https://doi.org/10.55453/rjmm.2026.129.6.3

Review

Artificial Intelligence for Radiographic Assessment of Peri-Implant Marginal Bone Loss: Current Evidence, Clinical Interpretation, and Limitations

Vladimir Năstasie[1], Ioan Sîrbu[2], Andreea Mihaela Custură[3,*], Elisei Adelin Radu[3,*], Dragoș Epistatu[4], Alexandra Tuță[2], Valentin Daniel Sîrbu[2]

[1] Doctoral School, Faculty of Dentistry, "Carol Davila" University of Medicine and Pharmacy, 050474 Bucharest, Romania
[2] Discipline of Implanto-Prosthesis Therapy, Faculty of Dentistry, "Carol Davila" University of Medicine and Pharmacy, 37 Dionisie Lupu Street, District 2, 020021 Bucharest, Romania
[3] Discipline of Dental Prostheses Technology, Faculty of Dentistry, "Carol Davila" University of Medicine and Pharmacy, 37 Dionisie Lupu Street, District 2, 020021 Bucharest, Romania
[4] Department of Dental Radiology, Faculty of Dentistry, "Carol Davila" University of Medicine and Pharmacy, 020021, Bucharest, Romania

*Correspondence: A.M.C - andreea-mihaela.custura@drd.umfcd.ro; A.E.R – elisei-adelin.radu@drd.umfcd.ro; Tel: +40743907259.

Abstract: Background: Artificial intelligence (AI) is increasingly investigated for radiographic assessment of peri-implant bone conditions, although radiographic marginal bone loss does not alone establish a clinical diagnosis of peri-implantitis. This narrative review aimed to critically synthesize current evidence on AI-assisted radiographic assessment of peri-implant marginal bone levels and bone loss. Methods: A structured literature search was conducted in PubMed/MEDLINE and Scopus. Eligible original studies applying AI or machine-learning methods to human dental radiographs for peri-implant bone assessment were included in a qualitative synthesis. Results: Seventeen studies were included. Periapical radiographs were the predominant imaging modality, while panoramic radiographs and CBCT were less frequently investigated. AI applications included implant and landmark localization, marginal bone-level measurement and quantification, bone-loss severity and defect-morphology classification, and prediction using imaging-derived features. Reported performance was generally promising, although substantial heterogeneity was observed in datasets, target tasks, reference standards, validation strategies, and performance metrics. Conclusions: Current radiograph-based AI systems show potential for image analysis, quantitative assessment, and clinical decision support. However, radiographic AI performance should not be interpreted as autonomous diagnostic accuracy for peri-implantitis. External validation, longitudinal assessment, and integration with relevant clinical parameters remain priorities for future research.

Keywords: artificial intelligence; convolutional neural network; peri-implant marginal bone loss; dental radiography

INTRODUCTION

Long-term implant maintenance requires surveillance of peri-implant soft tissues and supporting bone. Radiography is an essential objective method for identifying and longitudinally measuring peri-implant bone-level changes, but it does not directly demonstrate inflammation or disease activity. Periapical radiographs are generally preferred for standardized marginal bone-level assessment; panoramic radiographs may provide broader screening information, whereas cone-beam computed tomography (CBCT) is reserved for selected indications in which three-dimensional assessment is clinically justified [1].

Peri-implant health is characterized by the absence of clinical signs of inflammation. Peri-implant mucositis presents with bleeding and/or suppuration on gentle probing without bone loss beyond initial crestal remodeling. Peri-implantitis is a plaque-associated pathological condition characterized by inflammation of the peri-implant mucosa and subsequent progressive loss of supporting bone [1]. Thus, peri-implantitis is not a radiographic diagnosis alone. Diagnosis should integrate bleeding and/or suppuration on probing, increased probing depth relative to previous examinations, and radiographic bone loss beyond initial remodeling [2]. When previous records are unavailable, the 2017 World Workshop case definition uses

Citation: Năstase V, Sîrbu I, Custură AM, Radu EA, Epistatu D, Tuță A, Sîrbu VD. *Artificial Intelligence for Radiographic Assessment of Peri-Implant Marginal Bone Loss: Current Evidence, Clinical Interpretation, and Limitations.* R. J. Mil. Med. 2026, CXXIX(6): 610-623.https://doi.org/10.55453/rjmm.2026.129.6.3

Academic Editor: Octavian Vasiliu

Received: 12 September 2026
Revised: 26 October 2026
Accepted: 30 October 2026

bleeding and/or suppuration, probing depths of at least 6 mm, and bone levels at least 3 mm apical to the most coronal intraosseous portion of the implant [1]. The EFP S3 guideline likewise places radiographic findings within a clinical pathway that includes probing findings, patient- and implant-level risk assessment, and longitudinal comparison where available [3].

Marginal bone loss must therefore be distinguished from peri-implantitis-related bone resorption. Early crestal change may reflect physiological remodeling after implant placement and loading, and a single radiograph cannot determine the biological cause of a bone-level change. Baseline or previous radiographs are essential whenever available, because progression beyond expected remodeling is more informative than an isolated measurement [1, 3]. Consequently, radiographic marginal bone loss is an imaging finding, while peri-implantitis is a clinical diagnosis that cannot be determined solely from radiographic information [1, 3]. When evaluating the effectiveness of artificial intelligence models that have only been trained on radiographic images, this distinction is especially crucial.

Radiographic interpretation can nevertheless be challenging. Accurate measurement requires consistent identification of the implant platform, implant-abutment interface, implant apex, first bone-to-implant contact, and exposed threads. Projection error, overlapping anatomy, prosthetic components, image quality and implant design can obscure landmarks. Moreover, two-dimensional radiographs cannot fully depict buccal and lingual bone morphology. These challenges motivate automated methods that can reliably recognize anatomical and implant-related landmarks and measure changes at the peri-implant bone level [4, 5].

Artificial intelligence encompasses computational methods that perform tasks commonly associated with human reasoning or perception. Machine learning is a subset of AI in which algorithms learn patterns from data; deep learning is a subset of machine learning that uses multilayer neural networks. In dento-maxillofacial radiology, these approaches have been investigated to improve consistency, reduce repetitive measurement burden, and support image-based detection and classification [6, 7]. In peri-implant radiographic assessment, convolutional neural networks (CNNs) and related deep-learning architectures have been applied to several tasks, such as implant and tissue localization, landmark detection, bone-level measurement, radiographic severity grading, and defect-pattern classification [4, 5, 8-12]. These image-analysis tasks should not be conflated with the clinical diagnosis of peri-implantitis.

Recent systematic reviews and a meta-analysis have already summarized the performance of AI models applied to peri-implant radiographic assessment [13-15]. However, the available primary studies address heterogeneous imaging tasks, including implant and landmark localization, marginal bone-level measurement, severity grading, and defect-morphology classification. As a result, performance estimates for these specific tasks cannot be consistently interpreted as peri-implantitis diagnostic accuracy. The present review focuses on the precise radiographic task carried out by each AI model and the methodological elements that determine its potential clinical relevance, offering a clinically oriented interpretation of this varied evidence instead of another quantitative synthesis.

Accordingly, this narrative review aims to critically synthesize current evidence on AI-assisted radiographic assessment of peri-implant marginal bone levels and bone loss, with particular emphasis on the specific imaging tasks performed by existing models, their methodological and validation characteristics, and their potential clinical applications and limitations. A further objective is to clarify the distinction between AI-based assessment of radiographic bone changes and the multimodal clinical diagnosis of peri-implantitis.

MATERIALS AND METHODS

Review Design

This study was conceptualized as a narrative review that was complemented by a structured literature search. The objective was to critically synthesize the most recent evidence regarding the radiographic assessment of peri-implant marginal bone levels and bone loss using artificial intelligence (AI). With special attention to the particular imaging tasks carried out by each model, such as implant and peri-implant tissue localization, landmark detection, marginal bone-level measurement, bone-loss detection and severity grading, and defect-morphology classification, the review concentrated on original studies assessing AI-based analysis of dental implant radiographs.

The significant methodological variety across the available studies, especially with regard to imaging modalities, AI architectures, target tasks, reference standards, validation techniques, and reported performance measures, led to a narrative synthesis of the evidence. As a result, neither quantitative pooling nor meta-analysis was carried out.

While consensus reports and clinical practice guidelines established accepted definitions and diagnostic criteria for peri-implant health, peri-implant mucositis, and peri-implantitis, systematic reviews, meta-analyses, and other pertinent evidence syntheses were utilized to contextualize the results from the primary studies. These sources were taken into account independently of the original research that was part of the primary evidence synthesis.

The distinction between the clinical diagnosis of peri-implantitis, which necessitates the integration of radiographic results with clinical and, if available, longitudinal data, and AI-based assessment of radiographic marginal bone alterations was given special attention.

Information Sources and Search Strategy

A structured literature search was conducted in two electronic databases, PubMed/MEDLINE and Scopus, to identify studies investigating artificial intelligence (AI)-based radiographic assessment of peri-implant marginal bone levels and bone loss. The databases were searched between 15th of July and 15th of August 2026. No restrictions regarding publication date, language, or article type were applied at the database-search stage.

The search strategy was developed around three main concepts: (1) artificial intelligence and machine-learning methods; (2) dental implants and peri-implant tissues; and (3) peri-implant marginal bone levels, bone loss, remodeling, and related bone changes. Database-specific search strategies were used to account for differences in indexing systems and search syntax.

For PubMed/MEDLINE, Medical Subject Headings (MeSH) and free-text terms searched within the title and abstract fields were combined using Boolean operators. The final PubMed/MEDLINE search strategy was:

("Artificial Intelligence"[Mesh] OR "artificial intelligence"[Title/Abstract] OR "Machine Learning"[Mesh] OR "machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "convolutional neural network*"[Title/Abstract] OR CNN[Title/Abstract] OR "object detection"[Title/Abstract]) AND ("Dental Implants"[Mesh] OR "dental implant*"[Title/Abstract] OR "peri-implant"[Title/Abstract] OR "periimplant"[Title/Abstract] OR "peri-implantitis"[Title/Abstract] OR "periimplantitis"[Title/Abstract]) AND ("marginal bone loss"[Title/Abstract] OR "marginal bone level*"[Title/Abstract] OR "peri-implant bone loss"[Title/Abstract] OR "periimplant bone loss"[Title/Abstract] OR "peri-implant bone level*"[Title/Abstract] OR "periimplant bone level*"[Title/Abstract] OR "marginal bone remodeling"[Title/Abstract] OR "marginal bone remodelling"[Title/Abstract] OR "peri-implant bone remodeling"[Title/Abstract] OR "peri-implant bone remodelling"[Title/Abstract] OR "peri-implant bone defect*"[Title/Abstract] OR "periimplant bone defect*"[Title/Abstract] OR "peri-implantitis"[Title/Abstract] OR "periimplantitis"[Title/Abstract]). This search retrieved 80 records.

For Scopus, the search strategy was adapted to the database-specific TITLE-ABS-KEY syntax and was intentionally broader to account for differences in database indexing and terminology. The final Scopus search strategy was: TITLE-ABS-KEY (("artificial intelligence" OR "machine learning" OR "deep learning") AND ("dental implant" OR "peri implant" OR "peri implantitis")). This search retrieved 58 records.

Thus, a total of 138 records were identified through electronic database searching before duplicate removal and eligibility screening. A supplementary targeted search was also performed to identify potentially relevant studies not retrieved by the predefined database search strategies. Additional records identified through this process were assessed according to the same eligibility criteria. Records identified through the database and supplementary searches were subsequently evaluated according to the predefined eligibility criteria described below.

Eligibility Criteria

Studies were considered eligible for the primary evidence synthesis if they met the following criteria: (1) original peer-reviewed research investigating artificial intelligence, machine learning, or deep-learning methods; (2) evaluation of dental implants and/or peri-implant tissues using radiographic images; (3) assessment of at least one imaging task directly related to peri-implant marginal bone status, including peri-implant landmark or tissue detection, marginal bone-level measurement, marginal bone-loss detection or quantification, radiographic bone-loss severity grading, prediction of marginal bone loss using radiographic or imaging-derived features or peri-implant bone-defect morphology classification; (4) use of clinical human radiographic images, including periapical, panoramic, or cone-beam computed tomography (CBCT) images; and (5) availability of the full text in English.

Studies describing AI-based peri-implantitis detection or peri-implantitis classification were considered eligible when the model assessed radiographic bone changes relevant to the scope of this review. However, such studies were classified according to the actual imaging task performed and were not considered to provide evidence of AI-based clinical diagnosis of peri-implantitis unless the reference standard incorporated appropriate clinical diagnostic parameters in addition to radiographic findings.

Studies were excluded if they (1) investigated AI applications exclusively for implant planning or placement, anatomical segmentation for surgical planning, implant identification, implant-brand or implant-system recognition, or prosthetic planning without assessment of peri-implant marginal bone status; (2) evaluated exclusively non-radiographic data; (3) were based exclusively on animal, in vitro, phantom, or synthetic datasets; (4) did not assess an outcome directly related to peri-implant marginal bone level, bone loss, bone remodeling, or peri-implant bone-defect morphology; or (5) were case reports, conference abstracts or proceedings, editorials, letters, commentaries, protocols, or other non-original publications.

Systematic reviews, meta-analyses, scoping reviews, and narrative reviews were excluded from the primary evidence synthesis. Consensus reports and clinical practice guidelines were used to establish clinical definitions and diagnostic criteria but were not considered part of the primary AI evidence set.

Study Selection

All records retrieved from PubMed/MEDLINE and Scopus were screened by title and abstract for potential relevance to the predefined eligibility criteria. The database searches identified a total of 138 records, including 80 records from PubMed/MEDLINE and 58 records from Scopus. After comparison of the records retrieved from the two databases, no duplicates were identified; therefore, all 138 records proceeded to title and abstract screening.

During title and abstract screening, records were excluded when they clearly did not meet the eligibility criteria, including studies based exclusively on non-radiographic clinical or demographic data, large language models applied to simulated clinical scenarios, microbiome or genomic datasets, and other non-imaging approaches to peri-implant disease prediction. Following this screening, 17 articles were considered potentially eligible and underwent full-text assessment.

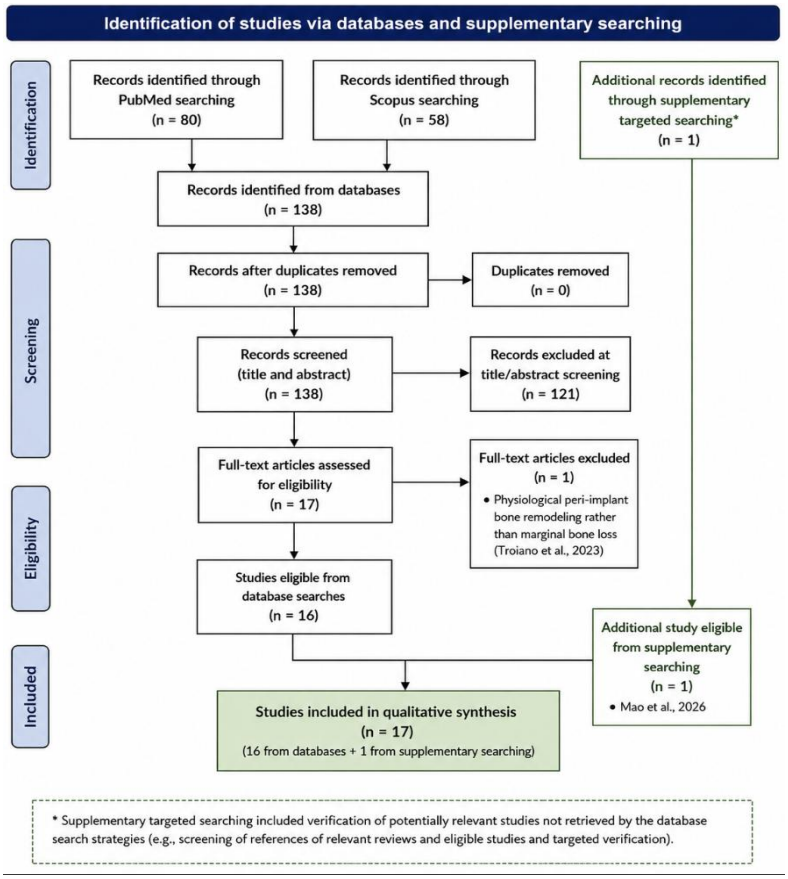
The full texts of these 17 potentially eligible articles were assessed against the predefined inclusion and exclusion criteria. One study was excluded at this stage because the investigated outcome was physiological peri-implant bone remodeling rather than marginal bone loss. Consequently, 16 studies retrieved through the database search fulfilled the eligibility criteria. None of the records retrieved exclusively through the Scopus search met the eligibility criteria for progression to full-text assessment. One additional eligible study was identified through supplementary targeted searching, resulting in a final sample of 17 original studies included in the qualitative synthesis.

Studies describing peri-implantitis detection based on radiographic images were retained when their actual imaging task involved peri-implant marginal bone assessment, bone-defect classification, or related radiographic outcomes. These studies were subsequently categorized according to the specific radiographic task performed rather than being interpreted uniformly as studies of clinical peri-implantitis diagnosis.

Any disagreement regarding eligibility or study classification was resolved through discussion among the authors until consensus was reached.

The study identification, screening, eligibility assessment, and final inclusion process is summarized in Figure 1.

Figure 1: Flow diagram of the literature search and study selection process across PubMed/MEDLINE and Scopus. AI, artificial intelligence; ML, machine learning; DL, deep learning.



Data Extraction and Evidence Synthesis

Data were extracted from each eligible study using a structured framework designed to characterize both the clinical imaging context and the methodological features of the artificial intelligence models. The extracted variables included: first author and year of publication; study design; sample and dataset characteristics; radiographic modality; AI or machine-learning approach; specific radiographic task; reference standard and annotation procedure; training, validation, and test strategy; performance metrics; comparator, when available; and the main findings relevant to peri-implant marginal bone assessment.

Particular attention was given to the specific imaging task performed by each model, rather than relying solely on the diagnostic terminology used by the original authors. Accordingly, eligible studies were categorized according to whether the AI approach was designed for implant or peri-implant landmark localization, marginal bone-level detection or measurement, marginal bone-loss detection or quantification, radiographic severity grading, peri-implant bone-defect morphology classification, or prediction of marginal bone loss using radiographic or imaging-derived features.

For studies using the terms peri-implantitis detection, peri-implantitis diagnosis, or peri-implantitis classification, the reference standard was examined to determine whether peri-implantitis had been defined using radiographic findings alone or a combination of radiographic and clinical parameters. Models based exclusively on radiographic information were interpreted as assessing radiographic manifestations associated with peri-implant disease rather than independently establishing a clinical diagnosis of peri-implantitis.

Because of the heterogeneity of the included studies with respect to imaging modality, AI architecture, dataset characteristics, reference standards, target tasks, and reported performance measures, no quantitative meta-analysis was performed. The findings were therefore synthesized narratively and organized according to the principal radiographic tasks and methodological characteristics of the included models. Performance measures were reported as provided by the original studies and were not pooled across studies when the underlying tasks or reference standards differed.

Methodological and Clinical Appraisal

In addition to descriptive data extraction, the included studies were critically appraised with respect to methodological features that may influence the reliability, generalizability, and clinical interpretation of their findings. The appraisal focused on dataset size and composition, radiographic modality, definition of the target outcome, reference-standard and annotation procedures, data-partitioning strategy, internal and external validation, class distribution, reported performance metrics, and comparison with clinician assessment when available.

Particular attention was paid to the unit of data splitting. When multiple radiographs, implants, image sections, or longitudinal examinations originated from the same patient, the reported training, validation, and test procedures were examined to determine whether partitioning was performed at the patient level or at the image/implant level. The potential for data leakage and overly optimistic performance estimates was considered when independence between training and evaluation datasets could not be established from the reported methodology.

The reference standard used for model development and evaluation was also examined. Studies were assessed according to whether annotations or diagnostic labels were established by one or more clinicians, whether examiner expertise was reported, whether consensus procedures or repeated assessments were used, and whether clinical information was incorporated into the reference standard. Particular caution was applied to studies in which radiographic bone loss or defect morphology alone was labelled as peri-implantitis, because radiographic findings in isolation do not establish the clinical diagnosis of peri-implantitis.

Model validation was categorized according to whether performance was evaluated using an internal test set or an independent external dataset. The presence of clinician comparison, multicenter data, heterogeneous imaging sources, and inclusion of different implant systems was additionally considered when assessing the potential generalizability of each model. Studies relying on single-center datasets, limited implant systems, or internally derived test sets were interpreted cautiously with respect to clinical transferability.

Because the included studies investigated heterogeneous AI tasks and reported different performance measures, model performance was interpreted within the context of the specific task and reference standard used in each study. Accuracy, sensitivity, specificity, precision, recall, F1-score, intersection over union, Dice coefficient, average precision, and other reported metrics were therefore not considered directly interchangeable across studies. High numerical performance was not interpreted as evidence of clinical diagnostic accuracy unless the model outcome and reference standard supported such an interpretation.

The appraisal was used to support the narrative interpretation of the evidence rather than to generate a numerical methodological quality score. No studies were excluded solely on the basis of methodological limitations identified during this appraisal; instead, these limitations were considered when evaluating the strength and clinical applicability of the available evidence. A structured risk-of-bias instrument, such as QUADAS-2 or PROBAST, was not applied, because the included studies addressed heterogeneous AI tasks (localization, measurement, severity classification, and prediction) with correspondingly different reference standards and validation

designs, limiting the applicability of a single standardized tool across all studies. Instead, the domains most relevant to these instruments — reference-standard adequacy, patient-level data partitioning, and validation strategy — were appraised narratively for each study within the context of its specific task.

RESULTS

Study Characteristics

Seventeen original studies met the predefined eligibility criteria and were included in the qualitative synthesis. The included studies were published between 2020 and 2026, reflecting the relatively recent application of artificial intelligence and machine-learning methods to radiographic assessment of peri-implant marginal bone conditions. The studies differed substantially in terms of imaging modality, dataset size, AI architecture, target task, reference standard, and reported performance measures.

Most studies used two-dimensional dental radiographs, particularly periapical radiographs, either alone or in combination with bitewing or panoramic images. For example, Cha et al. evaluated a modified region-based convolutional neural network (R-CNN) on 708 periapical radiographs for detection of the marginal bone level and implant landmarks, subsequent calculation of the peri-implant bone-loss ratio, and classification of bone-resorption severity [4]. Liu et al. used 1,670 periapical radiographs to train and evaluate a Faster R-CNN model for detecting marginal bone loss around dental implants [8]. Vera et al. investigated a combined deep-learning and image-processing approach using 2,920 intraoral periapical and bitewing radiographs to localize implant-related structures and automatically determine peri-implant bone loss [5].

The scope of the included AI applications extended beyond direct marginal bone-loss detection. Some models were developed to identify implants, peri-implant tissues, or radiographic landmarks required for subsequent measurements, whereas others quantified marginal bone levels, classified the severity of radiographic bone loss, characterized peri-implant bone-defect morphology, or distinguished radiographic patterns labelled by the original investigators as peri-implantitis. More recent studies also investigated three-dimensional CBCT-based assessment and prediction of marginal bone loss using imaging-derived features. Consequently, the included studies represented several related but methodologically distinct AI tasks rather than a homogeneous group of diagnostic models.

Convolutional neural networks and related deep-learning architectures predominated among the included studies. The investigated approaches included region-based CNNs, Faster R-CNN, YOLO variants, U-Net-based segmentation, conventional CNN classifiers, and combined object-detection and image-processing pipelines. A smaller subset used conventional machine-learning algorithms applied to quantitative imaging-derived features. This methodological diversity was accompanied by substantial variation in model outputs and evaluation metrics, ranging from landmark localization and segmentation measures to sensitivity, specificity, precision, recall, F1-score, accuracy, and agreement with clinician assessments.

An important source of heterogeneity concerned the definition of the target condition. While some studies explicitly evaluated marginal bone level or marginal bone loss as a radiographic outcome, others described their models as detecting or diagnosing peri-implantitis on radiographs. These studies were retained when their actual AI task fulfilled the eligibility criteria but were interpreted according to the radiographic feature assessed and the reference standard used. Therefore, the use of the term peri-implantitis by the original authors was not considered, by itself, evidence that the model established a clinical diagnosis of peri-implantitis.

The principal characteristics of the 17 included studies, including imaging modality, dataset characteristics, AI approach, radiographic task, reference standard, and principal outcomes are summarized in Table 1, and their methodological and validation features in Table 2.

Table 1: Main characteristics of the studies included in the qualitative synthesis

Study	Imaging modality / dataset	AI approach	Primary radiographic task	Reference standard / annotation	Main reported outcome
Su et al., 2026 [16]	Periapical radiographs; 708 images; 3i and Xive implants	YOLOv8 + custom segmentation + alveolar crest detection + EfficientNet-B3	Implant localization and classification of peri-implant bone-loss severity	Expert/manual radiographic assessment	YOLOv8 precision 99.39%, sensitivity 98.63%; 12-class accuracy 97.42%
Efeoglu et al., 2026 [17]	Panoramic radiographs; 144 patients	Mask R-CNN (ResNet-50) + radiomics/ML	Detection of radiographic peri-implant abnormalities	Regions segmented by two dentomaxillofacial radiology residents; ICC 0.97	Mask R-CNN F1-score 0.84; AUC 0.86
Adnan et al., 2026 [18]	Sequential periapical radiographs; 2,960 retained images	Six-model DL cascade	Implant/ROI detection and automated MBL quantification	Specialist-established reference standard	MBL model mAP 75.0%, precision 73.0%, recall 69.8%

Madani et al., 2026 [19]	CBCT; 699 2D sections	YOLOv8	MBL detection and severity grading	Four radiographic classes according to percentage of implant length affected	Accuracy, precision, recall and F1 ≈ 0.90; mAP@0.5 = 0.889
Gao et al., 2025 [20]	Periapical radiographs	YOLOv8-pose	Landmark/keypoint detection, MBL quantification and severity prediction	Radiographic keypoints associated with peri-implant bone loss	Keypoint precision >98%; severity accuracy 0.906 and 0.844 for the two implant sides
Khan et al., 2025 [21]	CBCT; 150 patients, 212 implants	CNN-based system	Implant classification and peri-implant bone-defect/disease detection	Clinical and radiographic assessment validated by two oral radiologists	Sensitivity 91.8%, specificity 93.4%, AUC 0.94
Lee et al., 2025 [12]	Panoramic; 1,075 radiographs, 2,250 implant sites	Ensemble YOLOv8	Bone-defect morphology and severity classification	Expert annotations; comparison with periodontal surgeons	AI overall accuracy 85.33% vs 75.6% mean surgeon accuracy
Kibcak et al., 2025 [22]	Orthopantomographs; 7,696 images	U-Net + CNN	Implant segmentation and radiographic peri-implantitis classification	Clinical identification/assessment by medical professionals	Segmentation DSC 0.986; classification F1-score 0.835
Lee et al., 2024 [10]	Periapical radiographs; 800 images	YOLOv7	Detection and classification of MBL severity	Radiographically defined severity classes	Specificity 100%, precision 100%, recall 94.44%, F1-score 97.10%
Vera et al., 2023 [5]	Periapical and bitewing radiographs; 2,920 images	YOLOv3 + image-processing pipeline	Implant/prosthesis localization, landmark detection and MBL measurement	Dentist-defined reference landmarks	mAP@0.5 varied according to detected structure; automated bone-loss measurements obtained
Chen et al., 2023 [23]	Periapical radiographs	CNN-based system	Implant localization and grading of peri-implant bone damage	Radiographic assessment according to implant thread involvement	Implant localization accuracy 89.31%; damage-grading accuracy 90.45%
Zhang et al., 2023 [24]	529 periapical + 551 panoramic images; 248 patients	Deep CNN/hybrid imaging model	Implant-outcome classification incorporating MBL	Radiographic classification of failure with MBL, failure without MBL, and success	Combined-image accuracy 87%; AUC 0.972 for failure with MBL
Jang et al., 2022 [9]	Periapical radiographs; 300 images	Faster R-CNN with transfer learning	Implant and peri-implant tissue localization	Manually labelled implant/peri-implant tissue regions	Precision 0.977, recall 0.992, F1-score 0.984, mean IoU 0.907
Liu et al., 2022 [8]	Periapical radiographs; 1,670 images	Faster R-CNN	Direct detection of peri-implant MBL	Clinician annotation with expert confirmation	Agreement with expert: κ=0.547 for bone-loss sites and κ=0.568 for implants with bone loss
Cha et al., 2021 [4]	Periapical radiographs; 708 images	Modified region-based CNN	Implant landmark detection, MBL measurement and severity classification	Clinician-defined implant top, apex and marginal bone landmarks	No significant difference between AI and clinician landmark localization; automated bone-loss ratio calculated
Zhang et al., 2020 [25]	CBCT; 81 patients	SVM, ANN, logistic regression and random forest	Prediction of severe MBL from CBCT-derived trabecular features	Severe-MBL versus control classification	Best SVM: AUC 0.967, sensitivity 91.67%, specificity 100%
Mao et al., 2026 [11]	Periapical radiographs; 780 images (1,210 dental implants)	YOLO-based deep learning + image processing	Peri-implant region detection and radiographic bone-loss severity assessment	Radiographic peri-implant bone-loss severity (three grades)	Automated bone-loss severity assessment; performance varied by severity category

Radiographic Modalities and Dataset Characteristics

Two-dimensional dental radiographs represented the predominant imaging source among the included studies. Periapical radiographs were the most frequently investigated modality and were used for implant and peri-implant landmark localization, marginal bone-level measurement, bone-loss detection and quantification, and radiographic severity classification [4,8–11, 16, 18, 20, 23].

Panoramic radiographs were evaluated either independently or in combination with intraoral images. Kibcak et al. used orthopantomographs for implant segmentation and classification of peri-implant conditions [22], whereas Lee et al. investigated panoramic images for automated classification of peri-implant bone-defect morphology and severity [12]. Efeoglu et al. combined

deep-learning-based segmentation with radiomic analysis of panoramic radiographs [17]. Zhang et al. incorporated both periapical and panoramic radiographs into a deep-learning framework for implant-outcome prediction involving marginal bone loss [24]. Vera et al. included both periapical and bitewing radiographs in an automated pipeline for implant-related landmark detection and marginal bone assessment [5].

Three-dimensional imaging was less frequently represented. Madani and Bashizadeh Fakhar applied a YOLOv8-based approach to two-dimensional sections derived from CBCT datasets for detection and grading of peri-implant marginal bone loss [19]. Khan et al. investigated a CNN-based approach using CBCT data for implant classification and peri-implant disease assessment [21]. Zhang et al. used CBCT-derived trabecular microstructural parameters as input variables for machine-learning models designed to predict severe marginal bone loss [25].

Dataset size varied substantially among the included studies. Zhang et al. evaluated CBCT-derived features in 81 patients [25], Jang et al. used 300 periapical radiographs [9], and Cha et al. and Su et al. each evaluated 708 periapical radiographs [4,16]. Liu et al. used 1,670 periapical radiographs [8]. Larger image datasets included 2,920 intraoral radiographs in the study by Vera et al. [5], 2,960 retained periapical images in the study by Adnan et al. [18], and 7,696 orthopantomographs in the study by Kibcak et al. [22]. The unit of analysis also varied across studies, with datasets reported according to patients, implants, radiographs, implant sites, or image sections.

AI Tasks in Peri-Implant Marginal Bone Assessment

The included studies addressed different radiographic tasks, including implant and peri-implant structure localization, landmark detection, marginal bone-level measurement, bone-loss detection and quantification, severity grading, bone-defect classification, and prediction based on imaging-derived features. The studies were therefore grouped according to the principal AI-assisted radiographic task investigated.

Implant, Tissue, and Landmark Localization

Several studies incorporated automated localization or landmark detection either as the primary task or as an initial component of a multi-stage bone-assessment workflow. Jang et al. developed a Faster R-CNN model using 300 periapical radiographs for automated detection of dental implants and peri-implant tissues. The model achieved a precision of 0.977, recall of 0.992, F1-score of 0.984, and mean intersection over union of 0.907 [9]. Cha et al. used a modified region-based CNN to identify the implant top, implant apex, and marginal bone level on periapical radiographs. These landmarks were subsequently used to calculate the ratio of marginal bone loss to implant length. The study reported no significant differences between model- and clinician-derived landmark localization measurements [4]. Vera et al. combined deep learning and image-processing techniques for automatic identification of implants, prosthetic components, and reference landmarks on periapical and bitewing radiographs. The detected landmarks were subsequently used for automated assessment of peri-implant marginal bone changes [5]. Localization was also incorporated into more recent multi-stage approaches. Adnan et al. developed a six-model deep-learning cascade that included image alignment, region-of-interest detection, implant-related structure identification, and subsequent marginal bone-loss estimation from sequential periapical radiographs [18]. Gao et al. used a YOLOv8-pose model for detection of radiographic keypoints required for automated MBL measurement, reporting keypoint-detection precision above 98% [20]. Su et al. combined YOLOv8-based implant localization with segmentation, alveolar crest detection, and subsequent classification of peri-implant bone-loss levels [16].

Marginal Bone-Level Measurement and Bone-Loss Quantification

Several studies investigated automated detection, measurement, or quantification of peri-implant marginal bone loss using two-dimensional radiographic images [4, 5, 8, 18, 20]. Cha et al. developed a region-based CNN approach for peri-implant bone-loss measurement on periapical radiographs. Following identification of the implant top, apex, and marginal bone level, the system calculated marginal bone loss relative to implant length [4]. Vera et al. combined deep learning with image-processing techniques to identify implants, prosthetic structures, and relevant reference points on periapical and bitewing radiographs. These landmarks were subsequently used to determine peri-implant marginal bone-level measurements [5]. Liu et al. evaluated a Faster R-CNN model using 1,670 periapical radiographs for automated detection of marginal bone loss around dental implants. Agreement between the model and expert assessment was reported with κ values of 0.547 for bone-loss sites and 0.568 for implants presenting bone loss [8]. Adnan et al. developed a multi-stage deep-learning cascade for automated marginal bone-loss assessment from sequential periapical radiographs. Following image alignment and localization steps, the MBL component achieved a mean average precision of 75.0%, precision of 73.0%, and recall of 69.8% [18]. Gao et al. applied a YOLOv8-pose architecture to detect key radiographic landmarks and calculate peri-implant marginal bone loss from periapical radiographs. Keypoint-detection precision exceeded 98%, while the subsequent severity assessment achieved accuracies of 0.906 and 0.844 for the two implant sides [20].

Bone-Loss Severity and Defect-Morphology Classification

Several studies applied deep-learning models to classify the severity or morphological characteristics of peri-implant bone loss [10–12, 16, 19, 21–23]. Lee et al. developed a YOLOv7-based model using 800 periapical radiographs for classification of radiographic peri-

implant bone-loss severity. The model achieved a specificity and precision of 100%, recall of 94.44%, and F1-score of 97.10% [10]. Because these figures were obtained on a 100-image internal test set, the near-ceiling specificity and precision should be interpreted with caution. Mao et al. combined deep learning and image-processing methods for automated assessment of peri-implant disease severity on periapical radiographs. The workflow included localization of peri-implant regions followed by classification according to predefined radiographic severity categories [11]. Su et al. combined implant localization, alveolar crest detection, segmentation, and classification components for automated assessment of peri-implant marginal bone-loss levels on periapical radiographs. The multiclass classification component achieved an accuracy of 97.42% [16]. Madani and Bashizadeh Fakhar applied YOLOv8 to 699 two-dimensional sections derived from CBCT datasets. Marginal bone loss was classified into four categories according to the proportion of implant length affected. Accuracy, precision, recall, and F1-score were approximately 0.90, with an mAP@0.5 of 0.889 [19]. Lee et al. evaluated an ensemble YOLOv8 model using panoramic radiographs for classification of peri-implant bone-defect morphology and severity. The model achieved an overall accuracy of 85.33%, compared with a mean accuracy of 75.6% among the periodontal surgeons included as comparators [12]. Kibcak et al. combined U-Net-based implant segmentation with CNN-based classification on orthopantomographs. The segmentation model achieved a Dice similarity coefficient of 0.986, whereas the classification component showed a precision of 0.777, recall of 0.903, and F1-score of 0.835 [22]. Khan et al. evaluated a CNN-based approach using CBCT data for dental implant classification and peri-implant disease detection. The reported sensitivity was 91.8%, specificity 93.4%, and AUC 0.94 [21]. Chen et al. used a CNN-based system for implant localization and grading of peri-implant bone damage on periapical radiographs, reporting accuracies of 89.31% for implant localization and 90.45% for grading of peri-implant damage [23].

Prediction Models Using Imaging-Derived Features

A smaller subset of the included studies investigated predictive applications of artificial intelligence using radiographic images or quantitative imaging-derived features rather than direct detection or measurement of marginal bone loss [24, 25]. Zhang et al. developed a deep-learning approach for dental implant outcome prediction using periapical and panoramic radiographs obtained from 248 patients. The dataset included 529 periapical and 551 panoramic images, and implant outcomes were categorized as failure with marginal bone loss, failure without marginal bone loss, or success. Models were developed using periapical and panoramic images separately and in combination. The combined-image approach achieved an overall prediction accuracy of 87%, while an AUC of 0.972 was reported for the classification of implant failure associated with marginal bone loss [24]. In a different predictive approach, Zhang et al. investigated whether trabecular microstructural parameters extracted from CBCT images could predict severe marginal bone loss around mandibular dental implants. The study included 81 patients, comprising 41 cases with severe marginal bone loss and 40 controls. Several machine-learning algorithms were evaluated, including support vector machine, artificial neural network, logistic regression, and random forest models. The support vector machine achieved the highest reported performance, with an AUC of 0.967, sensitivity of 91.67%, and specificity of 100% [25]. Thus, among the included studies, predictive approaches comprised both direct analysis of radiographic images for implant-outcome classification and machine-learning analysis of quantitative features extracted from CBCT datasets [24, 25]. Notably, neither predictive study was prospective or longitudinal: Zhang et al. [24] performed cross-sectional classification of implant outcomes, and Zhang et al. [25] used a 41-versus-40 case–control design, so the reported 100% specificity [25] most plausibly reflects the small, balanced internal sample rather than an ability to forecast future bone loss from a baseline radiograph. This reinforces that progression over time, rather than a single image, is the clinically decisive signal.

Model Performance and Validation Characteristics

The included studies used heterogeneous performance measures according to the specific AI task investigated. Object-detection and localization models were commonly evaluated using precision, recall, F1-score, intersection over union (IoU), and mean average precision (mAP), whereas classification models more frequently reported accuracy, sensitivity, specificity, F1-score, and area under the receiver operating characteristic curve (AUC) [8–12, 16–25]. Studies involving quantitative bone-level assessment additionally reported agreement with expert measurements or comparisons between automated and manually derived values [4, 5, 8, 18, 20].

Validation strategies also varied across studies. Several investigations divided available datasets into training, validation, and/or test subsets. Cha et al. used 508 periapical radiographs for training, 100 for validation, and 100 for testing [4]. Lee et al. divided 800 periapical radiographs into 600 training, 100 validation, and 100 test images [10], whereas Madani and Bashizadeh Fakhar used an 80:10:10 allocation for training, testing, and validation datasets [19]. Kibcak et al. used a 70:15:15 distribution for training, validation, and testing [22].

Other studies used different internal evaluation strategies. Jang et al. allocated 80% of 300 periapical radiographs to training and 20% to testing [9], while Lee et al. used 80% of their panoramic dataset for training and 20% for testing [12]. Vera et al. reported a 50:50 division for the deep-learning component of their image-analysis workflow [5]. Adnan et al. evaluated individual components of their multi-model cascade and additionally compared automated marginal bone-loss severity scores with specialist-derived reference assessments [18].

Comparison with human observers was incorporated into several studies. Cha et al. compared AI-derived landmark localization with clinician measurements [4], while Lee et al. compared the performance of their AI model for peri-implant bone-defect classification

with that of periodontal surgeons [12]. Liu et al. assessed agreement between model predictions and expert interpretation using Cohen's kappa [8].

Reference-standard and annotation procedures differed among the included investigations. Depending on the study, radiographic regions, landmarks, bone-loss measurements, or severity categories were established by dentists, dental specialists, oral and maxillofacial radiologists, periodontal specialists, or combinations of expert assessors [4, 5, 8, 12, 18, 21]. Some studies reported interobserver or agreement measures as part of the reference-standard procedure, whereas others provided more limited information regarding annotation reproducibility.

When these methodological features are tabulated across the 17 included studies (Table 2), several patterns become quantifiable. The reference standard was derived from radiographic information alone in 16 of the 17 studies; only Khan et al. [21] reported incorporating clinical assessment. No study reported independent external validation on a separate dataset, and none explicitly described patient-level data partitioning; where the partition was specified, data were divided at the image or radiograph level. A direct comparison against clinician performance was reported in four studies [4, 8, 12, 18]. Near-ceiling performance on small internal test sets was observed in several reports, including 100% specificity and precision in Lee et al. [10] (100-image test set) and 100% specificity in Zhang et al. [25] (41 cases, 40 controls). These counts substantiate the central caution of this review: with a single exception, current radiograph-based models are trained and evaluated against radiographic reference standards, so their reported performance reflects reproduction of radiographic labels rather than a validated clinical diagnosis of peri-implantitis.

Table 2: Methodological and clinical appraisal of the 17 included studies, derived from the characteristics reported in each primary study.

Study	Reference standard	Data split (unit)	External validation	Clinician comparison	Interpretive flag
Cha et al., 2021 [4]	Radiographic-only	Image-level (508/100/100)	No	Yes	—
Vera et al., 2023 [5]	Radiographic-only	Image-level (50:50)	No	No	—
Liu et al., 2022 [8]	Radiographic-only	NR	No	Yes	Moderate agreement (κ 0.55–0.57)
Jang et al., 2022 [9]	Radiographic-only	Image-level (80/20)	No	No	Near-ceiling F1 0.984
Lee et al., 2024 [10]	Radiographic-only	Image-level (600/100/100)	No	No	100% specificity/precision; test n=100
Mao et al., 2026 [11]	Radiographic-only	NR	No	No	—
Lee et al., 2025 [12]	Radiographic-only	Image-level (80/20)	No	Yes	vs periodontal surgeons
Su et al., 2026 [16]	Radiographic-only	NR	No	No	Near-ceiling precision 99.4%; accuracy 97.4%
Efeoglu et al., 2026 [17]	Radiographic-only	NR	No	No	—
Adnan et al., 2026 [18]	Radiographic-only	NR	No	Yes	vs specialist reference
Madani & Bashizadeh Fakhar, 2026 [19]	Radiographic-only (CBCT 2D sections)	Image-level (80:10:10)	No	No	—
Gao et al., 2025 [20]	Radiographic-only	NR	No	No	—
Khan et al., 2025 [21]	Radiographic + clinical	NR	No	No	Only study incorporating clinical data
Kibcak et al., 2025 [22]	Radiographic-only†	Image-level (70:15:15)	No	No	†label termed clinical but applied to radiographs
Chen et al., 2023 [23]	Radiographic-only	NR	No	No	—
Zhang et al., 2023 [24]	Radiographic-only	NR (248 patients)	No	No	Prediction but cross-sectional
Zhang et al., 2020 [25]	Radiographic-only	NR (81 patients)	No	No	Prediction but case–control; 100% specificity

Entries for external validation and data-partitioning unit reflect what was reported; cells marked NR were not explicitly reported and were not independently verified against the source datasets. RS, reference standard; NR, not reported; κ, Cohen kappa.

DISCUSSION

The present review identified 17 original studies investigating artificial intelligence for radiographic assessment of peri-implant bone conditions. The available evidence demonstrates a rapid expansion of AI applications in this field, particularly during the most recent years, but also reveals substantial heterogeneity in imaging modalities, dataset characteristics, model architectures, target tasks, reference standards, and performance metrics. Most studies relied on two-dimensional radiographs, particularly periapical images, while panoramic radiographs and CBCT-based approaches were less frequently investigated. Importantly, the reviewed models did not address a single diagnostic task. Their applications ranged from implant and landmark localization to marginal bone-level measurement, detection and quantification of bone loss, severity grading, defect-morphology classification, and prediction based on imaging-derived features [4,5,8–12,16–25].

This distinction between AI tasks is clinically important when interpreting the available evidence. Localization models can identify implants, peri-implant tissues, or anatomical landmarks with high technical performance, as demonstrated by Jang et al. [9] and by the localization components incorporated into several multi-stage workflows [4, 5, 16, 18, 20]. Such tasks provide the technical basis for subsequent automated measurements but do not themselves represent disease detection. Similarly, models developed for quantitative assessment provide estimates of marginal bone levels or the extent of bone loss [4,5,8,18,20], whereas classification models assign images to predefined radiographic severity or defect categories [10–12, 16, 19, 21–23]. Consequently, performance values reported for these different tasks should not be interpreted as measures of the same clinical outcome.

A central issue emerging from the reviewed literature is the distinction between radiographically detected marginal bone loss and clinically diagnosed peri-implantitis. According to the 2017 World Workshop, peri-implantitis is a plaque-associated pathological condition characterized by inflammation of the peri-implant mucosa and progressive loss of supporting bone [1]. Diagnostic assessment therefore requires integration of clinical inflammatory findings, probing information, and radiographic bone-level changes rather than radiographic evidence alone [1, 3]. Marginal bone loss identified on an image is consequently an imaging finding whose clinical significance depends on its magnitude, progression, timing, and association with clinical signs of disease. Early crestal remodeling after implant placement and loading must also be distinguished from progressive pathological bone loss.

This distinction is particularly relevant because several studies included in the present review use terms such as peri-implantitis detection, peri-implant disease detection, or peri-implantitis severity for models whose principal input consists of radiographic images [10–12, 21–23]. These studies provide valuable evidence regarding automated recognition of peri-implant bone alterations, but terminology should reflect the actual information available to the algorithm. When a model receives only a radiograph and no information regarding bleeding or suppuration on probing, probing depth, previous bone levels, or other clinical findings, its output is more appropriately interpreted as detection or classification of radiographic features associated with peri-implant bone loss rather than an autonomous clinical diagnosis of peri-implantitis. This distinction is especially important when previous radiographs are unavailable, because a single image cannot establish whether an observed bone level represents stable remodeling or progressive disease [1, 3].

The reported performance of the investigated models was generally promising, although direct numerical comparisons should be approached cautiously. Localization models reported high precision and recall, with Jang et al. achieving an F1-score of 0.984 [9], while several severity-classification systems reported accuracies or F1-scores approaching or exceeding 90% [10, 16, 19, 23]. Lee et al. reported an overall accuracy of 85.33% for AI-based classification of peri-implant bone-defect morphology and severity, compared with a mean accuracy of 75.6% among periodontal surgeons [12]. In contrast, Liu et al. reported only moderate agreement between their Faster R-CNN model and expert assessment, with κ values of 0.547 for bone-loss sites and 0.568 for implants presenting bone loss [8]. Adnan et al. reported a mean average precision of 75.0%, precision of 73.0%, and recall of 69.8% for the marginal bone-loss component of their multi-stage system [18]. These differences illustrate that apparently high performance in one technical task cannot be directly compared with performance obtained using another endpoint or metric.

Heterogeneity in reference standards represents another important consideration. Ground truth across the included studies was established using different approaches, including clinician-defined landmarks, manually annotated bone levels, study-specific severity thresholds, expert radiographic classifications, and, in some cases, combinations of clinical and radiographic findings [4, 5, 8, 12, 18, 21]. The expertise and number of annotators also varied. This is particularly relevant for supervised learning, because model performance is inherently linked to the labels used during training and evaluation. An algorithm trained to reproduce a radiographic severity classification can, at best, reproduce that classification with a certain degree of accuracy; its performance cannot automatically be extrapolated to a broader clinical diagnosis that was not represented in the reference standard.

Dataset characteristics and validation strategies further influence interpretation of the reported results. Dataset sizes ranged from relatively small patient cohorts to collections containing several thousand radiographs [4, 5, 8, 9, 16,18,22,25]. However, the number of images does not necessarily represent the number of statistically independent clinical observations, as individual patients may contribute multiple implants, bilateral measurements, sequential examinations, or several images. In AI research, the level at which data are separated into training, validation, and test sets is therefore important. If images originating from the same patient or implant

are distributed across different subsets, similarities between related observations may contribute to optimistic estimates of model performance. Clear reporting of patient-level data partitioning is consequently desirable in future investigations.

Another limitation of the current evidence is the predominance of internal validation. Most studies developed and evaluated their algorithms within datasets originating from a limited number of institutions, imaging systems, or patient populations [4, 5, 8–12, 16–25]. Models may therefore learn characteristics related not only to peri-implant bone morphology but also to acquisition protocols, image quality, implant design, or dataset-specific patterns. Independent external validation using images acquired in different clinical environments is necessary to establish whether reported performance can be reproduced across populations, radiographic systems, implant brands, and operators. This requirement is particularly relevant for peri-implant imaging because implant geometry, thread design, prosthetic components, image projection, and superimposition can substantially alter radiographic appearance.

The predominance of periapical radiographs among the included studies is clinically understandable because these images are routinely obtained during implant follow-up and provide relatively detailed visualization of mesial and distal marginal bone levels [4,8–11, 16, 18, 20, 23]. Nevertheless, two-dimensional imaging has intrinsic anatomical limitations. Buccal and lingual bone morphology cannot be directly represented, and projection geometry may influence the apparent relationship between the implant and marginal bone. CBCT-based AI approaches may provide additional three-dimensional information [19, 21, 25], but their role should be interpreted within established principles of radiological justification, since CBCT is not indicated as a routine substitute for intraoral radiography solely for peri-implant surveillance.

An especially relevant potential application of AI may therefore be measurement support and longitudinal monitoring rather than autonomous diagnosis. Automated identification of the implant platform, apex, marginal bone contact, and other reference landmarks could reduce the time required for repeated measurements and improve consistency between follow-up examinations [4, 5, 18, 20]. Systems capable of aligning sequential radiographs and quantifying changes over time, such as the approach investigated by Adnan et al. [18], may be particularly valuable because progression of bone loss provides more clinically meaningful information than an isolated radiographic measurement. AI-generated measurements could therefore function as decision-support information to be integrated with probing findings, inflammatory signs, risk factors, and previous examinations.

The present findings are broadly consistent with previous evidence syntheses reporting promising performance of AI methods for peri-implant radiographic analysis [13–15]. However, systematic reviews that aggregate diagnostic-performance estimates face an important conceptual challenge when the underlying studies evaluate different target tasks. Localization, segmentation, marginal bone-level measurement, severity classification, defect-morphology recognition, and prediction are technically and clinically distinct endpoints. Pooling or comparing performance estimates without accounting for these differences may obscure what an algorithm has actually been validated to accomplish. The present review therefore emphasizes task-specific interpretation of the available evidence rather than treating all radiographic AI models as diagnostic systems for peri-implantitis.

A practical consequence of this task-based reading is a figure that pooled-accuracy syntheses [13–15] do not surface: 16 of the 17 studies established their reference standard from radiographic information alone (Table 2), so a single summary estimate of diagnostic accuracy risks describing agreement with radiographic labels rather than detection of clinical disease. The review itself has several limitations. Only PubMed and Scopus were used for the structured database search, and one additional eligible study was identified through supplementary targeted searching. Although the search strategy was designed around AI, dental implants, and peri-implant bone outcomes, relevant studies using different terminology may not have been retrieved. The heterogeneity of the included studies precluded meaningful quantitative pooling and supported a narrative synthesis. Furthermore, differences in study design, imaging modality, reference standards, dataset composition, model architecture, and reported performance measures limited direct comparisons between individual algorithms. The rapidly evolving nature of AI research should also be considered, as recently developed models may not yet have undergone independent replication or external clinical validation.

Future research should prioritize transparent reporting of dataset composition, patient-level separation of training and test data, clearly defined annotation protocols, interobserver agreement where appropriate, and independent external validation. Multicenter datasets incorporating different radiographic systems and implant designs would improve assessment of model generalizability. Longitudinal datasets are particularly important for distinguishing stable peri-implant bone levels from progressive changes. Finally, future systems intended specifically to support peri-implantitis diagnosis should integrate radiographic analysis with relevant clinical variables, including bleeding and/or suppuration on probing, probing depth, previous bone-level measurements, and patient- and implant-level risk information. Such multimodal approaches would more closely correspond to the clinical definition of peri-implantitis than models based solely on radiographic appearance [1, 3].

CONCLUSION

Artificial intelligence shows promising potential for supporting the radiographic assessment of peri-implant marginal bone conditions. Current applications include implant and landmark localization, marginal bone-level measurement and quantification, radiographic severity and defect-morphology classification, and prediction based on imaging-derived features. However, the available evidence

remains heterogeneous with respect to imaging modalities, datasets, reference standards, AI architectures, validation strategies, and reported performance metrics.

Importantly, AI performance in detecting or classifying radiographic bone alterations should not be interpreted as equivalent to diagnostic accuracy for peri-implantitis. Marginal bone loss is a radiographic finding, whereas peri-implantitis is a clinical condition whose diagnosis requires integration of inflammatory signs, probing findings, and progressive bone loss [1, 3]. Accordingly, current radiograph-based AI systems are better regarded as tools for image analysis, quantitative assessment, and clinical decision support rather than autonomous diagnostic systems.

Future studies should prioritize standardized outcome definitions, transparent annotation and validation procedures, patient-level data partitioning, independent external validation, and multicenter datasets. Longitudinal and multimodal models integrating radiographic changes with relevant clinical parameters may provide a more appropriate basis for the development of AI-assisted systems for peri-implant disease assessment.

Conflicts of interest and sources of funding

The authors declare no conflict of interest.

This research received no external funding.

Authors' contribution

Conceptualization, A.M.C., V.N. and I.S.; methodology, A.M.C. and E.A.R.; literature search and data curation, A.M.C., E.A.R. and A.T.; formal analysis, A.M.C., V.N. and D.E.; visualization, A.M.C. and E.A.R.; writing—original draft preparation, A.M.C.; writing—review and editing, V.N., I.S., E.A.R., D.E., A.T. and V.D.S.; supervision, V.N., I.S. and V.D.S.; project administration, A.M.C. All authors have read and agreed to the published version of the manuscript.

Ethics approval and consent to participate

Not applicable.

References

1. Berglundh, T.; Armitage, G.; Araujo, M.G.; Avila-Ortiz, G.; Blanco, J.; Camargo, P.M.; Chen, S.; Cochran, D.; Derks, J.; Figuero, E.; et al. Peri-implant diseases and conditions: Consensus report of workgroup 4 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. J. Clin. Periodontol. 2018, 45 (Suppl. 20), S286-S291. https://doi.org/10.1111/jcpe.12957.

2. Renvert, S.; Hirooka, H.; Polyzois, I.; Kelekis-Cholakis, A.; Wang, H.L.; Working Group 3. Diagnosis and non-surgical treatment of peri-implant diseases and maintenance care of patients with dental implants—Consensus report of Working Group 3. Int. Dent. J. 2019, 69 (Suppl. 2), 12–17. https://doi.org/10.1111/idj.12490.

3. Herrera, D.; Berglundh, T.; Schwarz, F.; Chapple, I.; Jepsen, S.; Sculean, A.; Kebschull, M.; Papapanou, P.N.; Tonetti, M.S.; Sanz, M.; EFP workshop participants and methodological consultant. Prevention and treatment of peri-implant diseases-The EFP S3 level clinical practice guideline. J. Clin. Periodontol. 2023, 50 (Suppl. 26), 4-76. https://doi.org/10.1111/jcpe.13823.

4. Cha, J.Y.; Yoon, H.I.; Yeo, I.S.; Huh, K.H.; Han, J.S. Peri-implant bone loss measurement using a region-based convolutional neural network on dental periapical radiographs. J. Clin. Med. 2021, 10, 1009. https://doi.org/10.3390/jcm10051009.

5. Vera, M.; Gómez-Silva, M.J.; Vera, V.; López-González, C.I.; Aliaga, I.; Gascó, E.; Vera-González, V.; Pedrera-Canal, M.; Besada-Portas, E.; Pajares, G. Artificial intelligence techniques for automatic detection of peri-implant marginal bone remodeling in intraoral radiographs. J. Digit. Imaging 2023, 36, 2259-2277. https://doi.org/10.1007/s10278-023-00880-3.

6. Hung, K.F.; Ai, Q.Y.H.; Leung, Y.Y.; Yeung, A.W.K. Potential and impact of artificial intelligence algorithms in dento-maxillofacial radiology. Clin. Oral Investig. 2022, 26, 5535-5555. https://doi.org/10.1007/s00784-022-04477-y.

7. Farajollahi, M.; Safarian, M.S.; Hatami, M.; Esmaeil Nejad, A.; Peters, O.A. Applying artificial intelligence to detect and analyse oral and maxillofacial bone loss: A scoping review. Aust. Endod. J. 2023, 49, 720-734. https://doi.org/10.1111/aej.12775.

8. Liu, M.; Wang, S.; Chen, H.; Liu, Y. A pilot study of a deep learning approach to detect marginal bone loss around implants. BMC Oral Health 2022, 22, 11. https://doi.org/10.1186/s12903-021-02035-8.

9. Jang, W.S.; Kim, S.; Yun, P.S.; Jang, H.S.; Seong, Y.W.; Yang, H.S.; Chang, J.S. Accurate detection for dental implant and peri-implant tissue by transfer learning of Faster R-CNN: A diagnostic accuracy study. BMC Oral Health 2022, 22, 591. https://doi.org/10.1186/s12903-022-02539-x.

10. Lee, W.F.; Day, M.Y.; Fang, C.Y.; Nataraj, V.; Wen, S.C.; Chang, W.J.; Teng, N.C. Establishing a novel deep learning model for detecting peri-implantitis. J. Dent. Sci. 2024, 19, 1165-1173. https://doi.org/10.1016/j.jds.2023.11.017.

11. Mao, Y.C.; Chen, C.A.; Lin, Y.J.; Chang, Y.J.; Wei, S.T.; Chen, S.L.; Chen, T.Y.; Li, K.C.; Tu, W.C.; Abu, P.A.R. Automated assessment of peri-implant disease severity by deep learning and image processing in periapical radiographs. J. Dent. Sci. 2026, 21, 963-972. https://doi.org/10.1016/j.jds.2025.08.015.

12. Lee, J.H.; Kim, Y.T.; Schwendicke, F. Performance of artificial intelligence-based diagnosis and classification of peri-implantitis compared with periodontal surgeon assessment: A pilot study of panoramic radiograph analysis. J. Periodontal Implant Sci. 2025, 55, 436-446. https://doi.org/10.5051/jpis.2500280014.

13. Bonfanti-Gris, M.; Ruales, E.; Salido, M.P.; Martinez-Rus, F.; Özcan, M.; Pradies, G. Artificial intelligence for dental implant classification and peri-implant pathology identification in 2D radiographs: A systematic review. J. Dent. 2025, 153, 105533. https://doi.org/10.1016/j.jdent.2024.105533.

14. Mugri, M.H. Accuracy of artificial intelligence models in detecting peri-implant bone loss: A systematic review. Diagnostics 2025, 15, 655. https://doi.org/10.3390/diagnostics15060655.

15. Atieh, M.A.; Shah, M.; Hakam, A.; AlAli, F.; AlSayed, S.; Tawse-Smith, A.; Alsabeeha, N.H.M. Diagnostic accuracy of deep learning models in detecting peri-implant marginal bone loss: A systematic review and meta-analysis. Clin. Oral Implants Res. 2026, 37, 385-401. https://doi.org/10.1111/clr.70090.

16. Su, Y.S.; Chen, C.A.; Chang, L.H.; Su, H.K.; Sung, C.Y.; Lin, Y.J.; Chen, S.L. Automated detection of marginal bone loss levels in implant brands using deep learning on periapical radiographs. J. Formos. Med. Assoc. 2026, ahead of print. https://doi.org/10.1016/j.jfma.2026.06.011.

17. Efeoglu, S.; Karahan, E.; Gokdeniz, S.T.; Incebeyaz, B.; Gonuldas, F.; Aksoy, S.; et al. Radiomics analysis of panoramic radiographs using machine learning for the detection of peri-implantitis. BMC Oral Health 2026, 26, 964. https://doi.org/10.1186/s12903-026-08190-0.

18. Adnan, N.; Ahmed, S.M.F.; Nooruddin, A.; Sadiq, A.; Khalid, A.; Haseeb, M.; Ghori, M.H.; Umer, F. A web-based deep learning cascade for automated detection and quantification of marginal bone loss. BMC Oral Health 2026, 26, 590. https://doi.org/10.1186/s12903-026-07962-y.

19. Madani, Z.; Bashizadeh Fakhar, H. Detection and classification of peri-implant marginal bone loss in cone-beam computed tomography using a deep learning approach. Clin. Exp. Dent. Res. 2026, 12, e70308. https://doi.org/10.1002/cre2.70308.

20. Gao, T.; Tan, R.; Tian, S.K.; Chen, H.; Wang, H.R.; Xu, J.G. An auxiliary measurement method for marginal bone loss around dental implants based on periapical radiographs. Chin. J. Dent. Res. 2025, 28, 261–276. https://doi.org/10.3290/j.cjdr.b6745502.

21. Khan, G.S.; Aaliyah, S.; Pal, S.; Tiwari, R.; M, S.; Dixit, H.; M C, P.; Ali, T. Artificial intelligence for dental implant classification and peri-implant disease detection: A clinical study. Bioinformation 2025, 21, 2361–2364. https://doi.org/10.6026/973206300212361.

22. Kibcak, E.; Buhara, O.; Temelci, A.; Akkaya, N.; Ünsal, G.; Minervini, G. Deep learning-driven segmentation of dental implants and peri-implantitis detection in orthopantomographs: A novel diagnostic tool. J. Evid. Based Dent. Pract. 2025, 25, 102058. https://doi.org/10.1016/j.jebdp.2024.102058.

23. Chen, Y.C.; Chen, M.Y.; Chen, T.Y.; Chan, M.L.; Huang, Y.Y.; Liu, Y.L.; Lee, P.T.; Lin, G.J.; Li, T.F.; Chen, C.A.; et al. Improving dental implant outcomes: CNN-based system accurately measures degree of peri-implantitis damage on periapical film. Bioengineering 2023, 10, 640. https://doi.org/10.3390/bioengineering10060640.

24. Zhang, C.; Fan, L.; Zhang, S.; Zhao, J.; Gu, Y. Deep learning based dental implant failure prediction from periapical and panoramic films. Quant. Imaging Med. Surg. 2023, 13, 935–945. https://doi.org/10.21037/qims-22-457.

25. Zhang, H.; Shan, J.; Zhang, P.; Chen, X.; Jiang, H. Trabeculae microstructure parameters serve as effective predictors for marginal bone loss of dental implant in the mandible. Sci. Rep. 2020, 10, 18437. https://doi.org/10.1038/s41598-020-75563-y.