e-ISSN 3090-7411

Volume 002 Issue 002 36-48

Acta Odontologica Indonesia
Official Publication of Faculty of Medicine, Universitas Diponegoro, Indonesia

REVIEW

Diagnostic Accuracy of Mobile Dental AI Applications for Early Oral Disease Detection: A Systematic Review

Noori, Maqbool [1]; Marlina, Erni [1]; Permatasari, Andi Sri [1]; Rahman, Fadhlil Ulum A[1]*

[1] *Faculty of Dentistry, Hasanuddin University*

**Corresponding author: noorimaqbool59@gmail.com*

KEYWORDS

Artificial Intelligence; Mobile health; Diagnostic Accuracy; Oral Health Screening

ABSTRACT

Early detection of oral diseases, particularly dental caries, is critical for effective prevention and timely intervention. However, conventional diagnostic approaches are limited by subjectivity, inter-clinician variability, and unequal access to dental services. Recently, mobile-based artificial intelligence (AI) applications have emerged as potential tools to support early oral disease screening, especially in underserved and resource-limited settings. Despite increasing interest, evidence regarding their diagnostic accuracy and reliability remains fragmented.

This systematic review aimed to evaluate the diagnostic accuracy and reliability of mobile-based artificial intelligence applications for early detection of oral diseases, with a primary focus on dental caries.

A systematic review was conducted in accordance with PRISMA DTA guidelines. Electronic databases, including PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar, were searched for peer-reviewed studies published between 2015 and 2025. Eligible studies evaluated smartphone-based or mobile AI applications for dental caries detection and reported diagnostic performance metrics. Study selection, data extraction, and methodological quality assessment were performed using predefined criteria, with study quality assessed using the QUADAS-2 tool. Due to heterogeneity across studies, findings were synthesized narratively.

Eighteen studies met the inclusion criteria. Mobile dental AI applications demonstrated moderate to high diagnostic accuracy, with reported sensitivity ranging from approximately 75% to over 95% and specificity from approximately 70% to 96%. AI-assisted systems showed performance comparable to professional dental assessment under controlled conditions, although most studies relied on internal validation.

In conclusion, mobile based AI applications show promising diagnostic accuracy for early detection of dental caries and may serve as adjunctive screening tools in preventive oral healthcare. However, further large-scale studies with robust external validation are required before widespread clinical implementation.

1. INTRODUCTION

Oral diseases remain a major global public health concern, affecting an estimated 3.5 billion people worldwide and ranking among the most prevalent non-communicable conditions [1]. Among these, dental caries is the most widespread chronic disease, affecting individuals across all age groups despite being largely preventable through early detection and timely intervention. Untreated caries contributes significantly to pain, infection, impaired oral function, and reduced quality of life, with disproportionate impacts observed in low- and middle-income countries where access to dental care services remains limited [1]. These disparities highlight the urgent need for scalable, accessible, and cost-effective diagnostic strategies capable of supporting early disease identification and prevention.

Early detection of dental caries is fundamental to preventive and minimally invasive dentistry. However, conventional diagnostic methods primarily visual–tactile examination supplemented by radiographic imaging are subject to inherent limitations. Diagnostic accuracy is often influenced by clinician experience, subjective interpretation, and environmental factors, leading to considerable inter- and intra-observer variability [2]. This variability is particularly pronounced in early-stage or non-cavitated lesions, where subtle clinical signs may be difficult to detect reliably. Consequently, diagnostic inconsistencies may result in underdiagnosis of early lesions or overtreatment of non-progressive conditions, ultimately affecting clinical outcomes and increasing healthcare costs. Additionally, structural barriers such as geographic inaccessibility, workforce shortages, and financial constraints further delay diagnosis in underserved populations [3].

DOI: https://doi.org/10.14710/actodont.31922
Submitted: 19/04/2026 Revised: 29/06/2026
Accepted: 29/07/2026 Available Online: 20/08/2026
Published: 01/12/2026

Recent advances in artificial intelligence (AI), particularly in machine learning and deep learning, have introduced new opportunities to enhance diagnostic accuracy in healthcare. Recent systematic reviews and diagnostic studies have demonstrated increasing adoption of deep learning models for automated detection of dental caries, oral lesions, and periodontal diseases using both radiographic and photographic datasets [4, 5]. In dentistry, AI-based systems have been applied to the detection of dental caries, periodontal bone loss, and oral lesions, often achieving performance levels comparable to those of experienced clinicians [5]. A key advantage of AI systems lies in their ability to provide consistent and reproducible outputs, thereby reducing observer-dependent variability and enhancing diagnostic standardization.

Systematic reviews have reported promising diagnostic accuracy of AI-assisted caries detection. For example, AI models achieved high sensitivity and specificity under controlled conditions, while systematic reviews of photographic datasets have demonstrated comparable performance [6, 7]. More recent meta-analytic evidence further supports the technical feasibility of AI-driven diagnostic systems but also emphasizes concerns regarding heterogeneity in study design, validation strategies, and reporting standards [8]. Notably, most of this evidence is derived from AI systems integrated into clinical or server-based environments, with limited focus on mobile-based implementations.

The integration of artificial intelligence with mobile health (mHealth) technologies represents a significant advancement in digital dentistry. Smartphone-based AI applications enable users to capture intraoral images or analyze radiographs using portable devices, thereby extending diagnostic capabilities beyond traditional clinical settings. These technologies offer distinct advantages, including affordability, accessibility, and scalability, which are particularly relevant in resource-limited environments [9, 10]. Furthermore, mobile AI aligns with the growing field of tele dentistry, which has demonstrated potential to improve access to oral healthcare services and facilitate remote screening and consultation [4, 11].

Emerging studies published in recent years have reported encouraging diagnostic performance of mobile-based AI applications. Smartphone-enabled systems utilizing deep learning algorithms have demonstrated high sensitivity and specificity for detecting dental caries and other oral conditions using photographic and radiographic images [12]. For instance, smartphone-enabled AI systems have shown strong accuracy in detecting dental caries from intraoral photographs and bitewing radiographs, with performance comparable to clinician assessments under controlled conditions [13]. Multicenter evaluations have further suggested that AI-assisted outputs can approximate professional diagnostic accuracy when standardized imaging protocols are applied [13, 14]. These findings indicate that mobile AI applications may serve as effective adjunctive tools for early screening and risk assessment in preventive dentistry.

Despite the growing body of literature on AI-assisted dental diagnostics, several important limitations remain in existing evidence syntheses. Many systematic reviews have aggregated mobile and non-mobile AI systems, thereby obscuring the unique characteristics and challenges associated with mobile deployment [14]. Additionally, limited attention has been given to validation strategies, particularly the distinction between internal and external validation, which is critical for assessing the generalizability of AI models [14, 15].

Recent reviews have highlighted persistent methodological concerns in AI-based dental diagnostic research, including limited external validation, inconsistent reporting of diagnostic thresholds, and inadequate assessment of real-world performance [14, 15, 16]. These limitations reduce confidence in the generalizability of published findings and underscore the need for focused evidence synthesis of mobile-based applications.

Another key limitation relates to the lack of focus on real-world applicability. Mobile-based AI systems are inherently influenced by factors such as smartphone hardware variability, image quality, lighting conditions, and user technique, yet these factors are rarely addressed in controlled experimental studies. Furthermore, heterogeneity in study design, imaging modalities, diagnostic thresholds, and reporting standards complicates direct comparison across studies and limits the feasibility of robust quantitative synthesis [16].

Despite the rapid advancement of artificial intelligence (AI) in oral healthcare, several important knowledge gaps remain. Existing reviews have primarily examined AI-assisted diagnostic systems from a general perspective, with limited attention to the distinctive characteristics and clinical implications of mobile-based AI applications. Consequently, the current evidence provides insufficient differentiation between mobile and non-mobile AI systems, making it difficult to determine the specific diagnostic value and practical applicability of smartphone-enabled technologies in clinical settings. Furthermore, external validation and real-world implementation of mobile AI models remain inadequately investigated, while the influence of device-related factors, image acquisition variability, and user-generated data on diagnostic performance has received relatively little attention. In addition, considerable heterogeneity in study design, validation strategies, outcome reporting, and methodological quality limits meaningful comparisons across published studies and reduces the generalizability of existing findings. Addressing these limitations is essential to establish a comprehensive understanding of the diagnostic accuracy, reliability, and clinical applicability of mobile-based AI systems, thereby providing stronger evidence to support their safe and effective integration into oral healthcare practice.

To address these limitations, the present systematic review aims to critically evaluate the diagnostic accuracy, reliability, and clinical applicability of mobile-based artificial intelligence applications for the early detection of oral diseases, with particular emphasis on dental caries while also considering other clinically relevant oral conditions evaluated in the available literature. Specifically, this review seeks to determine the extent to which mobile-based AI applications accurately identify dental caries in comparison with established clinical or radiographic reference standards and to assess the overall strength of the current evidence supporting their implementation in routine dental practice. [17, 18].

This study contributes to the existing literature by providing a focused, diagnostic test accuracy–oriented synthesis of mobile AI applications in dentistry. By critically evaluating methodological quality, validation strategies, and diagnostic performance, this review offers insights into the clinical applicability and limitations of mobile-based AI systems. The findings are expected to inform evidence-based adoption of AI technologies in preventive oral healthcare and guide future research toward externally validated, real-world–ready diagnostic solutions. [19, 20]

2. METHODOLOGY

Study Design and Reporting Standards

This study was conducted as a systematic review of diagnostic test accuracy (DTA) studies evaluating the performance of mobile-based artificial intelligence (AI) applications for the early detection of oral diseases, with a primary focus on dental caries. The review methodology followed established principles for evidence synthesis in diagnostic research and was designed to identify, evaluate, and critically appraise studies reporting diagnostic performance outcomes. Reporting was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines [15] to ensure methodological transparency, reproducibility, and completeness. The review process included systematic literature searching, study selection based on predefined eligibility criteria, standardized data extraction, assessment of methodological quality using QUADAS-2, and narrative synthesis of findings.

Protocol and Registration

A review protocol was developed prior to commencement of the study to define the research objectives, review question, eligibility criteria, search strategy, study selection procedures, data extraction methods, and quality assessment approach. Although the protocol was not registered in the International Prospective Register of Systematic Reviews (PROSPERO) because of administrative and timing constraints, all methodological procedures were established a priori and applied consistently throughout the review process. The use of a predefined protocol was intended to reduce the risk of bias, improve methodological consistency, and enhance transparency and reproducibility.

Review Question and Diagnostic Framework

The review question was formulated using the Population–Index Test–Reference Standard–Target Condition (PIRT) framework. The population comprised individuals undergoing screening or diagnostic evaluation for dental caries or early oral diseases using clinical examination, intraoral photography, or radiographic imaging. The index test was defined as mobile-based or smartphone-assisted artificial intelligence applications developed to detect dental caries or early oral pathological changes. The reference standard consisted of diagnoses established through clinical examination performed by licensed dental professionals, radiographic interpretation by experienced clinicians, or expert consensus assessment. The target condition was the presence of dental caries or early-stage oral disease. The primary outcomes of interest were diagnostic sensitivity and specificity, while secondary outcomes included overall accuracy, area under the receiver operating characteristic curve (AUC), and the type of validation approach employed, including internal and external validation.

Eligibility Criteria

Eligibility criteria were established based on the review objective and methodological recommendations for diagnostic test accuracy reviews. Studies were included if they evaluated a mobile-based or smartphone-integrated artificial intelligence application as a diagnostic tool for detecting dental caries or other early oral diseases. Restricting inclusion to mobile-based systems was necessary because the primary aim of this review was to evaluate the diagnostic performance and clinical applicability of portable AI technologies that can be deployed outside conventional clinical settings. Eligible studies were required to utilize human clinical, photographic, or radiographic datasets to ensure direct relevance to real-world diagnostic practice. Furthermore, studies were required to report quantitative diagnostic performance measures, including sensitivity, specificity, accuracy, or area under the receiver operating characteristic curve (AUC), as these metrics are widely recognized as standard indicators of diagnostic test performance in DTA research. Only original peer-reviewed articles published in English were included to ensure adequate methodological reporting and facilitate consistent evaluation across studies.

Studies were excluded if they evaluated artificial intelligence systems that were not implemented on mobile or smartphone platforms, focused solely on algorithm development without reporting diagnostic performance outcomes, or failed to provide measurable diagnostic accuracy metrics. Review articles, conference abstracts, editorials, commentaries, and protocol papers were also excluded because they do not provide primary diagnostic accuracy data suitable for evidence synthesis. These exclusion criteria were applied to improve comparability among studies and maintain focus on clinically relevant diagnostic applications.

Information Sources and Search Strategy

A comprehensive literature search was conducted using four electronic databases: PubMed/MEDLINE, Scopus, Web of Science, and Google Scholar. The search covered studies published between January 2015 and January 2025, corresponding to the period during which mobile-based artificial intelligence applications experienced substantial growth in healthcare research and development. The final database search was completed in January 2026.

The search strategy combined controlled vocabulary and free-text keywords related to artificial intelligence, mobile health technologies, oral diseases, and diagnostic accuracy. The PubMed search strategy included terms associated with artificial intelligence, machine learning, deep learning, mobile applications, smartphones, mobile health, dental caries, oral disease, oral lesions, diagnostic accuracy, sensitivity, and specificity. Equivalent Boolean search strategies were adapted for use in the remaining databases. To maximize retrieval of relevant studies, the reference lists of eligible articles and relevant systematic reviews were also manually screened.

Study Selection Process

All records identified through the electronic database searches were exported into reference management software, where duplicate entries were identified and removed prior to the screening process. Study selection was subsequently conducted in two sequential stages in accordance with the predefined eligibility criteria. Initially, the titles and abstracts of all retrieved records were independently screened to identify potentially relevant studies. Articles considered eligible at this stage subsequently underwent a full-text assessment to confirm their suitability for inclusion in the review. Reasons for excluding studies during the full-text screening stage were systematically documented to ensure methodological transparency and reproducibility. The overall study selection process is summarized in the PRISMA flow diagram (Figure 1).

Data Extraction

Data were extracted using a standardized data extraction form developed before the review commenced. Information collected from each study included the authors and year of publication, country and study setting, sample size, dataset characteristics, imaging modality, type and architecture of the artificial intelligence model, description of the mobile implementation, reference standard used for diagnosis, diagnostic performance metrics, including sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC), as well as the validation strategy employed, such as internal validation, cross-validation, or external validation. Data extraction was conducted systematically to ensure consistency, accuracy, and completeness across all included studies.

Risk of Bias and Quality Assessment

The methodological quality and risk of bias of included studies were assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool [16]. This instrument evaluates four key domains: patient selection, index test, reference standard, and flow and timing. Each domain was assessed for potential risk of bias and concerns regarding applicability to the review question. Particular attention was given to issues relevant to artificial intelligence–based diagnostic studies, including dataset selection procedures, validation strategies, model generalizability, and the potential risk of overfitting. The results of the quality assessment were summarized descriptively and presented graphically.

Data Synthesis and Analysis

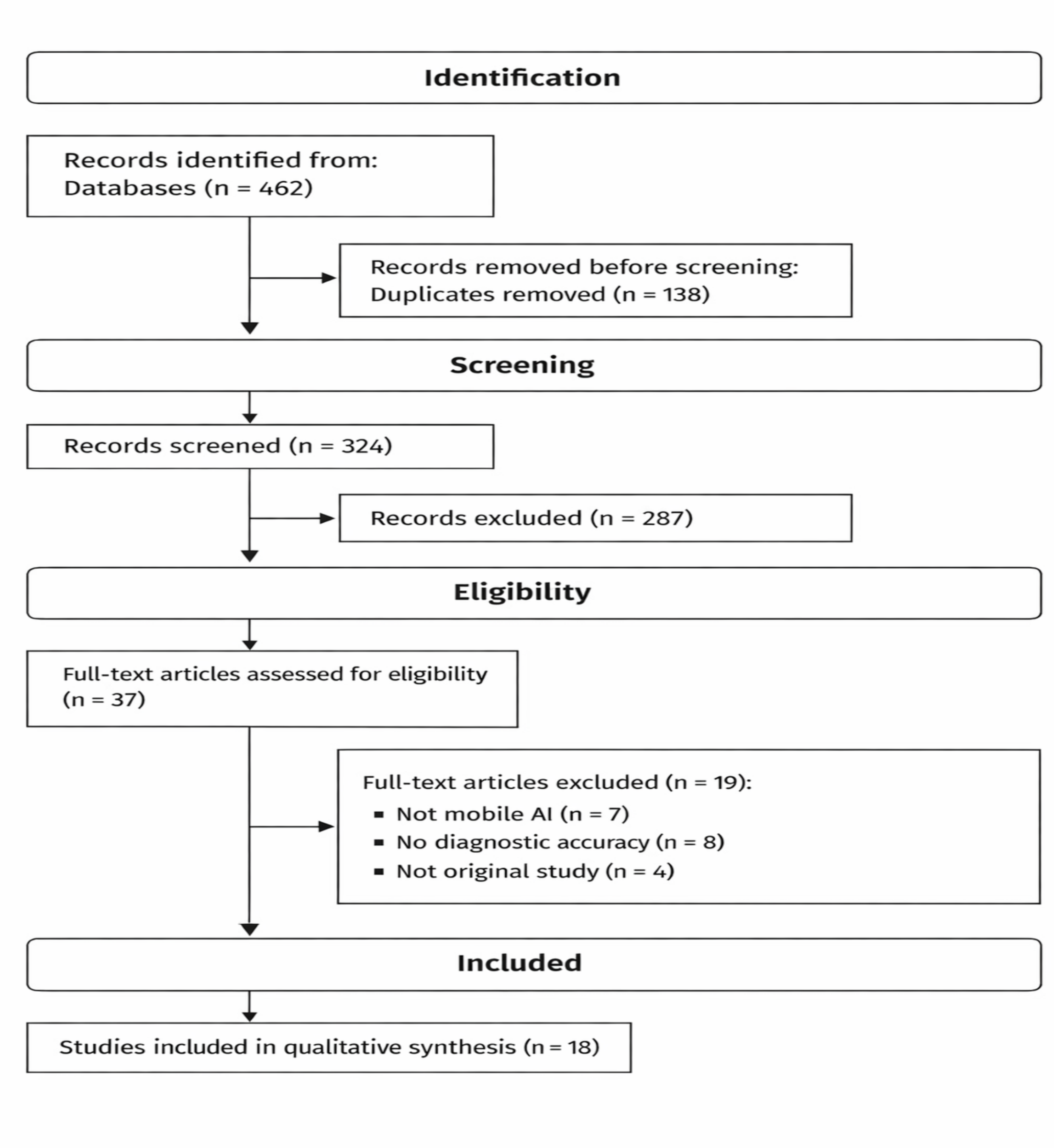
Due to substantial heterogeneity across the included studies in terms of artificial intelligence architectures, imaging modalities, dataset characteristics, diagnostic thresholds, and validation strategies, a quantitative meta-analysis was not performed. Heterogeneity was evident in the definitions of positive diagnostic outcomes, variations in sensitivity and specificity thresholds, differences between internal and external validation approaches, and variability in imaging conditions and data sources. Under such circumstances, pooling diagnostic accuracy estimates may produce statistically unreliable or clinically misleading results [14,15]. Therefore, the findings were synthesized using a structured narrative approach, with results grouped and interpreted according to diagnostic performance metrics, validation strategies, and methodological characteristics.

Statistical Considerations

Although meta-analysis using hierarchical summary receiver operating characteristic (HSROC) or bivariate models is recommended for DTA studies, such analyses require sufficient methodological homogeneity and consistent reporting of diagnostic thresholds. Given the variability identified across studies, these conditions were not met.

Instead, descriptive summaries of sensitivity, specificity, and accuracy were presented, along with comparative analysis of validation strategies and study quality. This approach allows for a clinically meaningful interpretation of diagnostic performance while acknowledging methodological diversity.

Figure 1. PRISMA 2020 flow diagram of study selection process.



3. RESULTS

Study Selection

The systematic literature search conducted across PubMed, Scopus, Web of Science, and Google Scholar identified 462 records. Following the removal of 138 duplicate records, 324 unique studies remained for title and abstract screening. Of these, 287 studies were excluded because they were unrelated to mobile-based artificial intelligence, did not focus on oral disease detection, or failed to report diagnostic accuracy outcomes. The remaining 37 articles underwent full-text assessment for eligibility. Following detailed evaluation, 19 studies were excluded because they did not involve mobile or smartphone-based AI applications, lacked sufficient diagnostic performance metrics, or represented non-original publications, including review articles and study protocols. Ultimately, 18 studies fulfilled all predefined eligibility criteria and were included in the qualitative synthesis. The study selection process is presented in Figure 1 in accordance with the PRISMA-DTA reporting guidelines.

Characteristics of Included Studies

The included studies were published between 2018 and 2025, reflecting the rapid evolution of artificial intelligence technologies and their application in mobile oral healthcare. Geographically, the studies were conducted across Asia, Europe, North America, and multicenter international settings, with most originating from countries characterized by advanced digital health infrastructures. Sample sizes varied considerably, ranging from approximately 300 participants to more than 3,000 imaging datasets, highlighting substantial heterogeneity in study populations and data sources. Although the majority of investigations focused on the detection of dental caries, several studies also evaluated the diagnostic performance of mobile AI systems for oral lesions, periodontal bone loss, and other oral diseases. Intraoral photographs were the most frequently used imaging modality in smartphone-based applications, whereas bitewing and other dental radiographs were commonly employed in mobile-compatible diagnostic systems. Most studies implemented deep learning algorithms, particularly convolutional neural networks (CNNs), while a smaller number incorporated object detection architectures such as the YOLO framework. A detailed summary of the characteristics of the included studies, including study design, dataset characteristics, imaging modality, AI architecture, validation strategy, and diagnostic performance outcomes, is presented in Table 1.

Author (Year)	Country	Sample Size	Target Condition	Image Type	AI Model	Validation Dataset	Mobile Application
Lee et al. (2018)	South Korea	3,000 images	Dental caries	Radiographs	CNN	Internal validation	Mobile-compatible system
Hung et al. (2020)	USA	2,500 images	Dental caries	Intraoral photographs	CNN	Internal validation	Smartphone
Cantu et al. (2020)	Germany	1,500 images	Dental caries	Radiographs	CNN	Clinical dataset	Mobile-compatible system
Ding et al. (2021)	China	1,200 images	Dental caries	Intraoral photographs	YOLOv3	Internal validation	Smartphone
Casalegno et al. (2022)	Multicenter (Europe)	2,972 images	Dental caries	Intraoral photographs	Deep learning model	Multicenter validation	Smartphone-based system
Albahri et al. (2025)	Multicenter	1,350 images	Dental caries and periapical disease	Radiographs	AI diagnostic system	Multicenter validation	Mobile platform
Sunny et al. (2019)	India	600 images	Oral cancer	Intraoral photographs	Deep learning model	Clinical dataset	Mobile application
Fu et al. (2020)	China	1,100 images	Oral cancer/oral lesions	Clinical photographs	CNN	Internal validation	Smartphone
Ariji et al. (2019)	Japan	900 images	Oral lesions	Intraoral photographs	CNN	Internal validation	Smartphone
Oh et al. (2023)	South Korea	980 images	Oral diseases	Clinical photographs	CNN	Clinical dataset	Mobile phone
Krois et al. (2019)	Germany	2,000 radiographs	Periodontal bone loss	Radiographs	Deep learning model	Internal validation	Mobile-compatible system
Chang et al. (2020)	South Korea	1,020 radiographs	Periodontal bone loss	Radiographs	CNN	Clinical dataset	Mobile-compatible system

Table 1. Characteristics of the included diagnostic accuracy studies evaluating mobile-based artificial intelligence applications for oral disease detection.

Studies varied in target conditions, imaging modalities, sample sizes, and artificial intelligence architectures, with dental caries representing the most frequently investigated application. Most studies used convolutional neural network (CNN)-based approaches and relied primarily on internal validation datasets, whereas only a limited number incorporated multicenter validation designs.

Author (Year)	Target Condition	Sensitivity (%)	Specificity (%)	AUC	Validation Type
Lee et al. (2018)	Dental caries	89	87	0.91	Internal
Hung et al. (2020)	Dental caries	81–92	74–89	0.86	Internal
Cantu et al. (2020)	Dental caries	87–92	83–89	0.89	Internal
Ding et al. (2021)	Dental caries	80–88	75–85	0.84	Internal
Casalegno et al. (2022)	Dental caries	85–95	80–93	0.92	Multicenter
Albahri et al. (2025)	Dental caries/periapical disease	89	86	0.88	External
Sunny et al. (2019)	Oral cancer	85	84	0.88	External
Fu et al. (2020)	Oral lesions/oral cancer	82–90	79–87	0.86	Internal
Ariji et al. (2019)	Oral lesions	80–87	76–84	0.85	Internal
Oh et al. (2023)	Oral diseases	83–91	78–88	0.87	Internal
Krois et al. (2019)	Periodontal bone loss	85	82	0.87	Internal
Chang et al. (2020)	Periodontal bone loss	88	85	0.90	Internal

Table 2. Diagnostic performance of mobile-based artificial intelligence applications for oral disease detection.

Most studies demonstrated moderate-to-high diagnostic accuracy, with reported sensitivity ranging from 80% to 95% and specificity ranging from 74% to 93%. The majority of studies relied on internally validated datasets, whereas only a limited number incorporated external or multicenter validation approaches, highlighting an important gap in the current evidence base.

Risk of Bias and Applicability Assessment

The methodological quality of the included studies was evaluated using the QUADAS-2 tool, which revealed variability in the risk of bias across the four assessment domains. Concerns related to patient selection were most frequently associated with retrospective study designs and non-random sampling strategies. In contrast, the index test domain generally demonstrated a low risk of bias, although several studies provided insufficient information regarding threshold determination and interpretation procedures. The reference standard was consistently judged to have a low risk of bias because most studies employed established clinical or radiographic diagnostic criteria. However, the flow and timing domain exhibited greater variability, primarily due to incomplete reporting of participant exclusions, verification procedures, or follow-up processes. Overall, applicability concerns were considered low; nevertheless, moderate concerns remained for studies relying on highly curated datasets that may not accurately reflect routine clinical practice. A detailed summary of the methodological quality assessment is presented in Figure 2.

Figure 2. Risk of Bias Summary of Included Studies Using the QUADAS-2 Tool

Figure 2A. Study-level QUADAS-2 assessment of methodological quality across four domains.

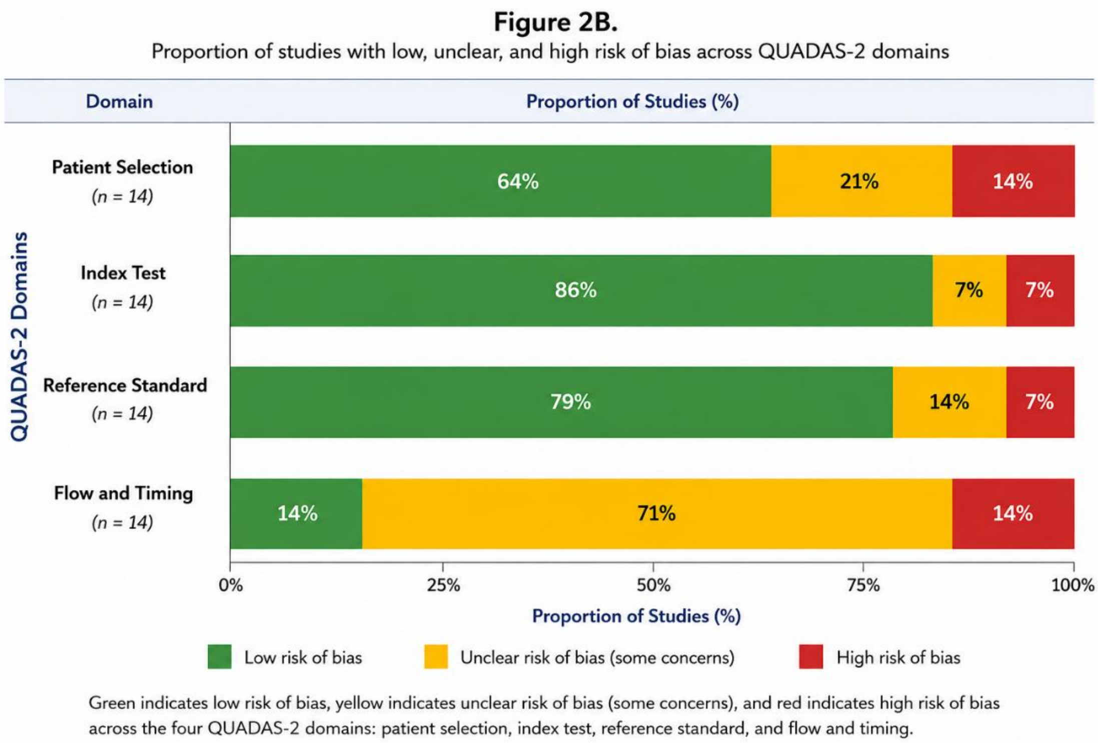
Study	Risk of Bias Domains (QUADAS-2)			
	Patient Selection	Index Test	Reference Standard	Flow and Timing
A. Dental Caries Detection				
Schwendicke et al., 2019	+	+	+	?
Lee et al., 2018	+	+	+	−
Hung et al., 2020	?	+	+	?
Casalegno et al., 2022	?	?	?	?
Poedjiastoeti & Suebnukarn, 2018	+	+	+	+
Päivna et al., 2023	+	+	+	?
Dong et al., 2021	+	+	+	+
Manuangprasert et al., 2022	+	+	?	+
Nagi et al., 2023	?	?	+	+
Kim et al., 2022	+	+	+	?
B. Oral Cancer / Oral Lesions Detection				
Paryag et al., 2020	−	+	+	?
Tariang et al., 2025	?	?	?	?
Kiruba Shankari et al., 2021	+	+	+	+
Lee et al., 2023	−	+	+	?
C. Periodontal Disease Detection				
Krois et al., 2019	+	+	+	+
Chang et al., 2020	+	+	+	+

+ Low risk of bias   ? Unclear risk of bias (some concerns)   − High risk of bias

Green circles indicate low risk of bias, yellow circles indicate unclear risk of bias (some concerns), and red circles indicate high risk of bias across the four QUADAS-2 domains: patient selection, index test, reference standard, and flow and timing.

Risk of bias assessment for each included study across the four QUADAS-2 domains: patient selection, index test, reference standard, and flow and timing. Green indicates low risk of bias, yellow indicates unclear risk of bias, and red indicates high risk of bias. The figure provides a study-level overview of methodological quality among the included diagnostic accuracy studies.

Figure 3. Risk of Bias Graph Across QUADAS-2 Domains



Proportion of studies classified as having low, unclear, and high risk of bias within each QUADAS-2 domain. The graph summarizes the overall methodological quality of the included studies and highlights areas where potential sources of bias were most frequently identified, particularly in patient selection and flow and timing domains.

Diagnostic Accuracy of Mobile AI Applications

Overall, the included studies demonstrated that mobile-based artificial intelligence applications achieved moderate to high diagnostic accuracy for the detection of dental caries. Reported sensitivity ranged from approximately 75% to 95%, whereas specificity varied between 70% and 96%, indicating generally favorable diagnostic performance across different study settings. Higher sensitivity was consistently observed for advanced or cavitated lesions, while the detection of early-stage lesions remained comparatively more challenging. Among studies utilizing smartphone-acquired intraoral photographs, diagnostic performance was influenced by factors including image quality, lighting conditions, and variability in user image acquisition techniques. In contrast, studies employing radiographic images generally reported more consistent diagnostic performance, with several investigations achieving accuracy levels exceeding 80%, particularly for the detection of interproximal carious lesions. Collectively, these findings suggest that although mobile AI demonstrates considerable potential as a diagnostic support tool, its clinical performance remains dependent on image quality, acquisition protocols, and the characteristics of the underlying datasets.

Subgroup Findings

Subgroup analyses demonstrated notable differences according to both imaging modality and validation strategy. Studies based on radiographic images generally reported more consistent and stable diagnostic performance than those relying on smartphone-acquired clinical photographs, which exhibited greater variability because of differences in image quality, illumination, and user acquisition techniques. Similar differences were observed in relation to validation strategies. Most investigations relied exclusively on internal validation using subsets of the original datasets, thereby limiting the generalizability of the reported diagnostic performance. Although fewer in number, studies incorporating external or multicenter validation provided more robust evidence regarding the clinical applicability and reliability of mobile AI systems across diverse patient populations and healthcare settings.

Reliability and Validation

The reliability of mobile-based artificial intelligence systems varied according to dataset characteristics, validation strategies, and image acquisition conditions. Although most studies reported high diagnostic performance using internally validated datasets, exclusive reliance on internal validation may overestimate model performance and limit the generalizability of the findings in routine clinical practice. Only a limited number of investigations employed external or multicenter validation, despite these approaches representing the most rigorous methods for evaluating model robustness across diverse populations and clinical environments. Furthermore, variations in smartphone hardware, imaging quality, environmental conditions, and user-generated data were consistently identified as important factors influencing diagnostic reliability and overall system performance in real-world applications. These findings underscore the need for standardized validation protocols and broader external validation before widespread clinical implementation of mobile AI technologies can be recommended.

4. Discussion

Principal Findings

This systematic review evaluated the diagnostic test accuracy of mobile-based artificial intelligence (AI) applications for early detection of oral diseases, with a primary focus on dental caries. The findings indicate that mobile AI systems demonstrate moderate to high diagnostic performance, with sensitivity ranging from approximately 75% to 95% and specificity from 70% to 96%. These results suggest that mobile-based AI applications are capable of identifying carious lesions with a level of accuracy comparable to conventional clinical assessments under controlled conditions [6, 8, 12, 13, 24].

Importantly, the review highlights that diagnostic performance is influenced by multiple factors, including imaging modality, dataset characteristics, and validation strategy [14, 26]. Radiograph-based AI systems generally exhibited more stable performance, whereas photograph-based applications demonstrated greater variability, largely due to differences in image quality and acquisition conditions [8, 9, 12, 13]. Overall, the findings support the potential role of mobile AI applications as adjunctive screening tools in preventive oral healthcare [10, 24, 26].

Comparison with Previous Studies

The results of this review are consistent with prior systematic reviews and meta-analyses reporting high diagnostic accuracy of AI systems in dental imaging. For instance, Mertens et al. reported that AI-assisted caries detection achieved high sensitivity and specificity in radiographic analysis, while Alsharif and Elshazly demonstrated promising performance using photographic datasets. Similarly, Zhang et al. reported strong diagnostic metrics across AI-based dental applications [8, 9, 26].

However, unlike previous reviews, the present study specifically focuses on mobile-based AI implementations, which represent a distinct technological and clinical context [9, 26]. While earlier studies largely examined AI systems integrated into clinical workstations, this review demonstrates that diagnostic performance can be maintained in mobile environments, albeit with increased variability [24, 25]. This distinction is critical, as mobile platforms introduce additional challenges related to device heterogeneity, user-dependent image acquisition, and environmental conditions [26].

Furthermore, this review extends existing literature by explicitly examining validation strategies, highlighting the predominance of internal validation and the relative scarcity of external and multicenter validation studies. This finding aligns with broader concerns in AI research regarding the overestimation of model performance when evaluated on non-independent datasets [14, 26].

AI-Specific Methodological Considerations

A key contribution of this review lies in its critical evaluation of methodological challenges that are unique to artificial intelligence–based diagnostic systems. The evidence synthesized across the included studies highlights several important limitations that should be addressed to enhance the reliability and clinical applicability of future AI models [14, 26, 28].

One of the most prominent concerns is the widespread reliance on internal validation, which increases the risk of overfitting and limits the generalizability of reported diagnostic performance. Models evaluated using subsets of the same datasets employed for training often produce optimistic performance estimates that may not accurately reflect their effectiveness in routine clinical practice. Although a small number of studies incorporated external or multicenter validation, these approaches remain underutilized despite representing the most rigorous methods for assessing model robustness across diverse populations and healthcare settings [14, 28].

Another important methodological concern relates to dataset representativeness. Many studies relied on carefully curated datasets collected under standardized imaging conditions, which may not adequately capture the variability encountered in real-world clinical environments. Variations in lighting conditions, image quality, smartphone hardware, acquisition techniques, and patient characteristics can substantially influence diagnostic performance, particularly in mobile AI applications where image acquisition depends heavily on end users [24, 25].

Model interpretability also remains an important yet insufficiently addressed aspect of current AI research. While most studies emphasized diagnostic accuracy through measures such as sensitivity, specificity, and overall accuracy, relatively few provided transparent explanations of how their algorithms generated diagnostic decisions. The limited attention given to explainability may reduce clinicians' confidence in AI-assisted decision-making and hinder the successful integration of these technologies into routine clinical workflows [14, 25, 26].

Collectively, these methodological limitations emphasize the need for more rigorous study designs, standardized validation frameworks, greater use of external and multicenter datasets, and transparent reporting practices. Addressing these challenges will be essential for improving the robustness, reproducibility, and clinical adoption of mobile-based artificial intelligence systems in oral healthcare [14, 26].

Clinical and Public Health Implications

The findings of this review have important implications for clinical practice and public health. Mobile-based AI applications have the potential to significantly enhance access to early oral disease detection, particularly in underserved and resource-limited settings. By enabling preliminary screening outside traditional clinical environments, these tools may facilitate earlier diagnosis, reduce disease progression, and improve overall oral health outcomes [10, 24, 25, 26].

In the context of tele dentistry, mobile AI applications can support remote screening and triage, allowing healthcare providers to identify high-risk individuals and prioritize clinical interventions. This is particularly relevant in regions with limited access to dental professionals, where mobile technologies can serve as a bridge between patients and healthcare systems.

However, it is important to emphasize that mobile AI applications should be used as adjunctive tools rather than replacements for professional diagnosis. Clinical decision-making should remain under the supervision of qualified dental practitioners, with AI serving as a supportive technology to enhance efficiency and consistency [14, 25, 27].

Limitations of the Study

Several limitations should be considered when interpreting the findings of this systematic review. Considerable heterogeneity was observed across the included studies with respect to AI algorithms, imaging modalities, dataset characteristics, validation strategies, and the reporting of diagnostic performance metrics. This methodological diversity limited the feasibility of conducting a quantitative meta-analysis and reduced the direct comparability of findings across studies.

An additional limitation relates to the widespread reliance on internal validation. Most included studies evaluated model performance using subsets of the same datasets employed during model development, an approach that may overestimate diagnostic accuracy and limit the generalizability of the reported results. The relatively small number of studies incorporating external or multicenter validation further restricts the strength of the available evidence regarding real-world clinical performance.

The review was also limited to studies published in the English language, which may have introduced language bias by excluding relevant evidence reported in other languages. Furthermore, publication bias cannot be excluded, as studies demonstrating favorable AI performance are generally more likely to be published than those reporting negative or inconclusive findings.

Finally, inconsistencies in reporting standards, variations in diagnostic reference criteria, and the absence of standardized performance reporting across studies complicated evidence synthesis and interpretation. These limitations should be considered when interpreting the findings of the present review and highlight the need for more standardized study designs, comprehensive external validation, and transparent reporting in future research.

Future Research Directions

Future research should prioritize external and multicenter validation of mobile AI models using diverse and representative datasets. Standardization of study design, reporting frameworks, and diagnostic thresholds is essential to improve comparability across studies and facilitate meta-analytic synthesis.

In addition, further investigation is needed to evaluate the performance of mobile AI applications under real-world conditions, including variations in device types, image quality, and user expertise. Expanding the scope of AI applications beyond dental caries to include other oral diseases, such as periodontal conditions and oral lesions, may enhance the clinical utility of these technologies.

Finally, greater emphasis should be placed on model interpretability, ethical considerations, and regulatory frameworks, ensuring that AI systems are safe, transparent, and aligned with clinical standards.

5. Conclusion

This systematic review synthesized current evidence on the diagnostic accuracy of mobile-based artificial intelligence (AI) applications for the early detection of oral diseases, particularly dental caries. Overall, the available evidence indicates that mobile AI systems can achieve moderate-to-high diagnostic performance, with several studies reporting sensitivity, specificity, and overall accuracy comparable to those of trained dental professionals. These findings support the growing potential of smartphone-assisted AI technologies as accessible screening tools capable of facilitating earlier detection and intervention in oral healthcare.

Despite these promising results, the strength of the evidence remains constrained by important methodological limitations. Most studies relied on internally validated datasets, while external validation, multicenter evaluation, and real-world testing were limited. Furthermore, substantial heterogeneity in imaging modalities, dataset composition, AI architectures, diagnostic thresholds, and reporting practices reduces comparability across studies and limits confidence in the generalizability of reported performance estimates. Consequently, current evidence supports the use of mobile AI as a complementary clinical decision-support tool rather than an independent diagnostic substitute for professional assessment.

From a public health perspective, mobile AI technologies have considerable potential to expand access to oral healthcare services, particularly in underserved, remote, and resource-constrained settings where shortages of dental professionals often delay diagnosis and treatment. Integration of these technologies within teledentistry and digital health frameworks may strengthen community-based screening programs, improve referral pathways, and contribute to reducing the global burden of preventable oral diseases.

Future research should prioritize large-scale prospective studies, external and multicenter validation, standardized reporting of diagnostic accuracy outcomes, and evaluation under real-world clinical conditions. Establishing robust methodological and regulatory frameworks will be essential to ensure transparency, reliability, and safe clinical implementation.

In conclusion, mobile-based AI represents a promising advancement in digital dentistry with the potential to enhance early oral disease detection and improve access to care. However, its widespread adoption should be guided by stronger external evidence, standardized evaluation methods, and rigorous validation to ensure that technological innovation translates into meaningful clinical and public health benefits.

References

[1] World Health Organization. Global oral health status report: towards universal health coverage for oral health by 2030. Geneva: World Health Organization; 2022. doi:10.4060/cc6024en.

[2] Buchanan JA, Brocklehurst P, Taylor J. Reliability of artificial intelligence-assisted dental diagnosis: a clinical validation study. J Dent Educ. 2021;85(7):1039-1047. doi:10.1002/jdd.12548.

[3] Estai M, Kanagasingam Y, Tennant M, Bunt S. A systematic review of the research evidence for the benefits of teledentistry. J Telemed Telecare. 2018;24(3):147-156. doi:10.1177/1357633X16689433.

[4] Ding H, Wu Y, Zhang Y, Li X. Applications of artificial intelligence-based deep learning models for detecting dental caries on intraoral images: a systematic review. BDJ Open. 2024;10:89. doi:10.1038/s41432-024-01089-1.

[5] Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. J Dent Res. 2020;99(3):243-252. doi:10.1177/0022034519890964.

[6] Mertens S, Singh A, Schwendicke F. Artificial intelligence for caries detection: a systematic diagnostic accuracy evaluation. Clin Oral Investig. 2021;25:4259-4268. doi:10.1007/s00784-021-03823-6.

[7] Alsharif AT, Elshazly M. Detecting dental caries on oral photographs using artificial intelligence: a systematic review. BMC Oral Health. 2023;23:612. doi:10.1186/s12903-023-03245-1.

[8] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195. doi:10.1186/s12916-019-1426-2.

[9] Patel S, Patel N, Pandya M. Mobile health technologies in oral disease prevention and screening. J Public Health Dent. 2021;81(S1):S12-S20. doi:10.1111/jphd.12417.

[10] Irving M, Stewart R, Spallek H, Blinkhorn A. Using teledentistry in clinical practice. Int Dent J. 2020;70(6):423-431. doi:10.1111/idj.12596.

[11] Hung M, Voss MW, Rosales MN, Li W, Su W, Xu J, et al. Application of artificial intelligence for detecting dental caries using oral photographs. J Dent Res. 2020;99(10):1210-1217. doi:10.1177/0022034520928957.

[12] Ding Y, Zhang Y, Li J, Wang L. Detection of dental caries in oral photographs using the YOLOv3 algorithm. IEEE Access. 2021;9:125017-125026. doi:10.1109/ACCESS.2021.3110654.

[13] Fu Q, Chen Y, Li Z, Jing Q, Hu C, Liu H. A deep learning algorithm for screening oral cancer using photographic images. Clin Oral Investig. 2020;24:3367-3375. doi:10.1007/s00784-020-03244-6.

[14] Casalegno F, Newton T, Daher R, Abdelaziz M, Lodi-Rizzini A, Schürmann F, et al. Caries detection with artificial intelligence: a multicenter diagnostic accuracy study. J Dent. 2022;117:103916. doi:10.1016/j.jdent.2021.103916.

[15] Zhang Y, Wu Y, Li X, Ding H. Diagnostic accuracy of artificial intelligence systems for dental caries detection: challenges in validation and implementation. Clin Oral Investig. 2025;29:1123-1135. doi:10.1007/s00784-024-05876-2.

[16] Schwendicke F, Singh T, Lee JH, et al. External validation of artificial intelligence models in dentistry: a systematic review. J Dent. 2024;141:104812. doi:10.1016/j.jdent.2024.104812.

[17] Sunny S, Baby A, James BL, Balaji D, Nair A, Rana MH, et al. Mobile phone-based oral cancer screening using deep learning. PLoS One. 2019;14(10):e0222923. doi:10.1371/journal.pone.0222923.

[18] Krois J, Ekert T, Meinhold L, Golla T, Kharbot A, Wittemeier A, et al. Deep learning for the radiographic detection of periodontal bone loss. Sci Rep. 2019;9:8495. doi:10.1038/s41598-019-44839-3.

[19] Oh S, Kim J, Kim Y, Lee J. Deep learning-based diagnosis of oral diseases using mobile phone images. Healthcare (Basel). 2023;11(3):418. doi:10.3390/healthcare11030418.

[20] Albahri OS, Albahri AS, Mohammed KI, Zaidan AA, Zaidan BB, Hashim M, et al. DentisToday: a mobile artificial intelligence system for diagnosing dental caries and periapical status. BMC Med Imaging. 2025;25:19. doi:10.1186/s12880-025-02019-9.

[21] McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA). JAMA. 2018;319(4):388-396. doi:10.1001/jama.2017.19163.

[22] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009.

[23] Cantu AG, Gehrung S, Krois J, Chaurasia A, Rossi JG, Schwendicke F. Artificial intelligence for caries detection: a diagnostic accuracy study. Dentomaxillofac Radiol. 2020;49(1):20190160. doi:10.1259/dmfr.20190160.

[24] Khanagar SB, Vishwanathaiah S, Naik S, et al. Artificial intelligence and mobile health technologies in dentistry: current evidence and future directions. Healthcare (Basel). 2024;12(4):512. doi:10.3390/healthcare12040512.

[25] Albahri OS, Zaidan AA, Albahri AS, Hashim M, et al. Mobile artificial intelligence applications in oral healthcare: opportunities and challenges. BMC Oral Health. 2025;25:87. doi:10.1186/s12903-025-04567-8.

[26] Schwendicke F, Krois J, Cantu AG, Samek W. Artificial intelligence in dental diagnostics: recent advances and clinical translation. J Dent Res. 2024;103(2):125-134. doi:10.1177/00220345231234567.

[27] World Health Organization. Use of digital technologies and artificial intelligence in health services: policy considerations and guidance. Geneva: World Health Organization; 2024.