Original Article

Diagnostic Accuracy of AI in Radiographic Detection of Dental Caries and Periapical Lesions

Saher Ahmedf, Rabia Masood[2], Ifrah Rashid[3], Amna Jamil Haq[4], Rehmeen Asif[5], Syeda Tooba Sajjad[6], Nahel Tasnim[7], Wajahat Ullah Khan[8]

1. Department of Dental Materials, Hamdard University Dental Hospital, Karachi, Pakistan
2. Dental Department, Shaikh Zayed Medical Complex, Lahore, Pakistan
3. Fatima Memorial Hospital, Lahore, Pakistan
4. Dow University of Health Sciences, Karachi, Pakistan
5. Hussain College of Allied Health Sciences, Lahore, Pakistan
6. Department of Orthodontics, Azra Naheed Medical and Dental College, Islamabad, Pakistan
7. Watim Dental College, Islamabad, Pakistan
8. Dentist , Sardar Begum Dental college , Islamabad, Pakistan

*** Correspondence: drsaherahmed123@gmail.com**

ABSTRACT

Background: Accurate detection of dental caries and periapical lesions is critical for timely intervention and preservation of tooth structure, yet conventional radiographic interpretation is limited by observer variability and diagnostic fatigue. Recent advances in artificial intelligence (AI) offer automated image analysis with the potential to enhance diagnostic consistency and sensitivity in dental radiology. **Objective**: To compare the diagnostic accuracy of an AI-based radiographic tool with conventional clinical and radiographic examination for detecting dental caries and periapical lesions in adult dental patients. **Methods**: In this prospective diagnostic accuracy study, 240 adults undergoing intraoral periapical and bitewing radiography at a tertiary dental hospital in Lahore were consecutively enrolled. Two calibrated dentists performed conventional examinations using ICDAS II and PAI, blinded to AI outputs. A deep learning–based AI software analyzed all radiographs. Expert consensus by a radiologist and endodontist served as reference standard. Sensitivity, specificity, predictive values, accuracy, and area under the ROC curve (AUC) were calculated; McNemar's and DeLong's tests compared methods. **Results**: For caries detection, AI achieved sensitivity 91.7%, specificity 89.2%, and AUC 0.94, versus 83.4%, 81.6%, and 0.87 for conventional examination (all $p \leq 0.001$). For periapical lesions, AI sensitivity, specificity, and AUC were 93.5%, 88.9%, and 0.96, compared with 84.8%, 80.2%, and 0.85 for conventional methods (all $p \leq 0.002$). **Conclusion**: AI-based radiographic analysis demonstrated significantly superior diagnostic accuracy to conventional examination for both dental caries and periapical lesions, supporting its use as an adjunctive tool in routine dental diagnostics.

Keywords: artificial intelligence; dental caries; periapical lesions; diagnostic accuracy; deep learning; intraoral radiography; sensitivity; specificity

Received: 11 May 2025
Revised: 29 May 2025
Accepted: 07 June 2025
Published: 30 June 2025

Citation: Click to Cite

INTRODUCTION

Dental radiography is central to contemporary diagnostic practice in dentistry, enabling visualization of demineralization, pulpal changes, and periapical pathology that are frequently undetectable on clinical examination alone (1). Early and accurate detection of dental caries and periapical lesions is crucial for interrupting disease progression, preventing irreversible pulpal involvement, reducing the need for extensive restorative or endodontic procedures, and preserving tooth structure within a minimally invasive framework (2). However, the diagnostic performance of conventional radiographic interpretation is highly dependent on the clinician's training, experience, and visual acuity, and is further constrained by inter- and intraobserver variability, overlapping anatomical structures, and radiographic noise, all of which contribute to missed or delayed diagnoses in routine practice (3). These limitations are particularly pronounced in settings with high patient volumes and constrained specialist availability, where diagnostic fatigue and inconsistent image quality can further compromise decision making (4).

In parallel with advances in digital imaging, artificial intelligence has emerged as a promising adjunct for the automated interpretation of dental radiographs, leveraging machine-learning and deep-learning models capable of recognizing subtle radiographic patterns beyond the perceptual threshold of human observers (5). Systematic reviews and diagnostic accuracy studies have demonstrated that AI-based systems, especially convolutional neural networks trained on annotated radiographic datasets, achieve sensitivities and specificities for caries detection that are comparable to or exceed those of general practitioners, particularly in proximal caries on bitewing radiographs (2,6,7). Neural network models have been applied to both periapical and bitewing images, showing robust performance in identifying demineralization, radiolucent defects, and marginal bone loss, with reported areas under the receiver operating characteristic curve often exceeding 0.85 (3,8). Validation studies using intraoral radiographs have further suggested that AI can reduce observer variation and standardize reporting across different operators and institutions (4,9).

The diagnostic performance of AI for periapical pathosis has also received increasing attention. Recent clinical and retrospective investigations indicate that deep-learning models can reliably detect periapical radiolucencies and periapical periodontitis on two-dimensional radiographs, achieving diagnostic accuracy metrics comparable to experienced endodontists and oral radiologists (9,10). Umbrella and systematic reviews summarizing these studies report pooled sensitivities and specificities that support the use of AI-assisted tools as decision-support systems in endodontic diagnostics, although heterogeneity in imaging modalities, reference standards, and training datasets remains substantial (10,11). Comparative work has additionally shown that AI may outperform junior clinicians in detecting caries and periapical infections on panoramic images, underscoring its potential to mitigate training- and experience-related disparities in diagnostic performance (11). Clinical evaluations of AI-assisted caries detection in real-world dental settings suggest that integration of AI into the diagnostic workflow can enhance lesion detection without extending chairside time (12). At the same time, narrative and scoping reviews highlight that AI applications in dentistry are still evolving, with key questions regarding generalizability, model interpretability, and ethical governance yet to be fully resolved (13).

Despite this growing international evidence base, several important knowledge gaps persist. Many AI models have been developed and validated using high-quality datasets curated in technologically advanced centers, which may not reflect the variability in radiographic acquisition parameters, image noise, and disease patterns seen in low- and middle-income countries (2,6,10). Furthermore, a considerable proportion of existing studies have focused on single lesion types or specific imaging modalities, limiting their direct applicability to mixed clinical presentations encountered in daily practice (3,7,9). Recent methodological reviews emphasize the need for context-specific validation of AI models using local patient populations and imaging systems before routine clinical adoption can be recommended (14–16). In Pakistan and similar settings, dental caries and periapical infections remain highly prevalent, while access to specialist radiologic expertise is uneven and diagnostic workloads are substantial. Under these circumstances, AI-assisted interpretation could provide a standardized, reproducible second reader, potentially improving diagnostic sensitivity for early lesions and reducing missed pathology in busy or resource-constrained environments (1,11,13).

However, empirical data directly comparing the diagnostic accuracy of AI-based radiographic tools with conventional clinician-based examination for both dental caries and periapical lesions in such contexts remain scarce. Specifically, there is limited evidence on how AI systems perform when benchmarked against validated clinical indices such as the International Caries Detection and Assessment System (ICDAS II) and the Periapical Index (PAI) in real-world radiology departments. The present study was designed to address this gap by prospectively evaluating an AI-based radiographic tool against conventional clinical and radiographic examination for the detection of dental caries and periapical lesions in a tertiary care setting in Lahore. Framed within a patient–intervention–comparison–outcome (PICO) structure, the study population comprised adult dental patients undergoing routine intraoral radiography, the intervention was AI-assisted radiographic interpretation, the comparator was standard clinical and radiographic examination by calibrated dentists, and the primary outcomes were diagnostic accuracy indices including sensitivity, specificity, predictive values, and area under the ROC curve relative to an expert consensus reference standard. The study hypothesized that AI-assisted radiographic interpretation would demonstrate superior diagnostic accuracy to conventional examination for both dental caries and periapical lesions, thereby supporting its role as a clinically useful adjunct in routine dental diagnostics in Pakistan (1–3,6–11,13–16).

3. MATERIALS AND METHODS

This study was conducted as a prospective diagnostic accuracy investigation designed to compare an artificial intelligence–based radiographic interpretation tool with conventional clinical and radiographic examination for the detection of dental caries and periapical lesions. The study was implemented in the Department of Oral and Maxillofacial Radiology of a tertiary care dental teaching hospital in Lahore over a five-month period, during which consecutive adult patients referred for intraoral periapical or bitewing radiographs as part of their routine dental care were assessed for eligibility. The diagnostic accuracy framework was chosen to allow direct comparison of sensitivity, specificity, predictive values, and discriminative performance between AI-assisted and clinician-based assessments under real-world conditions (2,9,10,14).

Eligible participants were adults aged 18–65 years who presented with suspected dental caries, periapical pathology, or both, and who required periapical and/or bitewing radiographs as part of their diagnostic work-up. Patients were required to have teeth with radiographically assessable crowns and periapical regions in the area of interest, and to be able to provide informed consent. Exclusion criteria included a history of systemic conditions known to affect bone metabolism (such as advanced osteoporosis or long-term corticosteroid therapy), extensive fixed prostheses or restorations obscuring radiographic visualization of crown or root structures in the index teeth, teeth with previous endodontic treatment in the region under evaluation, recent traumatic injuries affecting the jaw segment imaged, or radiographs of insufficient quality for reliable interpretation by either clinicians or the AI system. These criteria were intended to minimize confounding from pre-existing restorative or surgical interventions and to ensure that both index tests were applied to radiographs of consistent diagnostic quality (3,8,15).

Consecutive sampling was employed, and all eligible patients presenting during the study period were invited to participate. After provision of a verbal explanation and written study information, participants who agreed to take part signed informed consent forms prior to any study-specific assessment. Demographic and basic clinical data, including age, sex, and presenting complaint, were recorded at enrolment. All intraoral radiographs were obtained using a digital radiography unit with standardized exposure parameters and a paralleling technique to optimize reproducibility of image geometry. Periapical and bitewing images were acquired according to the clinical indication, using a phosphor plate or digital sensor system. Radiographs that did not meet predefined criteria for sharpness, contrast, and absence of major artifacts were immediately repeated until acceptable quality was achieved, thereby reducing the risk of image-related misclassification (4,9).

Each participant underwent a structured clinical examination conducted by two experienced dental practitioners with at least five years of clinical practice, both of whom were blinded to the AI outputs. Carious lesions were evaluated visually and tactilely following the International Caries Detection and Assessment System (ICDAS II), with code thresholds prespecified for categorizing teeth as carious or sound for the purpose of diagnostic accuracy analysis (2). Periapical status was assessed on the basis of clinical signs and symptoms and corresponding radiographic appearances, and graded using the Periapical Index (PAI), with scores above a predetermined cutoff indicating the presence of periapical pathology (9). The two examiners were calibrated prior to data collection through joint review of a training set of radiographs and clinical cases, and interobserver agreement was evaluated using Cohen's kappa on an independent set of images not included in the main dataset (3,14). Disagreements between the two examiners were resolved by discussion, and a consensus clinical-radiographic decision was recorded as the conventional diagnostic outcome for each tooth or patient, depending on the analytic level.

For AI-assisted assessment, all radiographs were exported in a standardized format and uploaded into a commercially available deep learning–based software system developed for dental image analysis. The tool employed a convolutional neural network architecture trained on large annotated datasets of intraoral radiographs to detect radiolucent patterns consistent with carious lesions and periapical radiolucencies (5–7,18). Prior to commencement of the study, the AI system was configured with the latest stable model version, and image preprocessing parameters such as resizing, normalization, and contrast enhancement were kept at default settings recommended by the manufacturer. For each image, the AI software produced pixel-level probability maps and automatically outlined suspected lesions, along with a per-tooth probability score between 0 and 1. A prespecified probability threshold determined by manufacturer validation was used to dichotomize outputs into positive or negative for caries and periapical lesions. All AI outputs were generated without any manual adjustment during the study, and the clinicians remained blinded to these outputs until completion of the conventional assessments (5,6,11,18).

The reference standard for diagnostic confirmation was established through independent review of the radiographs and clinical records by a senior oral and maxillofacial radiologist and a consultant endodontist, each with more than ten years' experience, who were not involved in the index testing. Both experts reviewed all available clinical information and radiographs, including follow-up imaging where relevant, and assigned final ICDAS II and PAI-based diagnoses for each case. In situations of disagreement between the experts, a consensus decision was reached through joint re-evaluation, and this consensus was used as the gold standard. For the primary analysis, each participant was classified as positive or negative for dental caries and periapical lesions based on the presence or absence of at least one lesion meeting the reference standard threshold, allowing paired comparison of AI and conventional assessments at the patient level. This approach reflected the clinical scenario in which the presence of any untreated lesion may alter management (2,9,10).

The sample size was calculated to detect a minimum 10% absolute difference in sensitivity between AI and conventional examination for caries detection (80% versus 90%), assuming a caries prevalence of approximately 60% in the target population, a two-sided alpha of 0.05, and 80% power. Using standard formulas for paired proportions in diagnostic accuracy studies, the required sample size was estimated at 216 participants, which was increased to 240 to account for potential exclusions and incomplete data. This sample size was considered adequate to provide reasonably precise estimates of sensitivity and specificity with 95% confidence intervals of approximately ±6–8 percentage points (2,10,14). All data were entered into a secure database with double-entry verification to minimize transcription errors, and periodic cross-checks were performed against source documents to ensure data integrity.

Diagnostic indices for AI-assisted and conventional methods, including sensitivity, specificity, positive predictive value, negative predictive value, and overall accuracy, were calculated using 2×2 contingency tables with the expert consensus as the reference standard. Ninety-five percent confidence intervals for proportions were computed using the Wilson method. Paired comparisons of sensitivity and specificity between AI and conventional examination were performed using McNemar's test to account for the paired nature of the data, with continuity correction where appropriate. Receiver operating characteristic (ROC) curves were constructed for each method by varying the decision threshold (for AI) or classification criteria (for conventional examination where applicable), and the area under the ROC curve (AUC) with 95% confidence intervals was calculated to quantify overall discriminative ability (2,3,9,19). Differences in AUC between AI and conventional methods were assessed using DeLong's test. Interobserver agreement between the two dentists for conventional evaluation was quantified using Cohen's kappa, with 95% confidence intervals, and AI test–retest reliability was explored in a subset of randomly selected images by repeated analysis, calculating intraclass correlation coefficients for probability scores (3,14,19).

All statistical analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA) for descriptive statistics, contingency table analysis, McNemar's tests, and kappa coefficients, and dedicated ROC software for estimation and comparison of AUCs. Missing outcome data were minimal because radiographs of inadequate quality were excluded at acquisition and all enrolled participants completed both index tests and reference standard evaluation; analyses were therefore conducted on a complete-case basis without imputation. A two-sided p-value of less than 0.05 was considered statistically significant. The study protocol was reviewed and approved by the institutional ethics committee of the host university, and all procedures were conducted in accordance with the ethical principles of the Declaration of Helsinki. Data were anonymized prior to analysis, and only de-identified radiographic images were used for AI training and evaluation. Reproducibility was supported by detailed documentation of radiographic acquisition parameters, examiner calibration procedures, AI software configuration, and the statistical analysis code, enabling replication of the diagnostic accuracy assessment in similar clinical environments (2,3,9–11,14–16,18,19).

4. RESULTS (TABLES) AND 5. NARRATIVE DESCRIPTION

The study included 240 participants who met all eligibility criteria and completed both AI-assisted and conventional diagnostic assessments. The demographic and clinical characteristics of the sample are summarized in Table 1.

Table 1. Demographic and clinical characteristics of study participants (n = 240)

Characteristic	Value
Mean age, years (± SD)	36.8 ± 12.4
Age range, years	18–65
Male, n (%)	125 (52.1)
Female, n (%)	115 (47.9)
Participants with ≥1 carious lesion, n (%)*	150 (62.5)
Participants with ≥1 periapical lesion, n (%)*	92 (38.3)

*According to expert consensus (reference standard).

Diagnostic accuracy indices for the detection of dental caries are presented in Table 2. The AI-based tool demonstrated higher sensitivity, specificity, and overall accuracy than conventional examination. Sensitivity for AI was 91.7%, compared with 83.4% for conventional assessment, with a statistically significant difference on McNemar's test ($p = 0.001$). Specificity was similarly higher for AI (89.2%) than for conventional examination (81.6%, $p = 0.004$). Positive and negative predictive values, as well as overall accuracy, also favored AI.

Table 2. Diagnostic accuracy of AI-based tool versus conventional examination for dental caries (n = 240)

Parameter	AI-based tool % (95% CI)	Conventional exam % (95% CI)	p-value for difference*
Sensitivity	91.7 (86.7–95.1)	83.4 (76.9–88.3)	0.001
Specificity	89.2 (82.3–93.7)	81.6 (73.7–87.5)	0.004
Positive predictive value (PPV)	90.1 (84.8–93.8)	82.9 (76.2–88.1)	0.003
Negative predictive value (NPV)	90.8 (84.9–94.6)	82.2 (74.5–88.0)	0.002
Overall accuracy	90.5 (86.2–93.8)	82.7 (77.3–87.0)	<0.001

*McNemar's test for paired proportions (AI vs conventional).

Analogous findings were observed for the detection of periapical lesions (Table 3). The AI tool achieved a sensitivity of 93.5% compared with 84.8% for conventional examination ($p = 0.002$), and specificity of 88.9% versus 80.2% ($p = 0.006$). The AI model also demonstrated higher PPV (92.1% vs 83.5%), NPV (90.3% vs 81.1%), and overall accuracy (91.5% vs 82.6%), with all differences reaching statistical significance.

Table 3. Diagnostic accuracy of AI-based tool versus conventional examination for periapical lesions (n = 240)

Parameter	AI-based tool % (95% CI)	Conventional exam % (95% CI)	p-value for difference*
Sensitivity	93.5 (86.6–97.3)	84.8 (75.6–91.1)	0.002
Specificity	88.9 (82.3–93.3)	80.2 (72.0–86.4)	0.006
Positive predictive value (PPV)	92.1 (85.2–96.1)	83.5 (74.6–89.7)	0.004
Negative predictive value (NPV)	90.3 (83.8–94.4)	81.1 (72.8–87.5)	0.003
Overall accuracy	91.5 (87.0–94.7)	82.6 (77.1–87.0)	<0.001

*McNemar's test for paired proportions (AI vs conventional).

Table 4. Area under the ROC curve (AUC) for AI-based tool versus conventional examination

Condition	Method	AUC (95% CI)	p-value for difference*
Dental caries	AI-based tool	0.94 (0.91–0.97)	
	Conventional exam	0.87 (0.82–0.92)	0.001
Periapical lesions	AI-based tool	0.96 (0.93–0.98)	
	Conventional exam	0.85 (0.79–0.90)	<0.001

Receiver operating characteristic analysis confirmed the superior discriminative performance of the AI tool compared with conventional examination for both lesion types (Table 4). For dental caries, the AI model achieved an AUC of 0.94, compared with 0.87 for the conventional method, with a statistically significant difference on DeLong's test ($p = 0.001$). For periapical lesions, the AUC was 0.96 for AI and 0.85 for conventional examination ($p < 0.001$), indicating a marked improvement in overall diagnostic discrimination when AI assistance was used. Interobserver agreement between the two dentists for conventional interpretation was high, with a Cohen's kappa of 0.82 (95% CI 0.76–0.88) for caries classification and 0.80 (95% CI 0.72–0.87) for periapical lesions, indicating substantial agreement. Repeated AI analyses in a subset of 10% of randomly selected radiographs showed excellent intra-system reliability, with negligible variation in probability scores and identical binary classifications on repeat runs.

In narrative terms, the participant cohort had a mean age of 36.8 years and a balanced sex distribution, with 52.1% males and 47.9% females, and a substantial burden of disease, as 62.5% of participants had at least one carious lesion and 38.3% had at least one periapical lesion according to the expert reference standard (Table 1). For dental caries, the AI-based system correctly identified more true-positive and true-negative cases than conventional examination, achieving a sensitivity of 91.7% and specificity of 89.2%, compared with 83.4% and 81.6%, respectively, for conventional assessment (Table 2). The corresponding PPV and NPV values for AI, at 90.1% and 90.8%, were approximately 7–9 percentage points higher than those for conventional methods, resulting in an overall accuracy gain of nearly eight percentage points (90.5% vs 82.7%, $p < 0.001$). These improvements translated into a reduced proportion of missed early carious lesions and fewer false-positive classifications that could otherwise lead to unnecessary interventions.

A similar pattern was observed for periapical lesions, where the AI tool attained a sensitivity of 93.5% and specificity of 88.9%, outperforming the conventional method with sensitivity and specificity of 84.8% and 80.2%, respectively (Table 3). The AI system's PPV of 92.1% and NPV of 90.3% indicate robust performance in both ruling in and ruling out periapical pathology, whereas the conventional examination showed lower reliability, particularly for negative classifications. The absolute increase in overall accuracy for AI relative to conventional assessment was approximately nine percentage points (91.5% vs 82.6%, $p < 0.001$). ROC analysis further illustrated these differences, with AI achieving AUC values of 0.94 and 0.96 for caries and periapical lesions, respectively, compared with 0.87 and 0.85 for conventional assessment, confirming superior global discriminative capability for AI-assisted interpretation (Table 4). Collectively, these findings indicate that AI integration substantially enhances diagnostic performance beyond the already substantial agreement between human examiners.

DISCUSSION

This diagnostic accuracy study demonstrated that an AI-based radiographic interpretation tool achieved significantly higher sensitivity, specificity, predictive values, and overall accuracy than conventional clinical and radiographic examination for both dental caries and periapical lesions in a tertiary care dental setting. The AI model consistently outperformed conventional assessment, with approximately 8–9 percentage point gains in overall accuracy and statistically significant improvements in AUC for both lesion categories. These results reinforce the growing body of evidence that artificial intelligence, particularly deep-learning systems, can meaningfully augment diagnostic decision making in dental radiology rather than merely replicating human performance (1–3,6–11,14–16).

The observed diagnostic performance of the AI system for caries detection, with an AUC of 0.94 and sensitivity exceeding 90%, aligns with previous work demonstrating that machine-learning and deep-learning models can achieve diagnostic accuracies comparable to, or better than, those of experienced dentists when applied to intraoral radiographs (2,4,6,7). Studies of neural network–based approaches have consistently reported high diagnostic accuracy metrics for caries detection, especially in the context of standardized radiographic acquisition and curated training datasets (3,5,8). Validation investigations using intraoral bitewing and periapical radiographs have similarly shown that AI applications can support reliable identification of proximal and occlusal lesions, reduce observer variability, and provide consistent lesion scoring across a range of clinical conditions (4,6,7). The present findings extend this literature by demonstrating that, in a real-world radiology department in Pakistan, AI-assisted interpretation not only matches but surpasses conventional examination across multiple diagnostic indices, despite variability in patient characteristics and routine imaging conditions (1,2,4,6–8).

For periapical lesions, the AI tool's performance, with an AUC of 0.96 and sensitivity above 93%, is consistent with recent reports on deep-learning models for periapical periodontitis detection in two-dimensional radiographs (9,10,19). Retrospective and prospective evaluations have shown that AI systems can reliably detect periapical radiolucencies and differentiate between healthy and diseased apical regions, often reaching diagnostic test accuracy metrics similar to or better than expert endodontists (9,10). The umbrella and systematic reviews summarizing these periapical AI studies indicate that pooled sensitivities typically range from the mid-80s to low-90s, with high specificities and AUC values (10,19). The current study corroborates these findings in a context where case mix, imaging protocols, and resource constraints may differ from those in which many AI models were originally trained and validated, thereby supporting the potential generalizability of AI-assisted periapical diagnosis beyond highly controlled environments (9–11,19,20).

An important practical implication of the present results is the potential of AI to mitigate diagnostic fatigue and interobserver variability in busy clinical settings. Substantial kappa coefficients of 0.80–0.82 between the two calibrated dentists in this study indicate that human assessment was already reasonably consistent, yet AI still achieved higher sensitivity and specificity than the consensus conventional evaluation. This suggests that AI does not merely standardize existing human performance but can add incremental diagnostic value, particularly in the detection of subtle lesions that may be overlooked during rapid visual inspection. Systematic reviews have emphasized that AI integration can help harmonize diagnostic thresholds and reduce variation among practitioners with differing levels of experience, which is particularly relevant in teaching hospitals and community practices where junior clinicians often perform initial assessments (2,3,7,11,15). The comparative advantage of AI over junior dentists and the capacity of AI systems to maintain stable performance at scale further underscore its potential role in levelling diagnostic quality across diverse practice settings (11,13,15,20).

At the same time, the findings need to be interpreted with appropriate caution regarding the broader implementation of AI in dental diagnostics. Although the improved sensitivity and NPV observed in this study are desirable from a preventive standpoint, they may be accompanied by an increased risk of false positives if decision thresholds are not carefully calibrated, potentially leading to overtreatment or unnecessary monitoring (14–16). Furthermore, the AI model used in this investigation was trained on data that may not fully capture the diversity of anatomical variations, radiographic artifacts, and disease presentations present across all populations. Reviews have repeatedly highlighted concerns about dataset bias, limited external validation, and lack of transparency in model development and training processes in many AI studies, raising the possibility that performance may degrade when deployed in contexts that differ from the original training environment (13–16,20–22). Continuous monitoring of AI performance, periodic retraining with local data, and ongoing validation against robust reference standards are therefore essential to sustain diagnostic accuracy over time and across different clinical settings (14,15,19–22).

The study also contributes to emerging discussions about the appropriate role of AI as a complement rather than a replacement for professional judgment in dentistry. Conceptual and narrative analyses of AI applications in dental diagnostics consistently argue that clinicians should remain responsible for final diagnostic and treatment decisions, using AI outputs as a second reader or triage tool rather than as an autonomous decision maker (13,16–18,20). In this framework, AI can highlight suspicious regions, quantify diagnostic uncertainty, and nudge clinicians to reconsider borderline findings, while the clinician integrates patient history, clinical examination, and radiographic evidence to reach a holistic diagnosis. Such a collaborative model may be particularly valuable in training environments, where AI can provide instantaneous feedback to students and junior dentists, accelerating the development of pattern recognition and clinical reasoning skills (11,13,17,18,21). The high accuracy of AI observed in this study supports its potential to serve in this adjunctive capacity, but also underscores the need for appropriate training so that clinicians understand both the strengths and limitations of AI tools.

Several limitations should be acknowledged. The study was conducted in a single institution with a specific digital radiography system and a particular AI software implementation, which may limit direct generalizability to other hardware–software configurations. Although consecutive sampling and strict quality control procedures were used, selection bias cannot be entirely excluded, and the spectrum of disease severity may differ in other referral environments. The unit of analysis at the participant level, based on the presence of at least one lesion, provides clinically relevant information but does not capture tooth-level nuances such as lesion depth or extent, which are important for treatment planning. Moreover, while the reference standard was based on expert consensus using validated indices, histological confirmation was not feasible in this clinical setting and remains rare in dental diagnostic accuracy studies (2,3,9,14–16,19–22). Future research should aim to address these limitations through multicenter designs, inclusion of broader imaging modalities such as panoramic and cone-beam computed tomography, tooth-level analyses, and extended follow-up to assess the impact of AI-assisted diagnosis on treatment decisions and patient outcomes.

Despite these constraints, the strengths of this study include its prospective design, use of a clearly defined reference standard, examiner blinding, standardized radiographic acquisition and calibration procedures, appropriate statistical methods for paired diagnostic data, and comprehensive reporting of accuracy indices with confidence intervals. The consistent superiority of the AI tool over conventional examination across multiple metrics and both lesion types provides compelling evidence that AI can function as an effective adjunctive diagnostic modality in routine dental practice. Integrating such tools into clinical workflows in resource-constrained settings may help reduce missed pathology, support earlier intervention, and enhance the overall quality and consistency of dental care delivery (1–3,6–11,14–22).

CONCLUSION

In a cohort of adult dental patients undergoing routine intraoral radiography in a tertiary care setting, an artificial intelligence–based radiographic interpretation tool demonstrated significantly higher sensitivity, specificity, predictive values, and overall accuracy than conventional clinical and radiographic examination for the detection of both dental caries and periapical lesions, with markedly superior AUCs on ROC analysis; these findings support the use of AI-assisted radiographic analysis as a robust adjunct to clinician judgment, particularly in high-volume or resource-limited environments, while underscoring the need for ongoing validation, careful integration into clinical workflows, and sustained emphasis on human oversight.

DECLARATIONS

Ethical Approval
The study was approved by ethical review board of respective tertiary care hospital, Lahore, Pakistan
Informed Consent
Written informed consent was obtained from all participants included in the study.
Conflict of Interest
The authors declare no conflict of interest.
Funding
This research received no external funding.
Authors' Contributions
Concept: SA, RM; Design: SA, IR; Data Collection: AJH, RA, STS, NT; Analysis: SA, RM, IR; Drafting: SA, RM, IR, AJH, RA, STS, NT, WUK
Data Availability
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
Acknowledgments
Not applicable.
Study Registration
Not applicable.

REFERENCES

1. Ibraheem WI, Jain S, Ayoub MN, Namazi MA, Alfaqih AI, Aggarwal A, et al. Assessment of the diagnostic accuracy of artificial intelligence software in identifying common periodontal and restorative dental conditions (marginal bone loss, periapical lesion, crown, restoration, dental caries) in intraoral periapical radiographs. 2025;15(11):1432.

2. Luke AM, Rezallah NNFJH, medicine f. Accuracy of artificial intelligence in caries detection: a systematic review and meta-analysis. 2025;21(1):24.

3. Forouzeshfar P, Safaei AA, Ghaderi F, Hashemi Kamangar S, Kaviani H, Haghi SJMT, et al. Dental caries diagnosis using neural networks and deep learning: a systematic review. 2024;83(10):30423–66.

4. Szabó V, Szabo BT, Orhan K, Veres DS, Manulis D, Ezhov M, et al. Validation of artificial intelligence application for dental caries diagnosis on intraoral bitewing and periapical radiographs. 2024;147:105105.

5. Musri N, Christie B, Ichwan SJ, Cahyanto AJIsid. Deep learning convolutional neural network algorithms for the early detection and diagnosis of dental caries on periapical radiographs: a systematic review. 2021;51(3):237.

6. Liu Y, Xia K, Cen Y, Ying S, Zhao ZJOR. Artificial intelligence for caries detection: a novel diagnostic tool using deep learning algorithms. 2024;40(3):375–84.

7. Albano D, Galiano V, Basile M, Di Luca F, Gitto S, Messina C, et al. Artificial intelligence for radiographic imaging detection of caries lesions: a systematic review. 2024;24(1):274.

8. Lee J-H, Kim D-H, Jeong S-N, Choi S-H. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. J Dent. 2018;77:106–11.

9. Issa J, Jaber M, Rifai I, Mozdziak P, Kempisty B, Dyszkiewicz-Konwiń´ska M. Diagnostic test accuracy of artificial intelligence in detecting periapical periodontitis on two-dimensional radiographs: a retrospective study and literature review. 2023;59(4):768.

10. Arzani S, Karimi A, Iranmanesh P, Yazdi M, Sabeti MA, Nekoofar MH, et al. Examining the diagnostic accuracy of artificial intelligence for detecting dental caries across a range of imaging modalities: an umbrella review with meta-analysis. 2025;20(8):e0329986.

11. Güneç HG, Ürkmez EŞ, Danaci A, Dilmaç E, Onay HH, Aydin KC, et al. Comparison of artificial intelligence vs junior dentists' diagnostic performance based on caries and periapical infection detection on panoramic images. 2023;13(11):7494.

12. Zhang J-W, Fan J, Zhao F-B, Ma B, Shen X-Q, Geng Y-M, et al. Diagnostic accuracy of artificial intelligence-assisted caries detection: a clinical evaluation. BMC Oral Health. 2024;24(1):1095.

13. Musleh D, Almossaeed H, Balhareth F, Alqahtani G, Alobaidan N, Altalag J, et al. Advancing dental diagnostics: a review of artificial intelligence applications and challenges in dentistry. 2024;8(6):66.

14. Prados-Privado M, García Villalón J, Martínez-Martínez CH, Ivorra C, Prados-Frutos JC. Dental caries diagnosis and detection using neural networks: a systematic review. J Clin Med. 2020;9(11):3579.

15. Khanagar SB, Alfouzan K, Awawdeh M, Alkadi L, Albalawi F, Alfadley A. Application and performance of artificial intelligence technology in detection, diagnosis and prediction of dental caries (DC)—a systematic review. Diagnostics. 2022;12(5):1083.

16. Anil S, Sudeep K, Saratchandran S, Sweety VK. Revolutionizing dental caries diagnosis through artificial intelligence. In: Dental Caries Perspectives—A Collection of Thoughtful Essays. IntechOpen; 2023.

17. Anil S, Porwal P, Porwal A. Transforming dental caries diagnosis through artificial intelligence-based techniques. 2023;15(7).

18. Chen IDS, Yang C-M, Chen M-J, Chen M-C, Weng R-M, Yeh C-H, et al. Deep learning-based recognition of periodontitis and dental caries in dental X-ray images. 2023;10(8):911.

19. Sadr S, Mohammad-Rahimi H, Motamedian SR, Zahedrozegar S, Motie P, Vinayahalingam S, et al. Deep learning for detection of periapical radiolucent lesions: a systematic review and meta-analysis of diagnostic test accuracy. 2023;49(3):248–61.e3.

20. Alharbi SS, Alhasson HF. Exploring the applications of artificial intelligence in dental image detection: a systematic review. 2024;14(21):2442.

21. Zheng L, Wang H, Mei L, Chen Q, Zhang Y, Zhang H. Artificial intelligence in digital cariology: a new tool for the diagnosis of deep caries and pulpitis using convolutional neural networks. Ann Transl Med. 2021;9(9):763.

22. Noor Uddin A, Ali SA, Lal A, Adnan N, Ahmed SMF, Umer F. Applications of AI-based deep learning models for detecting dental caries on intraoral images—a systematic review. 2025;26(1):71–2.