UNIVERSITY OF CALIFORNIA

Los Angeles

Harnessing Artificial Intelligence for Caries Detection:

A New Paradigm in Dental Education

A thesis submitted in partial satisfaction

of the requirements for the degree Master of Science

in Oral Biology

by

Somyung Ji

2025

ABSTRACT OF THE THESIS

Harnessing Artificial Intelligence for Caries Detection:

A New Paradigm in Dental Education

by

Somyung Ji

Master of Science in Oral Biology

University of California, Los Angeles, 2025

Professor Sanjay M. Mallya, Chair

Dental caries remain highly prevalent worldwide, underscoring the need for more accurate and efficient diagnostic methods in dental education. Traditional radiographic interpretation, though essential, often suffers from variability and limited sensitivity, prompting exploration of artificial intelligence (AI) as a supportive tool. This study evaluated whether an AI platform (Second Opinion®) could enhance radiographic caries detection in a dental school setting by comparing its diagnostic performance to that of second-year dental students and by assessing its impact on faculty accuracy and consensus. AI performance was compared with caries detection exam results from second-year dental students in the 2023 cohort (Cohort 1). The same exam was later repurposed as a self-assessment quiz for the 2024 cohort (Cohort 2), and their performance was compared with that of the AI. Subsequently, a new AI-assisted caries detection exam was developed for Cohort 2, incorporating ≥75% faculty agreement as the gold standard for lesion classification. Diagnostic metrics (sensitivity, specificity, accuracy, precision, and F1 score) were calculated for students, AI, and faculty members—both without and with AI annotations. The AI platform outperformed both student cohorts, achieving higher sensitivity (89.5%) and accuracy (93.6%). Cohort 2 demonstrated significant improvement after structured self-assessment, with accuracy increasing from 40.4% in the self-assessment to 61.7% in the caries detection exam. Notably, Cohort 2 surpassed Cohort 1's first attempt pass rate (96.25% vs. 55.8%). Among faculty, three of four members showed increased sensitivity and accuracy with AI annotations, and unanimous (4/4) consensus improved from 73.33% to 86.67%. The AI platform consistently exhibited higher diagnostic performance than second-year dental students, reinforcing its potential as a reliable adjunct in caries detection training. Moreover, AI-assisted workflows streamlined exam development and improved faculty consensus. While AI provides diagnostic support, it should complement—rather than replace—clinician-led education and judgment. With careful curriculum integration, AI holds substantial potential for elevating diagnostic standards and refining dental training.

The thesis of Somyung Ji is approved.

Kumar C. Shah

Vinodh Bhoopathi

Sanjay M. Mallya, Committee Chair

University of California, Los Angeles

2025

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
AI	Artificial Intelligence
CNN	Convolutional Neural Network
D1	Dentin caries extending less than one-third of the dentin depth
D2	Dentin caries extending less than two-thirds of the dentin depth
D3	Dentin caries extending more than two-thirds of the dentin depth, with potential pulp involvement
DEJ	Dentinoenamel Junction
E0	Sound Enamel
E1	Enamel caries extending less than half of the enamel depth
E2	Enamel caries extending to the dentinoenamel junction
FMX	Full Mouth Series of Radiographs

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I extend my deepest gratitude to Dr. Sanjay M. Mallya, my principal investigator and Section Chair of Oral and Maxillofacial Radiology. His unwavering mentorship, guidance, and expertise have been invaluable throughout this project. Beyond this research, he has been more than an advisor—he has been a mentor and an inspiration, shaping my journey as an oral and maxillofacial radiologist in ways I never anticipated. His patience, generosity, and willingness to share his profound knowledge have not only guided me through this project but have also profoundly influenced my growth as a clinician and researcher.

I also extend my heartfelt appreciation to my committee members, Dr. Kumar C. Shah and Dr. Vinodh Bhoopathi, for their invaluable insights, encouragement, and dedication to my thesis. Learning from such distinguished faculty has been a privilege, and their perspectives have significantly enriched my work.

I am deeply thankful to Drs. Michelle Rappeport, Tin H. Nguyen, Elizabeth Pollak, and Emily Wenzel, whose expertise and continuous support have been essential to the success of this project—from its early stages to its completion.

My sincere gratitude also goes to the entire UCLA Oral & Maxillofacial Radiology Clinic, especially Dr. Sotirios Tetradis, Dr. Mohammed A. Husain, my co-residents, and the clinic staff. Their collaboration, encouragement, and support have made it possible to balance clinical training alongside research and academic commitments.

Lastly, I am profoundly grateful for the unwavering support of my family, especially my parents, Dr. Jooyeon Ji and Miseong Kim. Their belief in me has been the foundation of my success, and I could not have made it here without them.

INTRODUCTION

Dental caries—commonly referred to as tooth decay—continues to be one of the most prevalent oral health issues worldwide[1-5], affecting both pediatric and adult populations[2, 4-6]. In the United States, caries are present in more than 25% of children aged two to five, approximately 50% of those aged twelve to fifteen, and over 90% of adults beyond the age of forty[6]. Such a high prevalence highlights the significant burden that caries places on healthcare systems, the broader economic landscape, and patients' overall well-being[4]. Early detection and timely intervention are widely recognized as the keys to minimizing tooth morbidity, reducing the need for invasive treatments, and ultimately lowering overall treatment costs[1, 3, 4]. Conversely, consequences of inaccurate or delayed diagnoses range from continued pain and infection to tooth loss and compromised general health[3, 7-9].

1. Challenges in Radiographic Caries Detection

Bitewing radiography is the established standard to detect proximal caries, especially those that remain elusive during clinical examinations[1, 6, 10]. Radiographic imaging for proximal caries detection is limited by its low sensitivity (24% to 42%), especially for non-cavitated lesions[11, 12]. Beyond technical limitations, even experienced clinicians may overlook 20–40% of carious lesions when relying solely on bitewing radiographs[6]. Factors contributing to these diagnostic gaps include varied radiographic quality, overlapping anatomical structures[13], optical illusions (e.g., Mach band), artefacts such as cervical burnout[10, 12, 14], the clinician's training and experience[12, 15-17], and potential observer bias[18]. Moreover, inter-examiner and intra-examiner disagreements in diagnosing the presence and extent of caries are welldocumented, reflecting the nuanced and often subjective nature of radiographic interpretation[1, 6, 10, 19-21].

From an educational standpoint, such variability translates into conflicting guidance for dental students, who often rely on multiple faculty members for training. When faculty members interpret the same radiograph differently[10, 18], students may receive inconsistent feedback and learn divergent diagnostic criteria. This not only impedes the acquisition of reliable skills but can also propagate inaccuracies that extend into future clinical practice[10].

To address these inconsistencies, dental schools may organize faculty calibration sessions[18, 22], where instructors collectively analyze radiographs to refine diagnostic standards. While such activities can improve consistency, they demand substantial assets—human resources and time[3, 13, 18]. Calibration requires regular alignment sessions and consensus-building, which are made more challenging by diverse faculty backgrounds, scheduling conflicts, and varying clinical philosophies[3, 17, 18]. Moreover, as student populations often outpace the number of available faculty members, ensuring timely and uniform feedback can become a major hurdle[18]. In response, many dental schools have begun to explore technology-driven solutions to support and standardize caries detection training[12, 23].

2. The Emergence of Artificial Intelligence in Dentistry

Rapid advances in computer science have propelled artificial intelligence (AI) to the forefront of healthcare innovation[15, 24, 25], and dentistry is no exception[12, 25, 26] . AI encompasses computational methods[12]—often leveraging machine learning and deep learning—that enable systems to recognize patterns, predict outcomes, and even simulate aspects of human decision-making[27-29]. In the context of dental radiography, AI-driven algorithms have demonstrated considerable promise in identifying carious lesions with accuracy levels that can rival, and sometimes surpass, those of experienced clinicians[8, 30, 31].

Recent studies on AI-assisted detection of enamel-only proximal caries have reported improved sensitivity compared to traditional methods[30]. Additionally, several studies suggest that AI can enhance diagnostic consistency[28], potentially lowering false-negative and false-positive rates[3, 32]. Many of these studies illustrate how machine learning algorithms interpret bitewing radiographs to detect and annotate potential carious lesions[1, 6, 10, 12, 30, 31, 33]. By training convolutional neural networks (CNNs) on large datasets of labeled images, these systems effectively learn the radiographic signatures of carious lesions and can identify suspicious areas in new, unseen images almost instantly[1, 3, 6, 34]. This capability has spurred the development of several commercial AI platforms—such as Second Opinion®, Overjet, Denti.AI, and Diagnocat—and points toward AI's transformative potential in clinical care and dental education[28, 32, 35].

3. Integrating AI into Dental Education

One of the most compelling reasons to adopt AI in dental curricula is its potential to provide standardized, timely feedback[3, 18, 36]. In traditional settings, students typically wait for faculty members to review radiographs and offer guidance[15]—sometimes days or weeks after the initial exposure. By contrast, AI tools can supply immediate and consistent input, as algorithms are programmed to apply the same analytical rules across all images without fatigue or shifting diagnostic criteria[3, 10, 13, 18, 29, 37]. For students, this consistency can be invaluable. They can compare their initial assessments with AI suggestions in real-time[3], reinforcing proper technique and recognizing specific diagnostic pitfalls[18]. This human-machine synergy may expedite skill acquisition while encouraging trainees to question any discrepancies between their evaluations and the AI's findings[15]. Consequently, AI-supported learning may sharpen critical thinking[29, 36, 38] and foster a more thorough understanding of radiographic landmarks associated with dental caries[3, 6, 39].

Beyond student learning, AI systems could play a valuable role in faculty calibration[18]. By examining AI's consistent outputs, instructors can more readily identify and address divergent diagnostic practices, ultimately leading to more uniform teaching standards across different clinical settings[13, 29]. This approach allows faculty members to recalibrate[18] and converge on shared diagnostic criteria, thereby improving consistency in educational delivery[3, 18, 28].

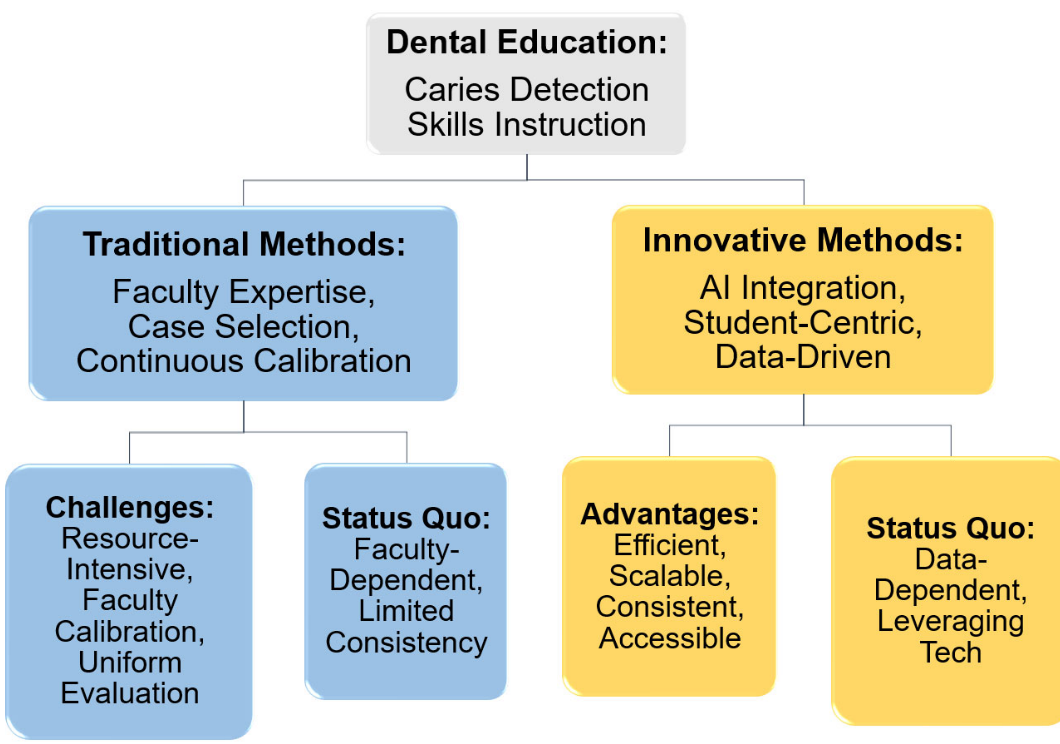


Figure 1. Overview of Traditional vs. Innovative Approaches to Caries Detection Instruction in Dental Education. This schematic compares traditional methods—characterized by reliance on faculty expertise, manual case selection, and continuous calibration—with more innovative, AI-driven methods that emphasize student-centric and data-driven processes.

4. Ongoing Barriers and Considerations

Despite AI's promise, multiple barriers stand in the way of seamless adoption in dental education. Ethical and legal questions—particularly around data privacy and liability for misdiagnosis—must be addressed[25, 28, 32]. Students may become overly dependent on AI-generated diagnoses[29], thereby neglecting the critical thinking and pattern-recognition skills that are essential for independent clinical practice[29, 36, 38]. Additionally, algorithms often require ongoing retraining and validation using region-specific or institution-specific data to maintain accuracy across diverse patient populations and radiographic techniques[8, 10, 16, 26, 39-41]. Financial constraints further complicate implementation, as commercial AI subscriptions and the requisite technological infrastructure can be costly[32, 36, 37].

Consequently, careful curriculum design is crucial—one that treats AI as a supportive tool rather than a replacement for thorough, clinician-led instruction[3, 29]. An incremental approach—introducing AI in controlled scenarios, assessing its impact, and gradually refining its integration—may be the most pragmatic path forward[27, 37, 38, 42]. By coupling AI-driven feedback with robust faculty guidance, dental schools can harness technology's benefits without sacrificing the essential human expertise at the heart of clinical education[13, 15].

OBJECTIVE AND SPECIFIC AIMS

OBJECTIVE

To evaluate the potential of artificial intelligence in enhancing caries detection within dental education by assessing its diagnostic performance, examining its impact on faculty diagnostic performance and consensus, and determining its feasibility as a supportive tool.

Specific Aims

1. **Assess AI Performance**
    - Analyze the diagnostic accuracy of the AI platform (e.g., Second Opinion®) in detecting dental caries on radiographs.
    - Compare the AI's diagnostic performance to that of second-year dental students (D2) from two consecutive cohorts to evaluate relative caries detection capabilities.
2. **Evaluate Faculty Integration**
    - Investigate how the AI platform influences faculty members' diagnostic performance by comparing outcomes without and with Second Opinion®.
    - Determine whether the AI platform affects faculty members' agreement on caries diagnosis, by comparing outcomes without and with Second Opinion®.
3. **Examine Feasibility in Education**
    - Assess the practicality and potential benefits of incorporating AI technology into routine dental education, particularly regarding improvements in student learning.

DESIGN AND METHODOLOGY

Study Design

The study was conducted at the UCLA School of Dentistry (UCLA SOD) using student performance data from a caries detection examination in the second-year predoctoral oral radiology course, Interdisciplinary Dental Sciences 201 (IDS 201). Data from two student cohorts from the academic years 2023 (Cohort 1, n = 86) and 2024 (Cohort 2, n = 80) were included. The study was reviewed and approved by the Institutional Review Board (IRB No. 21-000719) and conducted in compliance with all relevant guidelines and regulations.

1. Instruction in Radiographic Caries Detection

The IDS 201 oral radiology course is offered during the spring quarter of the second-year dental curriculum. This course prepares students to take intraoral radiographs and perform basic radiographic interpretation for detecting caries, periodontal bone loss, and apical periodontal inflammation. Each radiographic interpretive skill is taught and assessed independently. This study utilized only the data from the caries detection examinations.

Before enrolling in this course, students receive didactic instruction on general radiographic principles, intraoral radiographic techniques, and the pathogenesis of dental caries. In addition to classroom instruction, they watch instructional videos on systematic radiographic interpretation, with a focus on dental caries detection.

2. Artificial Intelligence in Radiographic Caries Detection

This study used an AI-powered radiographic interpretation tool, Second Opinion® (Pearl Inc., Los Angeles, CA, USA). The AI algorithm identifies and annotates potential carious lesions, periodontal bone loss, and periapical lesions on intraoral and panoramic radiographs using a sophisticated deep-learning model trained on a vast dataset of dental radiographs[28, 32]. As part of its annotation process, carious lesions are highlighted in pink, while periodontal bone levels are measured in millimeters (Figure 2).

The accuracy of AI detection was accepted or rejected through human intervention, in the form of faculty consensus and calibration. AI annotations of sound surfaces as carious lesions or overestimation of lesion depth were considered false positives, whereas false negatives represented missed caries detections and underestimated lesion depth (Figure 2).

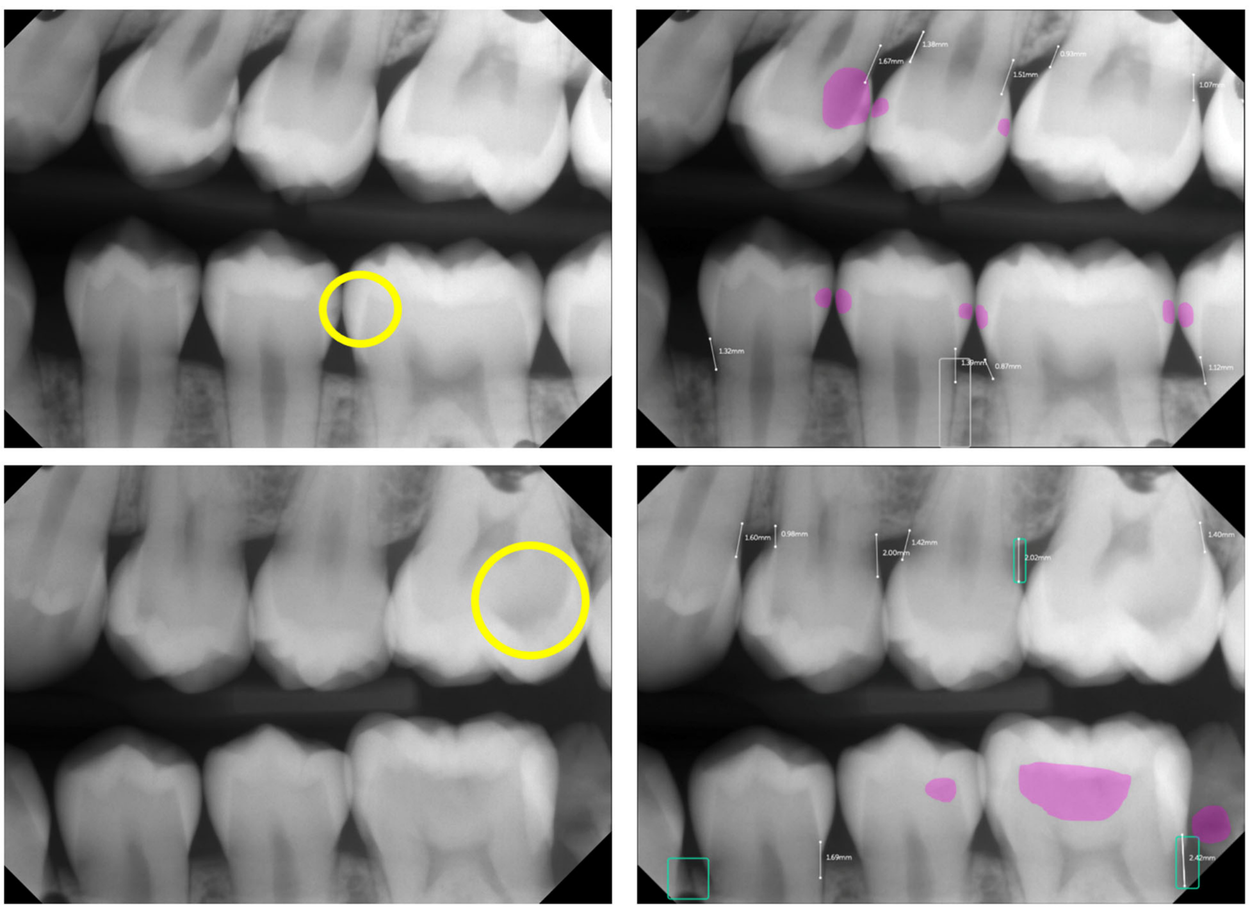


Figure 2. AI-Processed Radiograph Analysis and Examples of Misdiagnosis. The left panel shows an original radiograph, while the right panel displays the corresponding AI-processed analysis with carious lesions highlighted in pink and periodontal bone levels

measured in millimeters. The top row shows a false positive case where the AI incorrectly identified a sound surface as caries. The bottom row demonstrates a false negative case where the AI failed to detect a carious lesion.

3. Level 1 Caries Detection Examination

3.1 Examination Structure

While enrolled in the IDS 201 oral radiology course, students have limited clinical experience, as they are still in the early stages of their training. The course includes an examination designed to assess entry-level (Level 1) caries detection skills. Student performance is evaluated through an online examination administered via Canvas (Instructure, Salt Lake City, UT, USA), a learning management system.

The exam consists of 10 radiograph-based questions, requiring students to determine the presence or absence of caries based on the American Dental Association (ADA) Caries Classification System. To ensure content consistency, each student receives the same distribution of carious lesions—3 E0, 3 E1/E2, 3 D1/D2, and 1 D3—randomly selected from a question bank. The exam is time-limited to 10 minutes, with a passing criterion of ≥7 correct answers out of 10. The examination portal restricts reattempts to once every 24 hours, during which students are expected to review and prepare using self-assessment exercises.

3.2 Examination Administration

3.2.1 Cohort 1

The caries detection exam was administered to 86 D2 students, with unlimited attempts allowed to achieve a passing score. Students were given the opportunity to self-assess their knowledge through an ungraded, five-question quiz with no time limit. The self-assessment quiz was optional; however, students who did not pass the level 1 caries detection exam were required to complete the self-assessment quiz before reattempting the examination.

3.2.2 Cohort 2

The caries detection exam was administered to 80 D2 students. Students were allowed unlimited attempts, and passing the self-assessment quiz was a prerequisite for taking the level 1 caries detection exam. The level 1 examination from Cohort 1 was repurposed as a self-assessment quiz for Cohort 2, with a passing threshold of ≥70%.

3.3 Question Bank Development and Faculty Calibration

3.3.1 Cohort 1: Faculty-Calibrated Question Bank (No AI Assistance)

For the level 1 caries detection exam in Cohort 1, the course director selected radiographs of acceptable quality, which were then evaluated by IDS 201 course faculty (n = 5), each with over 20 years of clinical experience. Faculty members scored carious lesions according to the ADA Caries Classification System. The final diagnosis was established through faculty consensus, setting the gold standard for the examination.

3.3.2 Cohort 2: AI-Assisted Question Bank Development and Faculty Calibration

For the Level 1 caries detection exam in Cohort 2, Second Opinion® was utilized to develop a new question bank. Fifty de-identified intraoral radiographs were obtained from the UCLA SOD Oral Radiology Clinic. All radiographs were acquired using a Gendex Expert DC (Gendex Dental Systems, Hatfield, PA, USA) and XDR Anatomic Sensor size 1 and 2 (Cyber Medical Imaging Inc., Los Angeles, CA, USA), with the aid of RINN SCP-ORA film-holding devices (DENTSPLY Rinn, York, PA, USA).

An oral and maxillofacial radiology resident (S.J.) and a board-certified oral and maxillofacial radiologist with over 30 years of experience (S.M.) reviewed the radiographs, excluding those with unacceptable image quality. This process resulted in the selection of 10 periapical and 20 bitewing radiographs for further evaluation. The two observers assessed the radiographs and referred to AI-detected lesions from Second Opinion®. Since the AI platform highlights carious lesions but does not assign severity grading, lesions were categorized based on the extent of the AI annotation (Figure 3). The resulting classification included 20 E0, 25 E1/E2, 25 D1/D2, and 2 D3 lesions, as summarized in Table 1.

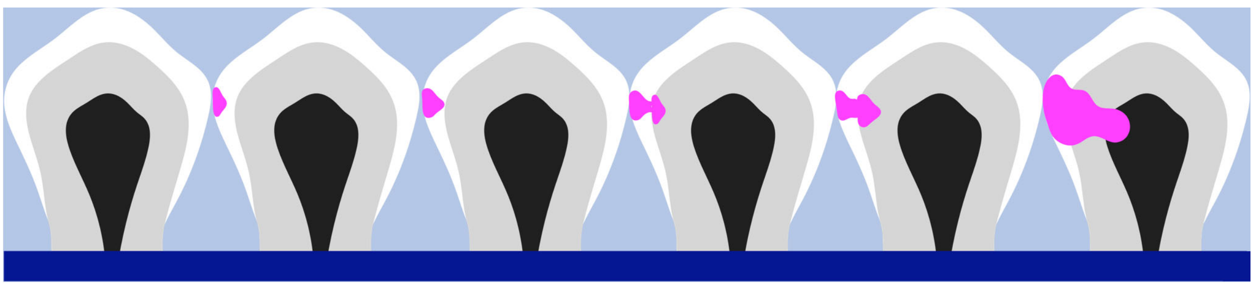


Figure 3. Caries Classification Criteria for Tentative Lesion Categorization. This diagram illustrates the classification system used to categorize AI-detected lesions. Since the Second Opinion® highlights lesions in pink without severity grading, this framework was applied to classify lesions based on depth and progression. Lesions extending up to the dentinoenamel junction (DEJ) were categorized as E2, distinguishing them from more advanced stages involving dentin.

Four faculty members independently reviewed these radiographs without AI annotations, classifying each lesion as E0, E1/E2, D1/D2, or D3. Radiographs were included if at least three out of four faculty members (≥75% agreement) classified them in the same category. A total of 5 periapical and 16 bitewing radiographs met the consensus threshold, finalizing the selection at 16 E0, 20 E1/E2, 22 D1/D2, and 2 D3 lesions, as shown in Table 1.

	E0	E1/E2	D1/D2	D3	Total
AI + Preliminary Review (n)	20	25	25	2	72
Final Faculty Consensus (n)	16	20	22	2	60
Difference in Lesion Count (n, %)	4 (20%)	5 (20%)	3(12%)	-	12(16.7%)

Table 1. Comparison of AI-Assisted Preliminary Review and Faculty Consensus in Lesion Classification. This table compares the lesion counts from the AI-assisted preliminary review to the final faculty consensus. Faculty evaluation refined classifications, reducing the total lesions from 72 to 60.

Three months later, the same four faculty members reassessed the radiographs with AI annotations. Faculty were instructed to refer to the AI annotations but had full discretion to accept or reject them. The structured workflow for developing the Level 1 caries detection exam in Cohort 2, from radiograph selection to final inclusion based on faculty agreement, is outlined in Figure 4.

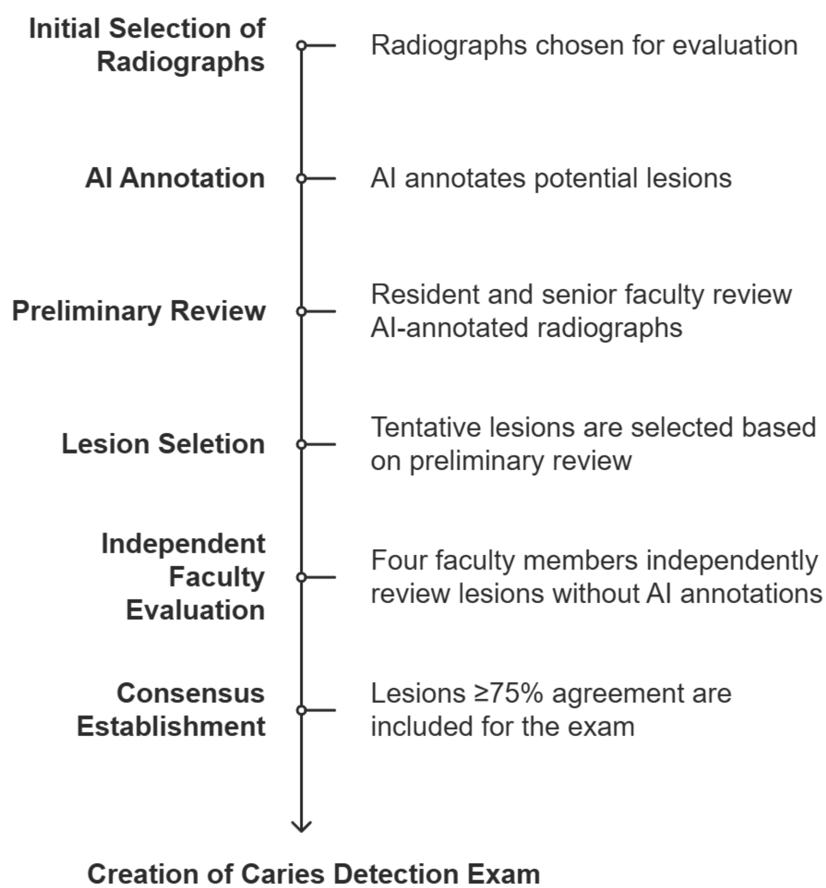


Figure 4. Workflow for Developing the Level 1 Caries Detection Exam in Cohort 2. This figure illustrates the process for creating the level 1 caries detection exam in Cohort 2. Selected radiographs were screened by AI to annotate potential lesions. A resident and senior faculty conducted a preliminary review, followed by tentative lesion selection. Four faculty members independently evaluated the lesions without AI assistance, and only those with ≥75% consensus were included in the level 1 caries detection exam in Cohort 2.

4. Data Analysis

4.1 Student Performance Metrics

Score distribution, average response time, and standard deviation were analyzed to evaluate overall performance on the caries detection exam in Cohorts 1 and 2, as well as the self-assessment quiz in Cohort 2. The number of students passing at each attempt was recorded to track progression and exam success patterns. No formal analysis was conducted for the Cohort 1 self-assessment quiz, as it was not designed to replicate the conditions of the caries detection exam.

4.2 Diagnostic Performance Metrics

The diagnostic accuracy of students, faculty, and the Second Opinion® AI platform was evaluated against the gold standard, defined by faculty consensus. Responses were classified into four categories:

- **True Positive** (TP): Correct identification of caries where both presence and extent (e.g., E0, E1/E2, D1/D2, D3) match the gold standard.
- **False Negative** (FN): A carious lesion that is either missed entirely (diagnosed as sound) or underestimated in severity (e.g., diagnosed as E1 when the gold standard is D1).

- **False Positive** (FP): A sound surface incorrectly identified as caries or a carious lesion overestimated in severity (e.g., diagnosed as D1 when the gold standard is E1).
- **True Negative** (TN): A sound surface correctly identified as sound.

From these classifications, the following performance metrics were calculated:

- **Sensitivity** = TP / (TP + FN)
- **Specificity** = TN / (TN + FP)
- **Accuracy** = (TP + TN) / (TP + TN + FP + FN)
- **Precision** = TP / (TP + FP)
- **F1 Score** = 2 × ((Precision × Sensitivity) / (Precision + Sensitivity))

4.3 Statistical Analyses

The Chi-square test or Fisher's Exact test ($p < 0.05$) was used to compare diagnostic performance among Cohort 1, Cohort 2, and the AI platform.

Faculty agreement on caries diagnosis was analyzed under two conditions—without and with AI annotations—by categorizing agreement as full (4/4), majority (3/4), or split (2/4). A Chi-square test with Yates' correction ($p < 0.05$) was conducted to assess statistical significance between conditions. Additionally, agreement levels were examined across lesion severities (E0, E1/E2, D1/D2, and D3) to evaluate variations in AI impact at different lesion depths.

Intra-observer reliability for each faculty member across the two time points using Cohen's κ[43]. Faculty diagnostic performance for different lesion severities (E0, E1/E2, D1/D2, D3) was also compared between assessments without and with AI annotations.

The Chi-square or Fisher's Exact test ($p < 0.05$) was used to compare diagnostic performance across lesion severities.

RESULTS

1. AI versus Student Diagnostic Performance

1.1 Diagnostic Metrics Comparison

The diagnostic performance of the AI platform (Second Opinion®) and the student cohorts is presented in Table 2. Across sensitivity, specificity, accuracy, and F1 score, Second Opinion® consistently outperformed both Cohort 1 and Cohort 2 student groups. The AI platform achieved a sensitivity of 89.5% and an accuracy of 93.6%, compared to Cohort 1 Caries Detection Exam results of 52.6% sensitivity and 51.1% accuracy. Similarly, students in Cohort 2 demonstrated lower sensitivity (31.6%) and accuracy (40.4%) in the self-assessment compared to AI. Although Cohort 1 and Cohort 2 students underwent different levels of training and assessment formats, the AI platform outperformed both groups.

A significant improvement was observed in Cohort 2 Caries Detection Exam performance following structured self-assessment. Sensitivity increased from 31.6% in the self-assessment to 50.0%, specificity rose from 46.4% to 93.8%, and precision improved from 28.6% to 95.7%. Accuracy also increased from 40.4% to 61.7%, while the F1 score more than doubled (30.0% to 65.7%). Despite AI maintaining the highest diagnostic performance, Cohort 2 students exhibited marked gains in accuracy and sensitivity following structured self-assessment, reinforcing the value of preparatory exercises before formal evaluations.

	AI (Second Opinion®)	Cohort 1 Caries Detection	Cohort 2 Self-Assessment	Cohort 2 Caries Detection
Sensitivity	89.5%	52.6%	31.6%	50.0%
Specificity	96.4%	50.0%	46.4%	93.8%
Accuracy	93.6%	51.1%	40.4%	61.7%
Precision	94.4%	41.7%	28.6%	95.7%
F1 Score	91.9%	46.5%	30.0%	65.7%

TABLE 2. Diagnostic Performance Metrics for AI, Cohort 1 Caries Detection, Cohort 2 Self-Assessment, and Cohort 2 Caries Detection Exam. AI outperformed both student groups across most metrics. Cohort 2 demonstrated significant improvements after completing the self-assessment, achieving higher accuracy and sensitivity in their caries detection exam.

1.2 Statistical Analysis

A Chi-square test confirmed that AI's diagnostic accuracy was significantly higher than that of the Cohort 1 Caries Detection Exam group ($\chi^2 = 21.26$, $p < 0.05$) and the Cohort 2 Self-Assessment group ($\chi^2 = 30.08$, $p < 0.05$). However, there was no statistically significant difference between the Cohort 1 Caries Detection Exam and Cohort 2 Self-Assessment results ($\chi^2 = 1.08$, $p = 0.299$), despite the latter exhibiting lower sensitivity (31.6% vs. 52.6%) and a lower F1 score (30.0% vs. 46.5%). These findings indicate that the most pronounced differences in caries detection performance were observed between AI and human raters during their initial or intermediate diagnostic attempts, rather than between the two student cohorts at the same stage of evaluation.

2. Student Performance Analysis

2.1 Overall Score Distribution and Timing

Figures 5A, 5B, and 5C compare the score distributions, average response times, and standard deviations for the Cohort 1 Caries Detection Exam, Cohort 2 Self-Assessment, and Cohort 2 Caries Detection Exam. The Cohort 1 Caries Detection Exam (Figure 5A) produced an average score of 81%, a low standard deviation of 0.95, and an average completion time of approximately 8 minutes 25 seconds. Given that each exam question set featured 10 surfaces, Cohort 1 students effectively spent about 51 seconds per surface.

The Cohort 2 Self-Assessment (Figure 5B) yielded a lower mean score of 71%, coupled with a higher standard deviation of 2.28, and required around 6 minutes in total—approximately 36 seconds per surface. This faster pace likely contributed to the observed variability in performance, as students completed the assessment more rapidly but with greater inconsistencies in their responses.

Figure 5A. Cohort 1 Caries Detection Exam

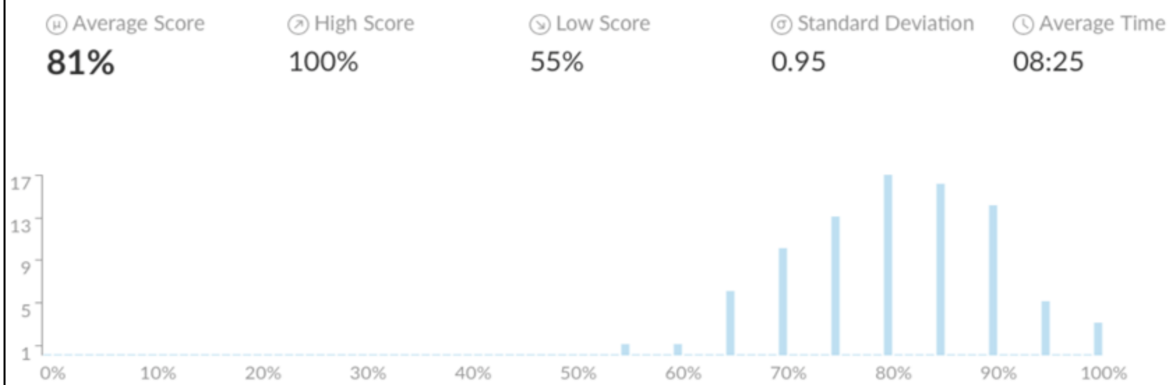


Figure 5B. Cohort 2 Self-Assessment Quiz

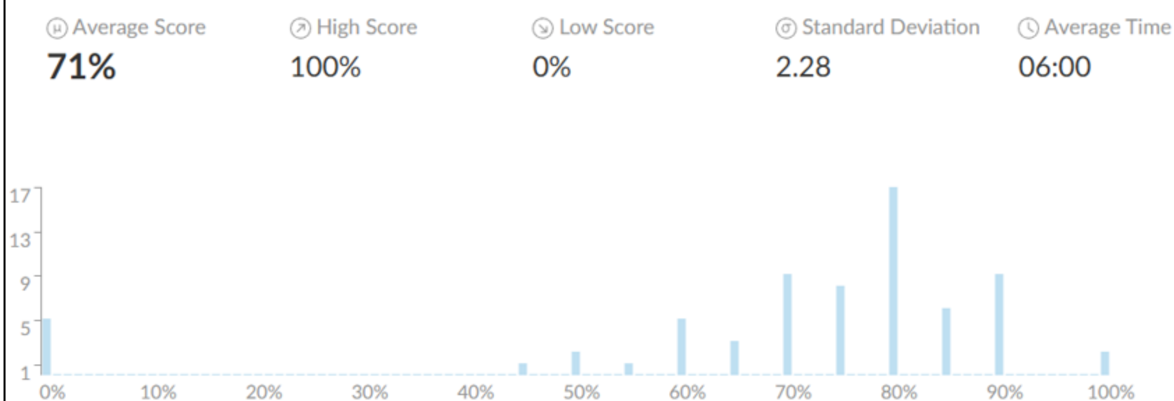
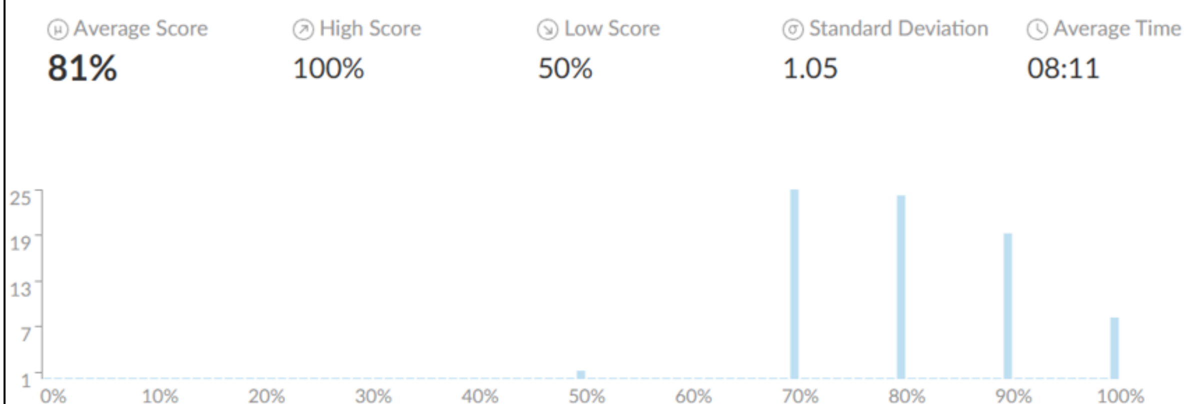


Figure 5C. Cohort 2 Caries Detection Exam



Figures 5A, 5B, and 5C. Score distribution, average response time, and standard deviation for Cohort 1 Caries Detection Exam, Cohort 2 Self-Assessment, and Cohort 2 Caries Detection Exam. These figures illustrate differences in student performance and variability. Figure 5A represents the score distribution for the Cohort 1 Caries Detection Exam. Figure 5B shows the Cohort 2 Self-Assessment, where students had unrestricted reattempts, resulting in greater variability in scores and a tendency for faster responses. Figure 5C presents the Cohort 2 Caries Detection Exam, demonstrating improved performance with higher scores and reduced variability, suggesting that prior exposure to the self-assessment contributed to better exam outcomes.

The Cohort 2 Caries Detection Exam (Figure 5C) achieved an average score of 81%, mirroring the mean performance of Cohort 1, but with a decreased standard deviation of 1.05, indicating more consistent results. Students took about 8 minutes 11 seconds (49 seconds per surface) to complete the exam, which is longer than the self-assessment setting yet more closely aligns with Cohort 1 timing. This more deliberate pace, combined with reduced variability, suggests that Cohort 2 students benefited from structured self-assessment prior to their formal exam. The overall improvement in score consistency, along with the longer completion time, highlights the self-assessment's role in familiarizing students with the exam format and reinforcing more careful diagnostic reasoning during the caries detection exam.

2.2 Pass Rate Comparisons and Attempt Analysis

A further contrast in cohort performance is evident from the number of students passing each exam on their first, second, and subsequent attempts, as shown in Table 3. In the Cohort 1 Caries Detection Exam, 48 of 86 students passed on the first attempt, with 33 more passing on the second, yielding a 94.2% cumulative pass rate by the second attempt. Four students proceeded to a third attempt, and one required a fourth. The Cohort 2 Self-Assessment produced a similar cumulative pass rate by the second attempt (46 of 80 students on the first attempt, plus another 29 on the second, for 93.75% total), but five students required additional attempts, with one needing as many as five. By contrast, in the Cohort 2 Caries Detection Exam, 77 of 80 students passed on their first attempt (96.25%), with only three requiring a second attempt. Notably, none of the Cohort 2 students required more than two attempts to pass. Although Cohort 2 students initially performed worse in the self-assessment than their Cohort 1 counterparts, they ultimately achieved a higher first-attempt pass rate (96.25% vs. 57.5%), emphasizing the impact of structured preparation before the final evaluation.

	Cohort 1 Students Who Passed Caries Detection Exam	Cohort 2 Students Who Passed Self-Assessment Quiz	Cohort 2 Students Who Passed Caries Detection Exam
1st Attempt (n, %)	48 (55.8%)	46 (57.5%)	77 (96.25%)
2nd Attempt (n, %)	33 (38.4%)	29 (36.25%)	3 (3.75%)
3rd Attempt (n, %)	4 (4.65%)	3 (3.75%)	-
4th Attempt (n, %)	1 (1.15%)	1 (1.25%)	-
5th Attempt (n, %)	-	1 (1.25%)	-

TABLE 3. Number of students passing at each attempt for the Cohort 1 Caries Detection Exam, Cohort 2 Self-Assessment, and Cohort 2 Caries Detection Exam. The table presents the number of students passing at each attempt, highlighting differences in performance between groups. While Cohort 1 and Cohort 2 Self-Assessment showed similar cumulative pass rates by the second attempt, Cohort 2 Caries Detection Exam exhibited the highest first-attempt pass rate, indicating an improvement after structured preparation.

A comparison between student performance and the AI platform provides additional perspective on the effectiveness of preparation and assessment, as shown in Table 4. This table categorizes first-attempt results as better than AI, equal to AI, or worse than AI. Among Cohort 1 Caries Detection Exam takers, 6 students (6.97%) outperformed AI and 7 students (8.13%) matched AI's performance, while the majority (84.9%) scored below AI. In the Cohort 2 Self-Assessment, no students outperformed AI, and only 5 students (6.25%) performed at an equivalent level, leaving 75 students (93.75%) below AI performance. By contrast, in the Cohort 2 Caries Detection Exam, 8 students (10.0%) scored higher than AI, and 19 students (23.8%) equaled AI, reducing the proportion below AI to 53 students (66.2%). These results demonstrate the marked progress made by Cohort 2 students after structured preparation, as they increasingly matched or surpassed AI performance by the time of their caries detection exam.

	Cohort 1 Caries Detection Exam Students	Cohort 2 Self-Assessment Students	Cohort 2 Caries Detection Exam Students
Better than AI (n, %)	6 (6.97%)	0 (0%)	8 (10.0%)
Equal to AI (n, %)	7 (8.13%)	5 (6.25%)	19 (23.8%)
Worse than AI (n, %)	73 (84.9%)	75 (93.75%)	53 (66.2%)
Percentages represent student performance on the first attempt only			

TABLE 4. Comparison of Cohort 1 and Cohort 2 Students' Performance Relative to AI Diagnosis (First Attempt Only). Student performance on their first attempt is categorized as better than AI, equal to AI, or worse than AI. Cohort 2 students in the caries detection exam demonstrated an improvement, with 10% performing better than AI and 23.8% performing equally to AI, compared to 0% and 6.25%, respectively, in the self-assessment.

3. Faculty Performance Without and With AI

3.1 Overall Diagnostic Metrics

Faculty diagnostic data before and after the incorporation of AI annotations are shown in Table 5. Three of the four faculty members achieved higher accuracy once AI-based lesion overlays were available, reflecting improvements in both sensitivity and specificity. Faculty D exhibited the most substantial increase in specificity (from 62.5% to 93.8%), underscoring a greater ability to correctly identify sound surfaces. Although Faculty A experienced a decline in specificity (from 100% to 93.8%), overall accuracy remained high.

	Condition	TP	FN	TN	FP	Sensitivity	Specificity	Accuracy	Precision	F1 Score	Cohen's κ
Faculty A	Without AI	44	0	16	0	100%	100%	100%	100%	100%	0.915
	With AI	43	1	15	1	97.7%	93.8%	96.7%	97.7%	97.7%	
Faculty B	Without AI	38	6	16	0	86.4%	100%	90.0%	100%	92.7%	0.916
	With AI	40	4	16	0	90.9%	100%	93.3%	100%	95.2%	
Faculty C	Without AI	42	2	16	0	95.5%	100%	96.7%	100%	97.7%	0.958
	With AI	43	1	16	0	97.7%	100%	98.3%	100%	98.9%	
Faculty D	Without AI	42	2	10	6	95.5%	62.5%	86.7%	87.5%	91.3%	0.705
	With AI	43	1	15	1	97.7%	93.8%	96.7%	97.7%	97.7%	

TABLE 5. Faculty Diagnostic Performance Metrics Without and With AI Annotation. Faculty diagnostic performance metrics, including sensitivity, specificity, accuracy, precision, and F1 score, are compared without and with AI annotation. AI annotation improved sensitivity, accuracy, and agreement (Cohen's κ) for most faculty members while maintaining high specificity and precision.

Improvements in sensitivity across multiple faculty members suggest that AI helped reduce missed lesions (false negatives). Cohen's κ values were strong for most faculty, indicating stable or enhanced intra-observer reliability with the addition of AI annotations. Faculty C attained the highest Cohen's κ value (0.958), signifying near-perfect agreement and highly consistent diagnostic interpretations with AI, while Faculty D had the lowest (0.705), falling slightly below the threshold for strong agreement but still within a moderately strong range. Although precision and F1 scores varied among individuals, there was no evidence that AI negatively impacted diagnostic performance. Overall, the data point toward a net gain in both consistency and accuracy for the majority of evaluators.

3.2 Performance by Lesion Severity

Results categorized by lesion severity (E0, E1/E2, D1/D2, D3) are shown in Table 6. Although the data reveal instances of improvement in diagnostic accuracy—particularly for Faculty D, who rose from 62.5% to 93.75% correct diagnoses at the E0 level (OR = 0.312, $p = 0.273$) and showed gains at D1/D2 (OR = 0.157, $p = 0.118$)—none of these changes reached statistical significance ($p > 0.05$). Faculty A, who approached 100% accuracy on multiple lesion types without AI, experienced slight declines in E0 and E1/E2 diagnoses after AI overlays were introduced (E1/E2, OR = 1.711, $p = 0.719$), but these differences were also non-significant. In the D3 category, which contained a very small sample size, there was no change in performance (OR =

1.000, p = 1.000), underscoring the difficulty of drawing conclusive statistical inferences from limited data. Although these findings suggest that AI-driven improvements may be more pronounced for certain lesion depths, modest sample sizes did not permit definitive statistical conclusions.

		Faculty A		Faculty B		Faculty C		Faculty D		Fisher's OR	p-Value
		Without AI	With AI	Without AI	With AI	Without AI	With AI	Without AI	With AI		
E0	True diagnose	16 100%	15 93.75%	16 100%	16 100%	16 100%	16 100%	10 62.5%	15 93.75%	0.312	0.273
	False diagnose	0 0%	1 6.25%	0 0%	0 0%	0 0%	0 0%	6 37.5%	1 6.25%		
E1/E2	True diagnose	20 100%	19 95%	19 95%	18 90%	18 90%	19 95%	20 100%	19 95%	1.711	0.719
	False diagnose	0 0%	1 5%	1 5%	2 10%	2 10%	1 5%	0 0%	1 5%		
D1/D2	True diagnose	22 100%	22 100%	18 81.82%	21 95.45%	22 100%	22 100%	20 90.91%	22 100%	0.157	0.118
	False diagnose	0 0%	0 0%	4 18.18%	1 4.55%	0 0%	0 0%	2 9.09%	0 0%		
D3	True diagnose	2 100%	2 100%	1 50%	1 50%	2 100%	2 100%	2 100%	2 100%	1.000	1.000
	False diagnose	0 0%	0 0%	1 50%	1 50%	0 0%	0 0%	0 0%	0 0%		

TABLE 6. Faculty Diagnostic Accuracy by Lesion Severity Without and With AI Annotation. Faculty diagnostic performance for different lesion severities (E0, E1/E2, D1/D2, D3) is compared without and with AI annotation. The table shows the proportion of true and false diagnoses, along with Fisher's odds ratio (OR) and p-values. AI annotation improved diagnostic accuracy for most lesion categories, particularly for Faculty D at the E0 and D1/D2 levels, while other faculty members showed consistent performance. The statistical analysis indicates no significant differences ($p > 0.05$) in accuracy with AI assistance.

To provide a broader perspective on diagnostic accuracy across different groups, Table 7 presents true diagnosis rates (%) by lesion severity for the AI platform, student cohorts, and faculty—both without and with AI annotation. Faculty with AI support generally achieved the highest correct diagnosis rates, particularly for E0 (96.9%) and D1/D2 (98.9%), surpassing their own performance without AI (90.6% and 93.2%, respectively) as well as outperforming student groups and the AI platform in certain categories. Second Opinion® consistently demonstrated strong diagnostic accuracy across lesion severities, particularly at the D1/D2 stage (96.0%). Among student groups, Cohort 2 exhibited higher accuracy than Cohort 1 for D1/D2 lesions (40.0% vs. 0%), reflecting the benefits of structured self-assessment and iterative learning. However, both cohorts struggled with D3 lesions, where neither group achieved successful identification, likely due to the small sample size.

	AI True Diagnosis %	Cohort 1 True Diagnosis %	Cohort 2 True Diagnosis %	Faculty Without AI True Diagnosis %	Faculty With AI True Diagnosis %
E0	94.5%	50.0%	63.6%	90.6%	96.9%
E1/E2	95.6%	76.9%	54.5%	96.3%	93.8%
D1/D2	96.0%	0%	40.0%	93.2%	98.9%
D3	80.0%	0%	0%	87.5%	87.5%

TABLE 7. True Diagnosis Rates (%) by Carious Lesion Severity Across AI, Students, and Faculty Groups. True diagnosis rates for different carious lesion severities (E0 to D3) are shown across AI, student cohorts, and faculty groups. Faculty with AI annotation generally achieved higher rates compared to AI, faculty without AI, and students.

3.3 Inter-Faculty Agreement

The shift in faculty agreement levels following AI annotation is reported in Table 8. Full consensus (4/4) increased from 73.33% without AI to 86.67% with AI, while 3/4 agreement decreased from 26.67% to 11.67%. Only 1.66% of cases fell into 2/4 agreement when AI was introduced, suggesting that the AI annotations resolved borderline discrepancies more often than they created new ones.

	Without AI (n, %)	With AI (n, %)	Percentage Difference (%)
Full agreement (4/4)	44 (73.33%)	52 (86.67%)	13.34%
Majority agreement (3/4)	16 (26.67%)	7 (11.67%)	-15.00%
Split agreement (2/4)	0 (0.0%)	1 (1.66%)	1.66%

TABLE 8. Faculty Agreement on Caries Diagnosis Without and With AI Annotation. The table compares faculty agreement levels on caries diagnosis without and with AI annotation, categorized as full agreement (4/4), majority agreement (3/4), and split agreement (2/4). AI annotation increased full agreement (4/4) by 13.34%, while partial agreement (3/4) decreased by 15%, suggesting greater consensus with AI assistance. The slight increase (1.66%) in 2/4 agreement indicates minor discrepancies introduced by AI, but overall alignment improved.

To further analyze these agreement shifts by lesion severity, Table 9 presents a breakdown of faculty agreement across different lesion categories (E0, E1/E2, D1/D2, D3). AI annotation led to increased 4/4 agreement in E0 and D1/D2, while E1/E2 remained largely stable (85% to 80%). D3 lesions showed no change in full agreement (50% both with and without AI), likely due to the small sample size. Notably, in the D1/D2 category, full agreement improved from 72.7% to 95.5%, while 3/4 agreement was significantly reduced (27.3% to 4.5%). These results suggest that AI annotation had the most pronounced effect on cases with moderate dentinal involvement, potentially aiding faculty consensus in more diagnostically challenging scenarios.

	Full agreement (4/4) (n, %)		Majority agreement (3/4) (n, %)		Split agreement (2/4) (n, %)	
	Without AI	With AI	Without AI	With AI	Without AI	With AI
E0	10 (62.5%)	14 (87.5%)	6 (37.5%)	2 (12.5%)	-	-
E1/E2	17 (85%)	16 (80%)	3 (15%)	3 (15%)	0 (0%)	1 (5%)
D1/D2	16 (72.7%)	21 (95.5%)	6 (27.3%)	1 (4.5%)	-	-
D3	1 (50%)	1 (50%)	1 (50%)	1 (50%)	-	-

TABLE 9. Faculty Agreement on Caries Diagnosis by Lesion Severity Without and With AI Annotation. The table details full agreement (4/4), majority agreement (3/4), and split agreement (2/4) levels for each lesion severity, illustrating how AI affected faculty consensus across different caries classifications.

A Chi-square test with Yates' correction indicated that the differences in faculty agreement levels between conditions without and with AI annotations were not statistically significant ($\chi^2 = 5.19$, $p = 0.075$). Overall, these findings indicate that most faculty benefited from AI support when interpreting ambiguous radiographs, achieving higher unanimity. The small fraction of additional 2/4 outcomes implies that a few cases prompted disagreement when AI annotations conflicted with an individual's initial impression. Nonetheless, the net effect was increased overall alignment, consistent with the observed gains in sensitivity, specificity, and accuracy for the majority of faculty members.

DISCUSSION

This study investigated whether an AI platform (Second Opinion®) could enhance caries detection training and diagnostic performance in a dental school setting. The results indicate that AI not only outperformed student groups in multiple diagnostic metrics but also enhanced diagnostic consistency among faculty. These findings align with growing interest in using AI to modernize and optimize dental education, a need recognized by both researchers and clinicians[1, 12, 34, 44].

A key motivation for this work arises from the longstanding challenges of detecting dental caries on radiographs, especially in early enamel lesions. Systematic reviews have reported sensitivities ranging from 24% to 42% for radiographic detection of enamel caries[11, 12], and a mean sensitivity of about 41% in proximal enamel-only caries, with specificity near 78%[12]. Such a modest performance underscores the difficulties both novice and experienced clinicians face, a point corroborated by other studies reporting wide ranges of diagnostic accuracy[13, 44, 45]. By providing a more objective approach to lesion identification, AI can reduce variability introduced by factors such as fatigue, cognitive biases, and inconsistencies in training[3].

Comparisons between the AI platform and second-year dental students (Cohort1 and Cohort2) align with other investigations suggesting that AI can outperform human raters in caries detection. In one study, Liu et al[8]. reported that an AI model outperformed junior dentists in diagnostic accuracy, while Cantu et al.[31] showed that a neural network achieved superior sensitivity and accuracy. The AI system in this study likewise exhibited consistently higher sensitivity and specificity, both critical for minimizing missed diagnoses and preventing unnecessary interventions.

Beyond demonstrating strong overall performance, this study also illustrates how students benefited from structured preparation and repeated practice. Because AI-assisted workflows expedited the creation of the new caries detection exam for Cohort 2, the Cohort 1 caries detection exam was repurposed as a self-assessment quiz for Cohort 2. Unlike the self-assessment for Cohort 1—which featured only five questions, no time limit, and no passing requirement—the self-assessment quiz for Cohort 2 mirrored the caries detection exam in both structure and rigor. This additional exposure—effectively a dress rehearsal matching the difficulty level of the final exam—helped improve Cohort 2's performance. Their sensitivity increased from 31.6% in the self-assessment to 50.0% in the caries detection exam, while accuracy improved from 40.4% to 61.7%. Such improvements parallel findings in medical education, where AI-driven training has yielded skill gains of up to 30%[23, 24]. Although Cohort 1 initially posted stronger diagnostic metrics, Cohort 2 ultimately matched or surpassed them in key measures (e.g., first-attempt pass rates), highlighting the value of formative assessments and iterative feedback loops in accelerating learning.

The development of the Cohort 2 caries detection exam was also streamlined by AI. By generating preliminary annotations of potential carious lesions, the AI helped faculty reach consensus more quickly, making a traditionally human-exclusive task more efficient. Such consistency aligns with studies showing that AI can expand educational resources without significantly increasing faculty's administrative burden [13, 29, 46]. The ability to create standardized test sets "on demand" could benefit both high-stakes exams and more frequent, lower-stakes assessments[18, 28, 34]. If AI can be used frequently for quizzes—always reflecting up-to-date diagnostic criteria and diverse cases—students may benefit from ongoing, adaptive learning cycles[15, 34]. While this investigation focused on caries detection, parallel efforts are examining AI for periodontal bone loss, endodontic lesions, and craniofacial anomalies[13, 26, 47-49].

Another significant finding is the improvement in faculty diagnostic consistency with AI. Although baseline faculty accuracy was already high, three out of four faculty members exhibited increased sensitivity after adopting AI, and the 4/4 agreement rate rose from 73.33% to 86.67%. One faculty member (Faculty D) experienced a notable jump in specificity (from 62.5% to 93.8%), suggesting that AI overlays may help experienced clinicians avoid “overcalling” cervical burn-out and Mach band optical illusions as genuine lesions. By providing a reliable “second set of eyes,”[28] AI encourages faculty to converge on unified interpretations, reducing the inconsistent feedback that can confuse students and diminish their confidence. Importantly, AI-assisted workflows may ease faculty workloads[13] by cutting down on repetitive tasks[26] like lesion-marking[8] and repeated calibration sessions, allowing instructors to devote more time to case-specific mentoring[15, 42].

A further breakdown of faculty agreement by lesion severity suggests that AI had its greatest impact on D1/D2 lesions. Full agreement for these cases rose from 72.7% to 95.5%, while 3/4 agreement decreased from 27.3% to 4.5%. This pattern indicates that AI was particularly effective in improving consensus on caries extending into initial to moderate dentinal areas. Additionally, for E0 lesions, AI increased full agreement from 62.5% to 87.5%, reinforcing its potential to help faculty consistently recognize the absence of caries. In contrast, agreement in E1/E2 remained relatively stable (85% to 80%), suggesting that AI did not substantially alter interpretations of early enamel lesions. For D3 lesions, agreement stayed at 50% both with and without AI, likely reflecting the small sample size.

It remains crucial, however, to protect the interpretive skills and autonomy of the dentist. AI can err when radiographs contain overlapping structures or poor image quality[1, 3, 13]. This study employed mostly high-quality radiographs and excluded those with severe artefacts, limiting its generalizability to more complex real-world conditions[3, 10]. A small fraction of cases shifted from unanimous faculty agreement to a 2/4 split after AI intervention, illustrating that technology can occasionally introduce ambiguity. Viewing such cases as opportunities to reinforce evidence-based reasoning can help students hone critical thinking[29, 36, 38]. Another limitation is that this work was conducted at a single institution—UCLA's School of Dentistry—using one radiographic system. Future multicenter studies could broaden patient demographics, incorporate diverse imaging equipment, and capture a wider range of lesion severities, including more advanced (D3) lesions[3, 18, 29, 45]. Additionally, gathering subjective feedback about participants' attitudes toward AI may clarify whether acceptance of the technology influences diagnostic confidence[24]. Further investigations could also compare junior-level clinicians (e.g., 5–10 years of experience)[8] with both novices and seasoned practitioners, offering a more nuanced view of how experience intersects with AI performance. Finally, measuring lesion-specific interpretation times might reveal which severity levels clinicians find most challenging and whether AI accelerates or slows detection across varying severities.

Overall, these findings confirm that AI-assisted caries detection offers tangible benefits for both students and faculty. The gains in student diagnostic performance, improved faculty consensus, and reduced retest rates suggest that AI can serve as a valuable adjunct to traditional dental education. However, careful implementation is necessary. Future efforts must prioritize rigorous validation, structured AI training for users[28, 36, 38, 50], and consideration of ethical implications[32] to ensure sustainable AI integration in clinical and educational settings[25]. In essence, AI should serve as a complement, not a replacement, for clinician-led education[3, 29]. With thoughtful curriculum design, AI can standardize and reinforce radiographic interpretation, guiding dental education toward a future in which advanced digital tools and expert human judgment converge to enhance patient outcomes[26, 27, 42]. By striking this balance, educators can harness the strengths of artificial intelligence to elevate diagnostic standards, reduce variability, and reimagine dental training for the twenty-first century[29].

CONCLUSIONS

- The AI platform consistently outperformed second-year dental students in sensitivity, specificity, and overall accuracy, reinforcing its potential as a reliable tool for caries detection.
- Incorporating AI-assisted exam development and structured practice opportunities enhances student engagement, promotes more deliberate diagnostic reasoning, and improves performance on high-stakes assessments.
- While Cohort 2 initially underperformed relative to Cohort 1, structured practice and repeated exposure led to substantial gains that ultimately matched or surpassed key metrics of Cohort 1.
- Using AI to annotate potential lesions streamlined the creation of the Cohort 2 caries detection exam, saved faculty time, reduced repeated calibration efforts, and demonstrated how AI can support test-set development in educational contexts.
- Introducing AI annotations increased unanimous faculty agreement, indicating that AI can serve as a valuable second opinion to reduce inter-faculty variability in radiographic caries detection, especially for D1/D2 lesions and E0 lesions.
- While most faculty achieved higher diagnostic accuracy and consistency with AI, a few lesions shifted from full consensus to partial disagreement, reflecting occasional ambiguities introduced by overlays and reinforcing the importance of clinical judgment.

• AI can greatly enhance detection but must complement rather than replace clinician-led diagnosis, reinforcing the importance of ethical and educational safeguards that protect critical thinking and professional accountability.

• High-quality radiographs and clear lesion classification were pivotal for successful AI integration. Future studies should address image artefacts, overlapping anatomical structures, and broader patient populations to ensure generalizability and applicability.

• Continued research should examine AI's utility across diverse clinical settings, lesion severities, and imaging modalities. By refining dental curricula, dental education can harness AI's advantages without undermining essential human expertise.

REFERENCES

1. Lee, S., et al., *Deep learning for early dental caries detection in bitewing radiographs.* Scientific Reports, 2021. **11**(1): p. 16807.
2. Adnan, N., et al., *Developing an AI-based application for caries index detection on intraoral photographs.* Sci Rep, 2024. **14**(1): p. 26752.
3. Alam, M.K., et al., *AI-Based Detection of Dental Caries: Comparative Analysis with Clinical Examination.* J Pharm Bioallied Sci, 2024. **16**(Suppl 1): p. S580-S582.
4. Kassebaum, N.J., et al., *Global burden of untreated caries: a systematic review and metaregression.* J Dent Res, 2015. **94**(5): p. 650-8.
5. Bagramian, R.A., F. Garcia-Godoy, and A.R. Volpe, *The global increase in dental caries. A pending public health crisis.* Am J Dent, 2009. **22**(1): p. 3-8.
6. Srivastava, M.M., et al., *Detection of tooth caries in bitewing radiographs using deep learning.* arXiv preprint arXiv:1711.07312, 2017.
7. Featherstone, J.D., *The science and practice of caries prevention.* J Am Dent Assoc, 2000. **131**(7): p. 887-99.
8. Liu, Y., et al., *Artificial intelligence for caries detection: a novel diagnostic tool using deep learning algorithms.* Oral Radiology, 2024. **40**(3): p. 375-384.
9. Amrollahi, P., et al., *Recent advancements in regenerative dentistry: A review.* Materials Science and Engineering: C, 2016. **69**: p. 1383-1390.
10. Karakus, R., M.U. Ozic, and M. Tassoker, *AI-Assisted Detection of Interproximal, Occlusal, and Secondary Caries on Bite-Wing Radiographs: A Single-Shot Deep Learning Approach.* J Imaging Inform Med, 2024. **37**(6): p. 3146-3159.

11. Schwendicke, F., M. Tzschoppe, and S. Paris, *Radiographic caries detection: A systematic review and meta-analysis.* J Dent, 2015. **43**(8): p. 924-33.

12. Schropp, L., et al., *Use of artificial intelligence software in dental education: A study on assisted proximal caries assessment in bitewing radiographs.* Eur J Dent Educ, 2024. **28**(2): p. 490-496.

13. Bahadir, H.S., et al., *Comparison of diagnoses made by dentistry students and by artificial intelligence dentists.* J Dent Educ, 2024.

14. Berry, H.M., Jr., *Cervical burnout and Mach band: two shadows of doubt in radiologic interpretation of carious lesions.* J Am Dent Assoc, 1983. **106**(5): p. 622-5.

15. Duong, M.T., et al., *Artificial intelligence for precision education in radiology.* Br J Radiol, 2019. **92**(1103): p. 20190389.

16. Lee, J.-H., et al., *Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm.* Journal of Dentistry, 2018. **77**: p. 106-111.

17. Vishwanathaiah, S., et al., *Artificial Intelligence Its Uses and Application in Pediatric Dentistry: A Review.* Biomedicines, 2023. **11**(3): p. 788.

18. Mahrous, A., et al., *The use of artificial intelligence and game-based learning in removable partial denture design: A comparative study.* J Dent Educ, 2023. **87**(8): p. 1188-1199.

19. Langlais, R.P., et al., *Interpretation of bitewing radiographs: Application of the kappa statistic to determine rater agreements.* Oral Surgery, Oral Medicine, Oral Pathology, 1987. **64**(6): p. 751-756.

20. Valachovic, R.W., et al., *Examiner reliability in dental radiography.* J Dent Res, 1986. **65**(3): p. 432-6.

21. Bader, J.D., D.A. Shugars, and A.J. Bonito, *Systematic reviews of selected dental caries diagnostic and management methods.* J Dent Educ, 2001. **65**(10): p. 960-8.

22. Masella, R.S. and T.J. Thompson, *Dental education and evidence-based educational best practices: bridging the great divide.* J Dent Educ, 2004. **68**(12): p. 1266-71.

23. Ayan, E., et al., *Dental student application of artificial intelligence technology in detecting proximal caries lesions.* J Dent Educ, 2024. **88**(4): p. 490-500.

24. Yang, Y.Y. and B. Shulruf, *Expert-led and artificial intelligence (AI) system-assisted tutoring course increase confidence of Chinese medical interns on suturing and ligature skills: prospective pilot study.* J Educ Eval Health Prof, 2019. **16**: p. 7.

25. Dave, M. and N. Patel, *Artificial intelligence in healthcare and education.* British Dental Journal, 2023. **234**(10): p. 761-764.

26. Surlari, Z., et al., *Current Progress and Challenges of Using Artificial Intelligence in Clinical Dentistry—A Narrative Review.* Journal of Clinical Medicine, 2023. **12**.

27. Negi, S., et al., *Artificial Intelligence in Dental Caries Diagnosis and Detection: An Umbrella Review.* Clin Exp Dent Res, 2024. **10**(4): p. e70004.

28. Anil, S., P. Porwal, and A. Porwal, *Transforming Dental Caries Diagnosis Through Artificial Intelligence-Based Techniques.* Cureus, 2023. **15**(7): p. e41694.

29. Chang, J., et al., *Artificial intelligence-assisted full-mouth radiograph mounting in dental education.* Journal of Dental Education, 2024. **88**(7): p. 933-939.

30. Devlin, H., et al., *The ADEPT study: a comparative study of dentists' ability to detect enamel-only proximal caries in bitewing radiographs with and without the use of AssistDent artificial intelligence software.* British Dental Journal, 2021. **231**(8): p. 481-485.

31. Cantu, A.G., et al., *Detecting caries lesions of different radiographic extension on bitewings using deep learning.* J Dent, 2020. **100**: p. 103425.

32. Roganovic, J., M. Radenkovic, and B. Milicic, *Responsible Use of Artificial Intelligence in Dentistry: Survey on Dentists' and Final-Year Undergraduates' Perspectives.* Healthcare (Basel), 2023. **11**(10).

33. Devlin, H., et al., *A Pilot Comparative Study of Dental Students’ Ability to Detect Enamel-only Proximal Caries in Bitewing Radiographs With and Without the use of AssistDent® Deep Learning Software.* medRxiv, 2020: p. 2020.06.15.20131730.

34. Al-Khalifa, K.S., et al., *The Use of Artificial Intelligence in Caries Detection: A Review.* Bioengineering (Basel), 2024. **11**(9).

35. Mladenovic, R., et al., *Importance of artificial intelligence in the analysis of children's CBCT imaging by dental students.* J Dent Educ, 2023. **87 Suppl 1**: p. 914-916.

36. Delgado-Ruiz, R., et al., *Generative Artificial Intelligence (Gen AI) in dental education: Opportunities, cautions, and recommendations.* J Dent Educ, 2024.

37. Dragan, I.F., et al., *ADEA-ADEE Shaping the Future of Dental Education III: The impact of scientific technologies and discoveries on oral health globally.* J Dent Educ, 2020. **84**(1): p. 111-116.

38. Kim, C.S., et al., *Artificial intelligence (A.I.) in dental curricula: Ethics and responsible integration.* Journal of Dental Education, 2023. **87**(11): p. 1570-1573.

39. Schwendicke, F., et al., *Artificial Intelligence for Caries Detection: Value of Data and Information.* Journal of Dental Research, 2022. **101**(11): p. 1350-1356.

40. Luan, H., et al., *Challenges and Future Directions of Big Data and Artificial Intelligence in Education.* Frontiers in Psychology, 2020. **11**.

41. Schwarzmaier, J., et al., *Validation of an Artificial Intelligence-Based Model for Early Childhood Caries Detection in Dental Photographs.* J Clin Med, 2024. **13**(17).

42. Zitzmann, N.U., et al., *Digital Undergraduate Education in Dentistry: A Systematic Review.* International Journal of Environmental Research and Public Health, 2020. **17**(9): p. 3269.

43. Cohen, J., *A Coefficient of Agreement for Nominal Scales.* Educational and Psychological Measurement, 1960. **20**(1): p. 37-46.

44. Mertens, S., et al., *Artificial intelligence for caries detection: Randomized trial.* J Dent, 2021. **115**: p. 103849.

45. Zhang, J.W., et al., *Diagnostic accuracy of artificial intelligence-assisted caries detection: a clinical evaluation.* BMC Oral Health, 2024. **24**(1): p. 1095.

46. Katne, T., et al., *Artificial intelligence: demystifying dentistry–the future and beyond.* Int J Contemp Med Surg Radiol, 2019. **4**(4): p. D6-D9.

47. Kierce, E.A. and R.J. Kolts, *Improving Periodontal Disease Management With Artificial Intelligence.* Compendium of continuing education in dentistry (Jamesburg, N.J. : 1995), 2023. **44**(6): p. e1-e4.
48. Aminoshariae, A., J. Kulild, and V. Nagendrababu, *Artificial Intelligence in Endodontics: Current Applications and Future Directions.* J Endod, 2021. **47**(9): p. 1352-1357.
49. Monill-González, A., et al., *Artificial intelligence in orthodontics: Where are we now? A scoping review.* Orthod Craniofac Res, 2021. **24 Suppl 2**: p. 6-15.
50. Islam, N.M., et al., *Adopting artificial intelligence in dental education: A model for academic leadership and innovation.* J Dent Educ, 2022. **86**(11): p. 1545-1551.