask predictions, also illustrated in Fig. 4, shows how the F1 score varies for different confidence thresholds. It is based on a harmonic balancing metric between precision and recall, providing an overview of model performance for segmentation tasks (Fig. 4).

General dice coefficient

The overall general Dice coefficient reached 0.7949, reflecting a high degree of overlap between the predicted and ground-truth regions. Considering that caries regions are constrained within enamel regions, the Dice coefficient was 0.6854, reflecting moderate overlap. The Dice coefficient for caries regions, regardless of enamel region, equaled the general Dice coefficient, 0.7949, indicating that the quality of the overlap does not depend on the enamel constraint.

Visualization of segmentation results

Samples of enamel and caries segmentation with overlapping representations are shown in Fig. 5. These visualizations provide a clear view of the difference between the original images and the segmentation results, that is, how the model identifies and differentiates between enamel and caries areas. This visualization helps in evaluating the quality of segmentation performance.

Discussion

This study aimed to develop a deep learning-based AI model, using YOLOv8, to detect interproximal caries in bitewing radiographs. Early identification of enamel caries is especially important, as it allows for timely

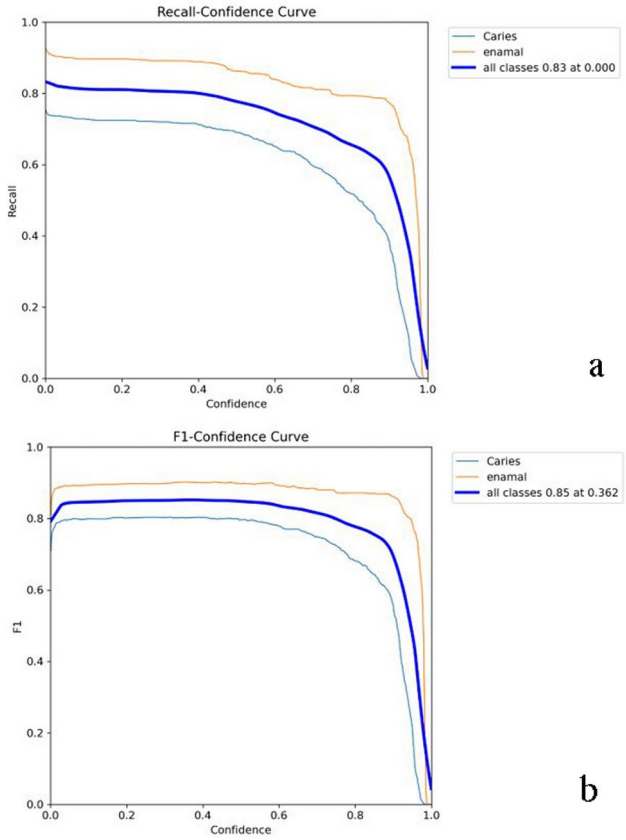

Fig. 2. (**a**) Recall–confidence curves for the mask predictions. The model has high recall, indicating its strong ability to detect true positives. (**b**): F1-confidence curve for box predictions. The curve shows the model's balanced performance across different confidence levels.

intervention before lesions reach the dentin, where treatment can become more complex, painful, and costly. Catching enamel caries early helps prevent disease progression, minimizes the need for invasive procedures, and conserves valuable resources that can be allocated to broader health services. With these benefits in mind, our model specifically focuses on detecting enamel caries.

Our dataset consisted of 1506 augmented images from an original set of 552 bitewing radiographs. This dataset size is greater than related to previous studies[6,19–21], and it has often returned improved performance metrics. For example, the model gives a precision of 84.83%, a recall of 79.77%, an F1 score of 82.22%, and a false negative rate of 20.22% for all caries detection. On the other hand, for dentin caries detection, the model had accuracy, recall, F1 score, and FNR values of 80.06%, 73.46%, 76.62%, and 26.53%, respectively. The performance improved so much when testing for caries limited to the enamel region that it achieved precision, recall, and an F1 score of 96.03%, with an FNR of 3.96%. These results portray the effectiveness of this model concerning the detection of incipient enamel caries, which is very important for early intervention and treatment. Following the findings of this study, a YOLOv8 model would significantly improve the accuracy and efficiency of caries detection in the clinical environment by providing an essential tool for dental practitioners. AI-based caries research in the literature can be divided into three primary categories: segmentation[19,21–23], classification[6,24,25], and detection[20,22,26–28]. In several earlier studies, U-Net was used to detect proximal dental caries in segmentation studies[19,21,23], with Cantu et al. reporting a recall of 75%[23]. Lee et al. achieved a precision of 63.29%, a recall of 65.02%, and an F1 score of 64.14%[21]. On the other hand, our YOLOv8 model performed better in terms of precision, recall, and F1 score, reflecting more robust performance in detecting carious lesions. Previous studies have relied on older versions of the YOLO algorithm for caries detection. Kunt et al. utilized YOLOv5-M and YOLOv5-L, recording both precision and recall rates between 74 and 79%, with an F1 score of 76%[29]. Bayraktar and Ayan used YOLOv3 and reported a recall of 72.26% for caries detection[26]. Chen et al. applied Faster R-CNN for proximal caries detection and achieved a recall of 72%[27]. In contrast, our YOLOv8 model has higher recall; therefore, it is more effective in detecting carious lesions by capturing more true positive cases. YOLOv8 has improved precision and recall over base variants, which means that it identifies positive classes with much discretion and a high amount of recall. Compared with YOLOv5, YOLOv8 effectively balances precision and recall.

Compared with several earlier studies, it provides high-performance metrics. Specifically, we can obtain an F1 score of 82.22% at the top compared with the F1 scores reported by Srivastava et al. (70.00%)[28], Lee et al. (64.14%)[21], and Kunt et al. (76%)[29]. A higher F1 score indicates a better balance between precision and

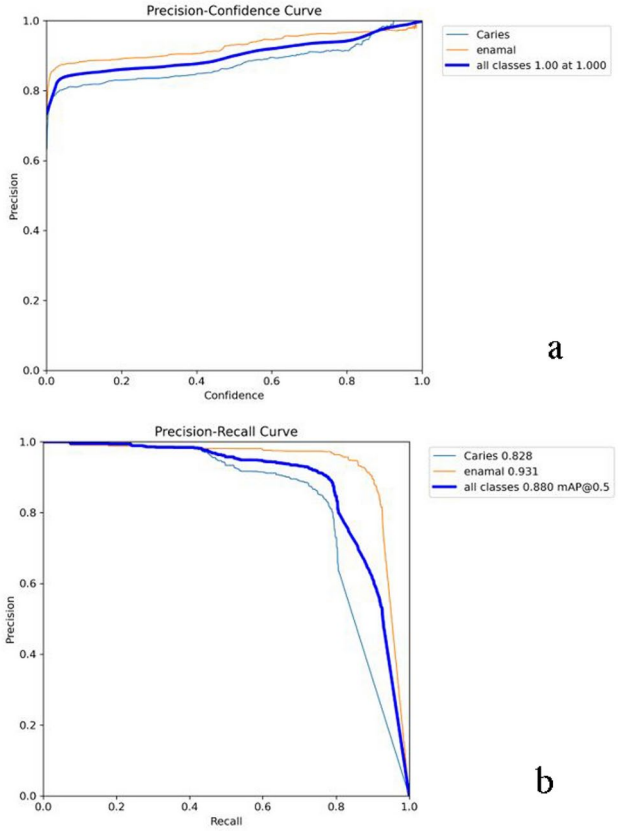

Fig. 3. (**a**) Precision–confidence curve for box predictions; The model maintains high precision across varying confidence thresholds. (**b**): Precision–recall curves for the box predictions; The high area under the curve indicates a strong balance between precision and recall.

recall, reflecting the model's effectiveness in correctly identifying true positives while minimizing false positives. Moreover, our recall of 79.77% outperformed the recall rates from Cantu et al. (75%)[23], Lee et al. (65.02%)[21], and Kunt et al. (74%)[29]. A higher recall reflects that the model is effective in classifying actual positive cases; therefore, fewer missing caries indicate an accurate diagnosis of caries. Furthermore, the precision is also significantly greater than those reported by Srivastava et al. (61.50%)[28], Lee et al. (63.29%)[21], and Kunt et al. (77% and 74%)[29]. High precision means that the most significant proportion of all positive predictions are true positives, which significantly reduces false positives and ensures that the identified caries is indeed present.

A systematic review by *BMC Oral Health* analyzed AI-based models for caries detection and reported a broad range of sensitivity values (0.44–0.86) (30). This variability may stem from several factors, including differences in radiographic modalities, annotation techniques, model training methods, and the varying expertise of the radiologists involved in the studies. In contrast, our study achieved a sensitivity of 0.92 for enamel caries, placing it at the upper end of this range. This underscores the effectiveness of the YOLOv8 model in delivering consistent and accurate caries detection. Our study also demonstrated high precision and recall rates for enamel caries, confirming the model's ability to identify early-stage caries, which is crucial for timely intervention and treatment. These findings align with the literature[5,6,30] and highlight the promise of deep learning models in enhancing the accuracy of dental caries detection. They also indicate that with adequate training and strong datasets, AI has the potential to minimize the variability in diagnosis that has been observed in earlier studies. The strong performance metrics noted in this study can be linked to the sophisticated architecture of the YOLOv8 model. This model features improved capabilities for extracting essential details and benefits from a more thorough training process than its predecessors. As a result, it effectively distinguishes between carious lesions in bitewing radiographs, supporting timely intervention.

Recent publications respect the outstanding development of the application of artificial intelligence in detecting enamel caries. However, the results still need to be divergent regarding accuracy and diagnostic reliability. W. Suttapack, Panyarak, et al. presented an adapted ResNet-18 model that reached an accuracy of 86.67%, with a sensitivity of 87% for enamel caries and an accuracy of 73.33%, with a sensitivity of 73% for dentin caries[31]. On the other hand, this research exhibited better performance in enamel caries detection with higher precision, recall, and F1 score than W. Suttapack et al. demonstrated the model efficiency in detecting early-stage enamel caries. Furthermore, we found that although our study was slightly better for recall in dentin caries, both studies very nearly showed sensitivity. In a study, Lian et al. demonstrated that their model, DenseNet121, had better performance than experienced dentists in detecting caries D1 lesions; it had higher sensitivity and better

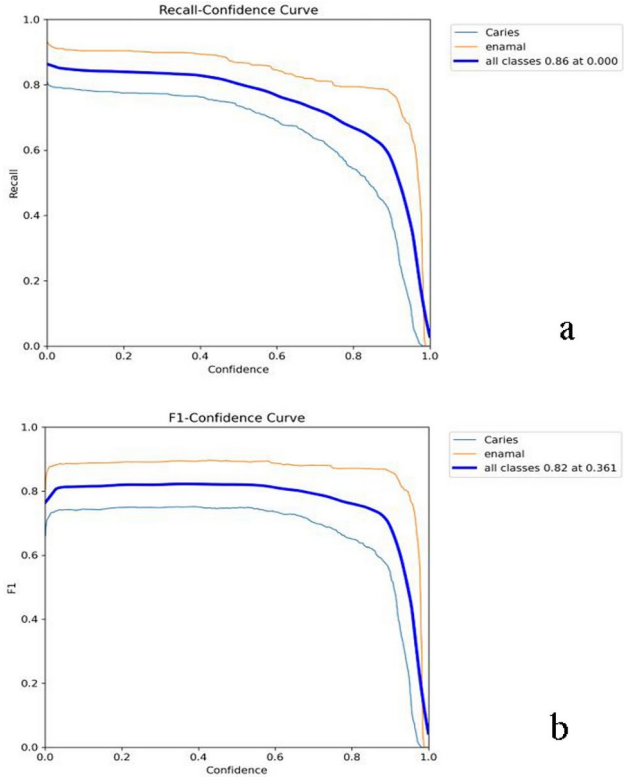

Fig. 4. (**a**) Recall–confidence curve for box predictions; The model demonstrates high recall, ensuring that most relevant instances are detected. (**b**): F1-confidence curves for the mask predictions; This curve highlights the model's overall performance in balancing precision and recall for segmentation.

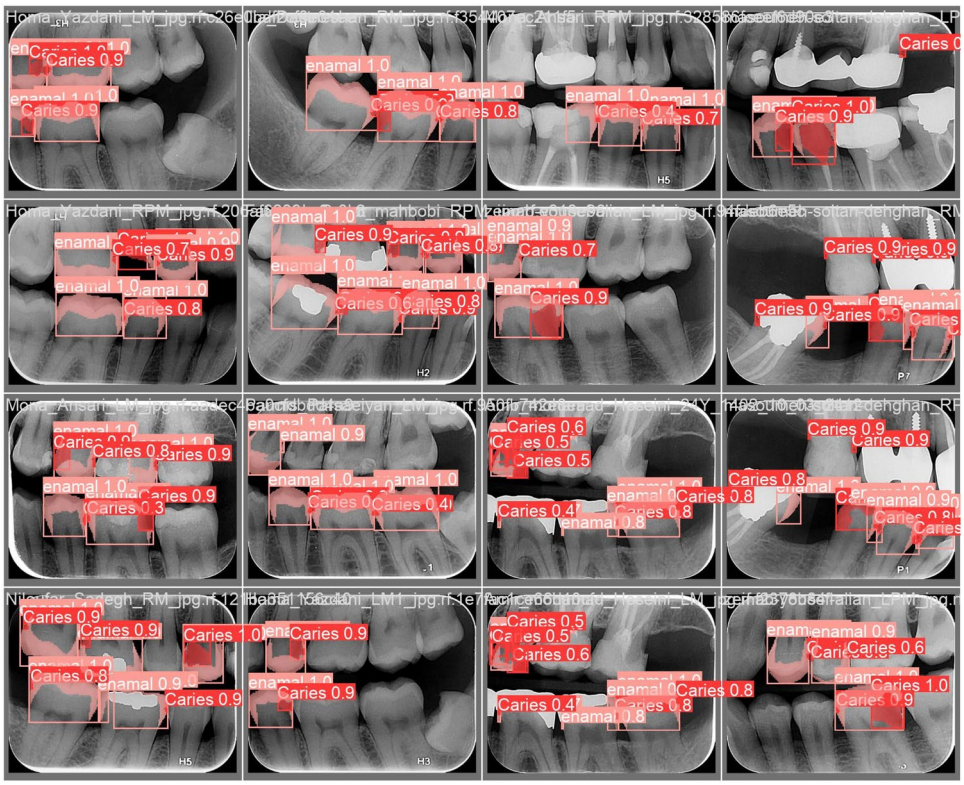

Fig. 5. Overlay comparison of the enamel and caries segmentation results.

performance for accuracy, recall, specificity, precision, NPV, F1 score, IoU, and Dice. This finding also indicated greater accuracy in detecting D1 lesions than in detecting more advanced caries. The variable performance of the dentists was measured against the consistency and reliability of the model. This could be due to the model being able to learn from subtle features on radiographs, which human dentists often cannot detect. At the same time, it was capable of differentiating them very easily. Other advantages noted regarding the model were consistency in performance and classification speed. The results revealed the possibility of improving caries detection and classification by combining the model results with the dentist's diagnosis. This study underscores AI's potential in revolutionizing dental imaging, offering more accurate and reliable caries detection[5]. Thus, for enamel caries detection, our YOLOv8 model had excellent performance, with precision and recall high enough for its results to be compared with those of previous studies[5,31], which further illustrates its effectiveness in identifying early-stage lesions. Our YOLOv8 showed extremely high precision and recall for enamel caries and a Dice coefficient of 0.6854 for the enamel lesion compared with the previous study's Dice coefficient of 0.663 for panoramic radiographs[5]. The stronger sensitivity of our model in detecting enamel caries (0.92) compared to its performance in identifying dentin caries can be understood through several factors. Enamel caries, while smaller, are surface-level lesions that tend to stand out more clearly in bitewing radiographs due to the sharp contrast between the enamel and surrounding tissues. This makes them more distinguishable and easier for the YOLOv8 model to detect, as it's particularly effective at identifying objects with clear boundaries and pronounced grayscale differences. In contrast, dentin caries affect a deeper and larger tooth area, often making them more diffuse and less defined. The reduced contrast between dentin and nearby tissues can make these lesions harder to spot. Moreover, dentin caries usually present more complex internal structures, making it difficult for the model to pick up on the subtle grayscale variations required for accurate detection.

In this study, the primary metric employed in evaluating the segmentation model was the Dice coefficient, which measures the intersection between the predicted and ground truth regions. To facilitate a more comprehensive evaluation and enable meaningful comparisons with a broader range of related studies, additional metrics were incorporated. Specifically, the False Negative Rate (FNR), F1 score, precision, and recall were calculated, providing a multifaceted assessment of model performance across various relevant dimensions. These metrics were chosen to align the evaluation process with metrics commonly reported in related literature, ensuring both robustness and comparability in the assessment of segmentation accuracy.

In the current study, the general Dice coefficient for caries detection was 0.7949, with a value of 0.6854, specifically for enamel caries. These results indicate a greater overlap in detecting carious regions compared with a ResNet-based study that reported a Dice coefficient of 0.663 via panoramic radiographs[5]. This finding suggests that the high resolution and specificity of bitewing radiographs allow our model, or for that matter, any model, to perform well in the detection of both general and enamel caries interproximal, which contradicts panoramic imaging with low resolution and overlapping surfaces affecting interproximal caries detection.

In this study, the false negative rate (FNR) aligns with findings from similar research[32], particularly regarding the challenges involved in detecting dentin compared to enamel caries. For instance, a study utilizing YOLOv5 reported an FNR of 15% for enamel caries and 21.5% for dentin,[32]. Our model achieved an FNR of 3.96% for enamel and 26.53% for dentin. This difference likely results from the clearer boundaries and higher contrast of enamel, which make it easier to identify. Reducing the FNR further could involve refining model architecture and training with datasets specifically designed to capture the nuances of complex dentin lesions, potentially enhancing detection accuracy for these more challenging cases.

The YOLOv8 architecture, with its sophisticated feature extraction abilities, performs exceptionally well in capturing the distinct outlines of enamel lesions. However, detecting more advanced dentin caries, which involve broader and more complicated textures, may need additional fine-tuning. This likely accounts for the slight discrepancy in the model's performance between detecting enamel and dentin caries in our study.

This is one of the first applications to adapt a model by YOLOv8 for interproximal caries detection in bitewing radiographs. More specifically, this work makes use of a variant architecture of the YOLOv8 model that supports real-time object detection, which is advantageously very fast for application within clinical settings where speed and accuracy are factors in diagnosis. YOLOv8 is an architecture that is both more advanced and efficient; however, it also involves comprehensive data augmentation, such as random horizontal flips, rotation, and color jittering. This helps greatly enhance model robustness and generalizability. In the proposed work, detailed dataset preparation is performed along with precise annotation tools such as Roboflow, which makes it different from most earlier studies.

Although our study has found promising results, it is important to acknowledge certain limitations that might affect the generalization of our results. First, relying on a single institutional dataset that exclusively utilized images taken from a specific device restricts data diversity. Data diversity is an important factor in establishing the performance of AI models under a wide range of diagnostic conditions. Future research should incorporate datasets from multiple institutions and imaging systems representative of public and private healthcare settings. The aim would be to facilitate an all-rounded view of the robustness of the AI model in disparate populations and real clinical world settings where variations in equipment and imaging protocols will always be present.

Also in this study, images were taken from only one type of radiographic device, with potential bias on the issue of standardization. The work should be extended to study radiographs from different imaging devices with diverse technical specifications. This would be a way to comprehensively test various diagnostic imaging technologies for model performance and help increase its adaptability in various clinical settings. It will also show how well the AI system can account for variations in image quality, contrast, and resolution for better real-world application of the AI system.

In this respect, further studies should develop a model that includes an even wider range of clinical variables such as the demographic data of the patient, state of oral health, and caries risk assessment. That will allow a more precise calculation of the model's performance, showing in which groups of patients or under which

conditions the model can detect caries well. Including variables would provide an AI model with more sensitivity to personalized diagnostics and thus will guarantee its efficiency in a wider circle of real-life situations. Lastly, to enhance model accuracy and reliability, additional pre-processing and augmentation techniques should be explored as future work. Advanced noise reduction, enhancement in contrast, and 3D reconstructions from 2D images may include especially useful techniques that would help further fine-tune the model in question for better detection of subtle carious lesions. Such investigation with due diligence could lead to the introduction of novel strategies aimed at maximizing diagnostic potentials by using AI models on dental radiography.

Conclusion

This study represents a significant advance in the application of artificial intelligence to dental diagnostics, with a special view to detecting interproximal caries in bitewing radiographs. Using the YOLOv8 model, including extensive data augmentation techniques, ensured especially high sensitivity of enamel caries detection. The success here points out the superiority of YOLOv8 over previous models; thus, overcoming the variability in diagnosis from previous studies and increasing their accuracy in diagnosing early caries. However, the results on dentin caries detection showed that further refinement of the model is necessary for more complicated lesions. To establish generalizability such studies should be replicated using various imaging systems and clinical settings. For clinically expanded use within a dental practice, this model needs further development to be capable of much higher complexity regarding texture and other variations in clinical conditions occurring in nature.

Data availability

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request. Data are located in controlled access data storage at Naaptech CO.

Received: 18 August 2024; Accepted: 26 December 2024
Published online: 07 February 2025

References

1. Haghanifar, A., Majdabadi, M. & Ko, S. Paxnet: Dental caries detection in panoramic X-ray using ensemble transfer learning and capsule classifier. https://arxiv.org/abs/2012.13666 (2020).
2. Park, E. Y., Cho, H., Kang, S., Jeong, S. & Kim, E.-K. Caries detection with tooth surface segmentation on intraoral photographic images using deep learning. *BMC Oral Health* **22**, 1–9 (2022).
3. Chanintonsongkhla, C. & Chaovatut, V. Detecting Caries Lesions with Bitewing Radiograph Using Ensemble of Convolutional Neural Network Model.
4. Lee, S., Kim, D. & Jeong, H.-G. Detecting 17 fine-grained dental anomalies from panoramic dental radiography using artificial intelligence. *Sci. Rep.* **12**, 5172 (2022).
5. Lian, L., Zhu, T., Zhu, F. & Zhu, H. Deep learning for caries detection and classification. *Diagnostics* **11**, 1672 (2021).
6. Moran, M. et al. Classification of approximal caries in bitewing radiographs using convolutional neural networks. *Sensors* **21**, 5192 (2021).
7. Bader, J. D., Shugars, D. A. & Bonito, A. J. Systematic reviews of selected dental caries diagnostic and management methods. *Journal of dental education* **65**, 960–968 (2001).
8. Khanagar, S. B. et al. Application and performance of artificial intelligence technology in detection, diagnosis and prediction of dental caries (DC)—a systematic review. *Diagnostics* **12**, 1083 (2022).
9. Foster Page, L. *et al.* The effect of bitewing radiography on estimates of dental caries experience among children differs according to their disease experience. *BMC oral health* **18**, 1–8 (2018).
10. Min, J. K., Kwak, M. S. & Cha, J. M. Overview of deep learning in gastrointestinal endoscopy. *Gut and liver* **13**, 388 (2019).
11. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *nature* **542**, 115–118 (2017).
12. Bossuyt, P. M. et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Clinical chemistry* **61**, 1446–1452 (2015).
13. Talpur, S. *et al.* Uses of Different Machine Learning Algorithms for Diagnosis of Dental Caries. *Journal of Healthcare Engineering* **2022** (2022).
14. Albano, D. et al. Artificial intelligence for radiographic imaging detection of caries lesions: a systematic review. *BMC Oral Health* **24**, 274 (2024).
15. Gupta, A. C. et al. Fully automated deep learning based auto-contouring of liver segments and spleen on contrast-enhanced CT images. *Scientific Reports* **14**, 4678 (2024).
16. Sohan, M., Sai Ram, T., Reddy, R. & Venkata, C. 529–545 (Springer).
17. Sohan, M., Sai Ram, T., Reddy, R. & Venkata, C. in *International Conference on Data Intelligence and Cognitive Informatics.* 529–545 (Springer).
18. Jocher, G. *Instance Segmentation*, <https://docs.ultralytics.com/tasks/segment/> (2024).
19. Baydar, O., Różyło-Kalinowska, I., Futyma-Gąbka, K. & Sağlam, H. The u-net approaches to evaluation of dental bite-wing radiographs: An artificial intelligence study. *Diagnostics* **13**, 453 (2023).
20. García-Cañas, Á., Bonfanti-Gris, M., Paraíso-Medina, S., Martínez-Rus, F. & Pradíes, G. Diagnosis of interproximal caries lesions in bitewing radiographs using a deep convolutional neural network-based software. *Caries Res.* **56**, 503–511 (2022).
21. Lee, S. et al. Deep learning for early dental caries detection in bitewing radiographs. *Sci. Rep.* **11**, 16807 (2021).
22. Bayrakdar, I. S. *et al.* Deep-learning approach for caries detection and segmentation on dental bitewing radiographs. *Oral Radiology*, 1–12 (2021).
23. Cantu, A. G. et al. Detecting caries lesions of different radiographic extension on bitewings using deep learning. *J. Dentistry* **100**, 103425 (2020).
24. Mao, Y.-C. et al. Caries and restoration detection using bitewing film based on transfer learning with CNNs. *Sensors* **21**, 4613 (2021).
25. Vimalarani, G. & Ramachandraiah, U. Automatic diagnosis and detection of dental caries in bitewing radiographs using pervasive deep gradient based LeNet classifier model. *Microprocessors Microsystems* **94**, 104654 (2022).
26. Bayraktar, Y. & Ayan, E. Diagnosis of interproximal caries lesions with deep convolutional neural network in digital bitewing radiographs. *Clin. Oral Investigations* **26**, 623–632 (2022).
27. Chen, X., Guo, J., Ye, J., Zhang, M. & Liang, Y. Detection of proximal caries lesions on bitewing radiographs using deep learning method. *Caries Res.* **56**, 455–463 (2022).

28. Srivastava, M. M., Kumar, P., Pradhan, L. & Varadarajan, S. Detection of tooth caries in bitewing radiographs using deep learning. https://arxiv.org/abs/1711.07312(2017).
29. Kunt, L., Kybic, J., Nagyová, V. & Tichý, A. Automatic caries detection in bitewing radiographs: part I—deep learning. *Clin.Oral Investigations* **27**, 7463–7471 (2023).
30. Karakuş, R., Öziç, M. Ü. & Tassoker, M. AI-Assisted Detection of Interproximal, Occlusal, and Secondary Caries on Bite-Wing Radiographs: A Single-Shot Deep Learning Approach. *Journal of Imaging Informatics in Medicine*, 1–14 (2024).
31. Suttapak, W., Panyarak, W., Jira-Apiwattana, D. & Wantanajittikul, K. A unified convolution neural network for dental caries classification. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)* **16**, 186–195, (2022).
32. Pérez de Frutos, J. et al. AI-Dentify: deep learning for proximal caries detection on bitewing x-ray-HUNT4 Oral Health Study. *BMC Oral Health* **24**, 344 (2024).

Author contributions
Conceptualization: M.B, B.A, F.M; Methodology: M.B, B.A, F.M, H.A; Acquisition: M.B, F.M; Formal analysis and investigation: M.B, B.A, F.M; Writing- original draft preparation: M.B, F.M; Writing- review and editing: M.B, B.A, F.M, H.A; Funding acquisition: M.B, BA; Resources: B.A, F.M, H.A; Supervision: B.A, H.A

Funding
This research received no specific grants from public, commercial, or not-for-profit funding agencies.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to F.M.

Reprints and permissions information is available at www.nature.com/reprints.

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