

Clinical Oral Investigations (2026) 30:207  
https://doi.org/10.1007/s00784-026-06882-z

RESEARCH



# Evaluating AI diagnostic accuracy in approximal dental caries detection on bitewing radiographs

Kostis Giannakopoulos<sup>1</sup> · Argyro Kavadella<sup>1</sup> · Dimitrios Paraskevis<sup>1,2</sup> · Aristidis Arhakis<sup>3</sup> · Miltiadis A. Makrygiannakis<sup>1,4</sup> · Eleftherios G. Kaklamanos<sup>1,3,5</sup>

Received: 4 September 2025 / Accepted: 14 April 2026  
© The Author(s) 2026

## Abstract

**Objectives** To evaluate the diagnostic accuracy of the Diagnocat™ artificial intelligence (AI) system for caries detection on bitewing radiographs compared with expert human examiners, with emphasis on differences between enamel and dentin lesions.

**Materials and methods** A sample of 100 digital bitewing radiographs (1540 surfaces) was retrospectively selected from the European University Cyprus dental clinic database using a systematic backward screening method. Radiographs were obtained with a standardized phosphor plate system and anonymized before analysis. Two independent experts (operative dentistry and oral radiology) established the reference standard. AI and human assessments were binarized (caries/no caries; enamel/dentin) and compared. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were calculated, and statistical significance was tested across detection categories.

**Results** Diagnocat™ showed high specificity (94.3%, 95% CI: 92.4%–96.0%) and NPV (96.1%, 95% CI: 94.7%–97.3%), with an overall accuracy of 91.6% (95% CI: 89.7%–93.4%). Sensitivity was moderate (73.1%, 95% CI: 65.9%–79.9%), and PPV was 64.7% (95% CI: 57.7%–71.5%). Agreement with the expert consensus was substantial (Cohen’s  $\kappa=0.638$ ). For enamel lesions, sensitivity and specificity were 73.3% (95% CI: 62.8%–82.7%) and 92.9% (95% CI: 91.0%–94.7%) with moderate agreement with the consensus (Cohen’s  $\kappa=0.492$ ) and for dentin lesions they were 72.8% (95% CI: 61.8%–83.8%) and 92.8% (95% CI 90.9%–94.6%) with moderate agreement with the consensus (Cohen’s  $\kappa=0.468$ ). NPV remained high ( $\geq 98.0\%$ ), while PPV was low (42.0% and 39.2%), across lesion types. Detection patterns differed significantly between AI