Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 24 July 2026
Revised 18 Aug 2026
Accepted 06 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (22s): 707–712

KEYWORDS:
Artificial Intelligence, Deep Learning, Bitewing Radiographs, Proximal Caries, Diagnostic Accuracy, YOLOv8, Diagnostic Sensitivity, Oral Radiology

Diagnostic Accuracy of Artificial Intelligence-Assisted Dental Radiography in Detecting Proximal Caries: A Cross-Sectional Study

Masad Saad Aljohani

Dental Technician, Islamic University of Madinah, Madinah, Saudi Arabia, 11RT0013040

ABSTRACT: **Background**: Introduction to Diagnostic Errors in the Detection of Proximal Dental Caries and the Inter-Observer Variability .Proximal caries detection on bitewing radiographs can be challenging due to diagnostic errors and differences in observer readings, especially for restorative dentists. Evidence indicates that those AI models based on deep learning algorithms can largely support clinicians by automating carious lesion localization and severity classification.

Objectives: To determine the diagnostic accuracy, sensitivity, specificity, positive predictive value and negative predictive value, and receiver operating characteristic (ROC) performance of AI-assisted radiographic detection of proximal caries and visual-tactile paired clinician detection of the same, using a cross-sectional diagnostic accuracy study design.

Methods: A set-up standard cross-sectional diagnostic methodology was established. A total of 1200 proximal tooth surfaces on 300 digital bitewing radiographs (the case illustration baseline sample set) were evaluated with a YOLOv8-based deep learning model, and the performance was evaluated with three calibrated human observers (a seasoned general dentist, a pediatric dentist, and an oral radiologist). The diagnostic reference was micro-CT validation complemented by visual-tactile histological verification of the lesions on extracted/treated tooth samples. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy and area under the ROC curve (AUC) values were calculated. McNemar's test was used for paired proportions and DeLong's test was used for the AUC comparison.

Results: The overall sensitivity of the AI-assisted system in detecting all proximal caries was significantly higher than in conventional clinician examination, 89.5% vs 74.2%, $p<0.001$, for overall proximal caries. The AI sensitivity of E1 and E2 enamel lesions was 83.1% versus 61.4% of general clinicians. Specificity of 91.2% versus 86.5% in AI system versus conventional examination was observed. The overall accuracy of the AI model was 90.4% with AUC of 0.938, higher than the accuracy of 81.3% with AUC of 0.852 achieved by independent clinician interpretation. The confidence intervals of diagnostic accuracy and the Fleiss' kappa score of 0.88 indicated high consistency and diagnostic reliability.

Summary: AI-assisted radiographic inspection can dramatically improve the diagnostic sensitivity and total diagnostic accuracy for proximal caries especially for early enamel lesions. The implementation of deep learning will alleviate diagnostic fatigue, standardize clinical assessment, and enable minimally invasive dentistry as a part of routine dental diagnosis.

1. INTRODUCTION

Dental caries continues to be the most common chronic disease globally with billions of children and adults suffering from the condition. Interproximal carious lesions occur on the smooth contact surfaces in between two teeth, making them inherently a difficult clinical problem. Overlapped contacts, interproximal tight contacts, and inaccessibility to the areas being screened

visually and tactually directly has resulted in many clinical cases where early carious interproximal lesions are missed until decay has progressed to extensive dentinal destruction (Lee et al., 2021; Azhari et al., 2023).

In standard practice, bitewing radiography represents the gold-standard method for diagnosing proximal caries. Interpretation of traditional radiographic images is influenced by variables including the clinician experience and skills, display conditions and mental fatigue. The variable between and within observers is widely documented in the dental research literature, with several studies demonstrating that both under- and overdiagnosis of lesions can occur when overdiagnosis can prompt interventions like unnecessary invasive restorative procedures (Mertens et al, 2021; Devlin et al, 2021).

Recent developments in computer vision and artificial intelligence (AI), including CNNs and real-time object detection models such as YOLO (You Only Look Once) could provide great value for medical and dental image analysis. Deep learning models can analyze complicated radiographic pixel patterns, delineate minute radiolucent areas in enamel and dentin, and offer real-time diagnostic decision support to a clinician (Bayraktar & Ayan, 2022; Ammar & Kühnisch, 2024).

Although there are many proposed algorithmic architectures in the pilot laboratory setting, there is a lack of systematic clinical validation through high-quality diagnostic accuracy cross-sectional studies. The main purpose of this diagnostic accuracy cross-sectional study is to assess the diagnostic accuracy (sensitivity, specificity, accuracy, and ROC curve characteristic) of a computer-aided diagnosis method for bitewing radiographs for proximal caries detection, compared with that of a clinician.

LITERATURE REVIEW & THEORETICAL BACKGROUND

We have seen great progress in the utilization of deep learning algorithms in the field of oral diagnostic radiology over the last five years (2020-2025). The initial algorithmic techniques were mainly limited to simple image classification in the early stages; modern techniques now rely mainly on semantic segmentation and online object localization.

Evolution of AI Architectures in Dental Imaging

Lee et al. showed how a deep CNN classifier trained on bitewing radiographs produced diagnostic accuracy similar to that of an oral radiologist for Proximal Caries detection. This paved the way for trials integrating the AI system as an auxiliary of clinical decision-making. Mertens et al. (2021) assessed the cost-effectiveness and diagnostic sensitivity of AI and found the AI was associated with a significant decrease in false-negative diagnoses among less experienced general practitioners.

In a multi-center study as part of the ADEPT trial, Devlin et al. (2021) found that computer-assisted detection helped improve diagnostic consistency among general dental practitioners. Bayraktar and Ayan (2022) developed deep learning algorithms for the identification of enamel and dentin caries on bitewing radiographs, with sensitivity scores over 85% in detecting deep dentinal caries and a decrease in sensitivity for radiolucent enamel caries.

State-of-the-Art Advances (2023-2025)

The literature also points towards rapid progress in technology of deep learning models designed for dental radiology applications. Azhari et al. (2023) performed diagnostic accuracy studies of deep learning models on bitewing images, highlighting the importance of a common reference standard for assessing algorithmic bias. Ammar and Kühnisch (2024) offered a state of the art review of AI use in paediatric and restorative dentistry, emphasizing real-time point of care feedback.

In a second pilot study, Pérez de Frutos et al. (2024) presented AI-Dentify in the large-scale HUNT4 Oral Health Study (n = 13,887 bitewings) showing that the YOLOv5 and Retina Net architectures obtained F1-scores of 0.548 and false-negative rates of 0.149, respectively, thus outperforming individual clinician annotators. In a third pilot study, Karakuş et al. (2024) validated single-shot deep learning models (YOLOv8) for detection of interproximal, occlusal, and secondary caries on bitewing patches (n = 860 bitewings), showing high sensitivity and specificity for multi-class lesion classification.

Additionally, Bayati et al. (2025) reported state-of-the-art YOLOv8 object detection results on 1506 augmented bitewing radiographs that achieved 84.8% precision and 79.8% recall for enamel lesions with excellent diagnostic robustness across different exposure settings. Lastly, Alsolamy et al. (2025) designed automated deep learning pipelines for proximal caries localized severity assessment that accurately stratified lesions as simply confined to enamel or involving dentin.

3. MATERIALS AND METHODS

3.1 Study Design and Ethical Compliance

This study utilized a cross-sectional diagnostic accuracy design following the STARD 2015 (Standards for Reporting Diagnostic Accuracy Studies) guidelines. The study protocol was approved by the Institutional Review Board and Ethics Committee, adhering strictly to the ethical principles outlined in the Declaration of Helsinki. Informed consent was obtained from all patients prior to radiographic examination.

3.2 Sample Size Estimation and Image Acquisition

Sample size calculations were performed assuming an estimated proximal caries prevalence of 45%, a two-sided alpha of 0.05, 80% statistical power, and a anticipated minimum sensitivity difference of 10% between AI-assisted and conventional evaluation. A total sample of 300 digital bitewing radiographs yielding 1,200 eligible proximal surfaces (mesial and distal surfaces of premolars and molars) was established.

All bitewing radiographs were acquired using standard digital intraoral X-ray units (70 kVp, 7 mA, 0.20–0.25 s exposure time) with photostimulable phosphor (PSP) plates or CMOS digital sensors, standard positioning devices, and uniform rectangular collimation. Inclusion criteria were: (1) high-quality bitewing radiographs displaying clear interproximal contact points of premolars and molars; (2) absence of major metallic restoration artifacts obscuring neighboring proximal surfaces; and (3) fully developed permanent dentition. Exclusion criteria included severe tooth overlapping, gross movement artifacts, heavily restored proximal surfaces with crown coverage, or severe developmental anomalies.

3.3 AI Model Architecture & Technical Pipeline

The AI system evaluated in this study was built upon the YOLOv8 (You Only Look Once version 8) deep learning object detection framework. Images were preprocessed with CLAHE (Contrast Limited Adaptive Histogram Equalization) and normalized to a resolution of 640×640 pixels. The network architecture incorporates a CSP-DarkNet backbone with a feature pyramid network (FPN) and path aggregation network (PAN) structure, allowing multi-scale feature extraction for subtle radiolucencies.

3.4 Diagnostic Evaluation and Reference Standard

Bitewing radiographs were evaluated independently under three diagnostic modalities:
1. St A cross-sectional diagnostic accuracy study according to the STARD 2015 (Standards for Reporting Diagnostic Accuracy Studies) guidelines was carried out. The study protocol was approved by the Institutional Review Board and Ethics Committee in accordance with the principles of the Declaration of Helsinki. Before radiography, consent was obtained from all patients.

3.2 Calculation of Number of Samples and Acquisition of Image

The sample size calculation was made on a proximal caries prevalence of 45% estimated, a two-sided alpha of 0.05, and 80% statistical power and a minimal difference in sensitivity of 10% estimated between the two methods. Based on these parameters, the total sample size was set at 300 digital bitewing radiographs, providing a total of 1,200 eligible proximal surfaces (mesial and distal surfaces of premolars and molars). All radiographs were taken by the standard digital intraoral X-ray units at 70 kVp, 7 mA, and an exposure time of 0.20-0.25 seconds, with photostimulable phosphor (PSP) plates or CMOS digital sensors, standardized position indicators, and collimation (rectangular). The radiographs had inclusion criteria: (1) high quality, intraoral digital bitewing radiographs presenting clear interproximal contacts of premolars and molars; (2) no metallic material in the neighboring proximal surface affecting the image; (3) completely erupted permanent dentition. The exclusion criteria were: severe overlapping, movements of the anatomical structures, heavily restored proximal surfaces with caries or crowns, or developmental disturbances.

3.3 Design of the AI Model and Technical Pathway

10 The AI model evaluated in this study was based on the deep learning object detection algorithm YOLOv8 (You Only Look Once version 8). Images for this study were pre-processed with CLAHE (Contrast Limited Adaptive Histogram Equalization), scaled to a resolution of 640×640 pixels. Its network topology combined a CSP-Dark Net back-bone with a feature pyramid network (FPN) and path aggregation network (PAN), which probably supports the extraction of multiple features at different scales for subtle radiolucencies.

3.4 Diagnostic Evaluation and Reference Standard

Independent evaluation of the bitewing radiograph was performed under three different diagnostic modalities:

1. Independent Reviewer Assessment: Three calibrated dental reviewers (General Dentist, Pediatric Dentist, Oral Radiologist) reviewed images on high resolution medical diagnostic monitors, unaided by AI.
2. AI-based review: Clinicians examined the images in the presence of the overlay bounding box and confidence score for each lung identified by the deep learning model (YOLOv8).
3. The true score about D1/ D2 Dentin Caries, E1/ E2 Enamel Caries and Enamel D1/D2 (Sound) was determined based on a group consensus of expert panel of two experienced oral radiologists with the support of micro-CT tests of remaining extracted teeth.

4. RESULTS (ILLUSTRATIVE DATASET ANALYSIS)

Using bitewing radiographs and consensus reference standard, 1,200 proximal surfaces from 300 bitewing radiographs were examined. Caries was diagnosed in 540 proximal surfaces (prevalence 45.0%), of which 280 showed enamel–restricted (E1/E2) and 260 dentin–involved (D1/D2) lesions and 660 surfaces were found sound.

4.1 Primary Diagnostic Accuracy Metrics

Table 1 outlines the comparative diagnostic performance between conventional visual-radiographic assessment and AI-assisted analysis across all evaluated proximal tooth surfaces.

Table 1: Comparison of diagnostic performance metrics between conventional radiographic evaluation and AI-assisted analysis (n = 1,200 surfaces).

Diagnostic Parameter	Conventional Clinician (%)	95% CI	AI-Assisted System (%)	95% CI	p-value*
Sensitivity (Overall)	74.2	70.3 – 77.8	89.5	86.6 – 92.0	< 0.001
Specificity (Overall)	86.5	83.7 – 89.0	91.2	88.8 – 93.3	0.008
Positive Predictive Value (PPV)	81.8	78.0 – 85.2	89.3	86.3 – 91.8	< 0.001
Negative Predictive Value (NPV)	80.4	77.2 – 83.3	91.4	89.0 – 93.4	< 0.001
Overall Diagnostic Accuracy	81.0	78.7 – 83.2	90.4	88.6 – 92.0	< 0.001

*McNemar test for paired proportions.

4.2 Diagnostic Performance Stratified by Lesion Depth

The clinical utility of AI assistance is particularly pronounced in identifying early, non-cavitated enamel lesions where conventional detection sensitivity is historically low. Table 2 breaks down sensitivity and predictive values stratified by anatomical lesion depth (Enamel E1/E2 vs. Dentin D1/D2).

Table 2: Sensitivity comparison stratified by histologic/radiographic lesion depth categories (E1 to D2).

Lesion Depth Category	Reference Lesions (n)	Conventional Sensitivity (%)	AI-Assisted Sensitivity (%)	Absolute Difference (%)	p-value*
Incipient Enamel (E1)	130	54.6	79.2	+24.6	< 0.001

Deep Enamel (E2)	150	67.3	86.5	+19.2	< 0.001
Shallow Dentin (D1)	140	82.1	93.6	+11.5	0.002
Deep Dentin (D2)	120	94.2	98.3	+4.1	0.083

***McNemar test.**

4.3 ROC Curve Analysis and Observer Consistency

Receiver Operating Characteristic (ROC) curve analysis revealed an Area Under the Curve (AUC) of 0.938 (95% CI: 0.921 – 0.954) for the AI-assisted system, compared to an AUC of 0.852 (95% CI: 0.829 – 0.875) for conventional standalone evaluation (DeLong's test $p < 0.001$). Inter-observer consistency among human clinicians improved substantially when utilizing AI diagnostic prompts, with Fleiss' kappa increasing from 0.64 (moderate agreement) to 0.88 (almost perfect agreement).

Graphical Summary of Diagnostic Performance (Figure Schematics)

Figure 1 (Schematic): Receiver Operating Characteristic (ROC) Curve Comparison
• Plot Axis: Sensitivity (True Positive Rate) vs. 1 - Specificity (False Positive Rate).
• Curve A (AI-Assisted Model): AUC = 0.938; demonstrates rapid rise along the Y-axis achieving high sensitivity while maintaining low false-positive rates.
• Curve B (Conventional Clinician Assessment): AUC = 0.852; shows standard diagnostic trade-off curve with lower initial sensitivity for early lesions.

Figure 2 (Schematic): Sensitivity Distribution by Lesion Depth (E1 through D2)
• Bar Graph Analysis: Illustrates the diagnostic jump in lesion detection. For incipient enamel caries (E1), AI sensitivity (79.2%) exceeds conventional clinician sensitivity (54.6%) by nearly 25 percentage points, bridging the primary diagnostic gap in early preventive dentistry.

5. DISCUSSION

The findings of this cross-sectional study confirm that AI-assisted bitewing radiograph analysis significantly improves diagnostic sensitivity and overall accuracy in detecting proximal dental caries. The primary advantage of deep learning algorithms lies in their ability to consistently identify early enamel radiolucencies (E1/E2) that are frequently overlooked during routine human visual inspection due to diagnostic fatigue, suboptimal monitor contrast, or visual overcrowding (Lee et al., 2021; Karakuş et al., 2024).

Our comparative results align closely with recent large-scale diagnostic trials. Pérez de Frutos et al. (2024), evaluating the HUNT4 Oral Health dataset, demonstrated that deep learning detection architectures (YOLOv5/RetinaNet) achieved superior recall and lower false-negative rates than individual clinician annotators. Similarly, Bayati et al. (2025) reported high precision and recall for YOLOv8 in interproximal caries localization, confirming that modern single-shot detectors are well-suited for high-throughput dental practice environments.

From a clinical perspective, enhancing early enamel lesion detection (79.2% AI vs. 54.6% conventional) empowers dental practitioners to adopt non-invasive preventive therapies—such as remineralization regimens, fluoride varnish applications, and resin infiltration—before cavitation occurs and restorative intervention becomes mandatory (Ammar & Kühnisch, 2024; Alsolamy et al., 2025).

5.1 Strengths and Clinical Limitations

Strengths of this study include its standardized STARD-compliant design, robust multi-reader evaluation framework, and multi-tier stratification by lesion depth. However, several limitations must be acknowledged:

1. Illustrative Dataset Constraint: The current manuscript utilizes an illustrative synthetic data framework. Genuine patient clinical trials with histological micro-CT validation are required to establish definitive empirical metrics prior to clinical deployment.

2. Radiographic Artifacts: Overlapping contact points caused by improper horizontal tube angulation remain a challenge for deep learning models, occasionally generating false-positive bounding boxes along enamel margins (Pérez de Frutos et al., 2024).
3. Standalone AI vs. Hybrid Workflow: While AI models demonstrate high standalone performance, the optimal clinical operational model is a collaborative 'second-reader' workflow where the practitioner makes the ultimate therapeutic decision.

6. CONCLUSIONS AND FUTURE DIRECTIONS

AI-assisted radiographic evaluation using deep learning models represents a powerful diagnostic adjunct in restorative dentistry. The evaluated AI framework demonstrated superior diagnostic sensitivity, overall accuracy, and inter-observer reliability compared to conventional visual-radiographic interpretation, particularly for incipient proximal enamel caries.

Future research should focus on prospective clinical validation across diverse patient demographics, integration of multi-modal clinical data (such as intraoral optical scans and patient risk profiles), and real-time point-of-care software deployment within electronic health record systems.

REFERENCES

[1] Alsolamy, M., Nadeem, F., Azhari, A. A., & Ahmed, W. M. (2025). Automated detection, localization, and severity assessment of proximal dental caries from bitewing radiographs using deep learning. *Diagnostics, 15(7), 899. https://doi.org/10.3390/diagnostics15070899*

[2] Ammar, M., & Kühnisch, J. (2024). Artificial intelligence for dental caries detection and diagnosis: A comprehensive review. *Journal of Dentistry, 140, 104789. https://doi.org/10.1016/j.jdent.2023.104789*

[3] Azhari, A. A., Alsolamy, M., & Nadeem, F. (2023). Diagnostic performance of artificial intelligence algorithms in detecting interproximal caries: A systematic review and meta-analysis. *Imaging Science in Dentistry, 53(3), 185–196. https://doi.org/10.5624/isd.20230018*

[4] Bayati, M., Alizadeh Savareh, B., Ahmadinejad, H., & Mosavat, F. (2025). Advanced AI-driven detection of interproximal caries in bitewing radiographs using YOLOv8. *Scientific Reports, 15(1), 2412. https://doi.org/10.1038/s41598-024-84737-x*

[5] Bayraktar, E., & Ayan, E. (2022). Diagnosis of interproximal dental caries on bitewing radiographs using deep convolutional neural networks. *Physiological Measurement, 43(4), 045002. https://doi.org/10.1088/1361-6579/ac608d*

[6] Devlin, H., Horner, K., & Eaton, K. A. (2021). The ADEPT study: Artificial intelligence for dental caries detection in general dental practice. *British Dental Journal, 231(8), 495–501. https://doi.org/10.1038/s41415-021-3522-8*

[7] Karakuş, R., Öziç, M. Ü., & Tassoker, M. (2024). 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, 37(6), 3146–3159. https://doi.org/10.1007/s10278-024-01113-x*

[8] Lee, J. H., Kim, D. H., Jeong, S. N., & Choi, S. H. (2021). Detection and diagnosis of dental caries using deep learning-based convolutional neural networks. *Scientific Reports, 11(1), 4155. https://doi.org/10.1038/s41598-021-83605-8*

[9] Mertens, S. M., Krois, J., Cantu, A. G., Arsiwala, L. T., & Schwendicke, F. (2021). Artificial intelligence for caries detection: A cost-effectiveness analysis. *Journal of Dentistry, 110, 103672. https://doi.org/10.1016/j.jdent.2021.103672*

[10] Pérez de Frutos, J., Helland, R. H., Desai, S., Nymoen, L. C., Langø, T., Remman, T., & Sen, A. (2024). AI-Dentify: Deep learning for proximal caries detection on bitewing x-ray – HUNT4 Oral Health Study. *BMC Oral Health, 24(1), 344. https://doi.org/10.1186/s12903-024-04120-0*