Diagnostic Accuracy of Artificial Intelligence- Based Models in Periodontitis: A Systematic Review Meta-Analysis

Dr. S. Deepa [1], Dr. Bhumimathan Swamikannu [2], Dr. R. Narasimha Raghavan [3], Dr. Keshavan [4], Dr. Karthikeyan Pandiyambakkam Rajendran[5], Gianna Sarah Theodore[6], Arumugam Shanmugavelayutham[7], Arunkumar Shanmugavelayutham[8], Dr. Raghavendran S[9], Dr. Divya Sneha M[10], Dr. Nandhini J[11]

[1] Research Scholar, Department of Dental Science, Bharath Institute of Higher, Education and Research, Chennai, Tamil Nadu, India. Email: deepasubramaniam09@gmail.com
[2] Registrar, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India. Email: bhumi.sbdch@gmail.com
[3] Director, RNR Research Services, Chennai, Tamil Nadu, India. Email: director@rnrresearch.co.in
[4] Professor, Thai Mookambigai Dental College & Hospital, Chennai, Tamil Nadu, India. Email: keshavan84@gmail.com
[5] Research Coordinator, Schizophrenia Foundation(I), Chennai, Tamil Nadu, India. Email: dentalprk@gmail.com
[6] Research Intern, Schizophrenia Research Foundation (I), Chennai, Tamil Nadu, India. Email: gia.teo323@gmail.com
[7] Founder, docFacet. Email: connectwithaaru@gmail.com
[8] ML Engineer, docFacet. Email: arunmeetsu123@gmail.com
[9] Intern, Sri Venkateswara Dental College and Hospital, Chennai, Tamil Nadu, India. Email: raghavendransathya25@gmail.com
[10] Intern, Sri Venkateswara Dental College and Hospital, Chennai, Tamil Nadu, India. Email: mdsdent07@gmail.com
[11] Intern, Sri Venkateswara Dental College and Hospital, Chennai, Tamil Nadu, India. Email: nandhinijawahar23@gmail.com

KEYWORDS

ABSTRACT

Introduction: Periodontitis is an irreversible disease caused by host-microbe interactions leading to the destruction of tooth-supporting structures. Its complex aetiology makes early diagnosis, staging, and treatment planning challenging but crucial to prevent disease progression. Artificial intelligence (AI) models, particularly Convolutional Neural Networks (CNNs), analyze complex variables, identify patterns, and make accurate predictions. Their use in periodontitis diagnosis can enhance diagnostic accuracy, reduce human error, and provide consistent results.

Objectives: This review evaluates the current landscape of AI applications in diagnosing periodontitis, with a focus on CNN-based models used directly or through proxy indicators.

Methods: A systematic literature search was conducted in PubMed, Web of Science, CINAHL, Embase, Cochrane Library, and ClinicalTrials.gov up to December 2019. Included studies assessed the diagnostic accuracy of AI models for periodontitis using cross-sectional, case-control, or cohort designs. Aggressive periodontitis cases were excluded. Risk of bias was assessed using the PROBAST tool, and results are presented as a narrative synthesis.

Results: AI models, particularly CNNs, demonstrated high diagnostic accuracy for periodontal bone loss using radiographic evidence, often surpassing expert performance. Models like DenseNet and U-Net excelled in segmentation and classification. Challenges included poor image quality, imbalanced datasets, and reliance on proxy indicators, highlighting the need for multivariable approaches.

Discussion: AI shows promise in standardizing and scaling periodontitis diagnosis, addressing manpower shortages, and improving outcomes. However, future research should focus on integrating multivariable diagnostic approaches and refining model interpretability for clinical applicability.

1. Introduction

Periodontitis is a disease characterized by inflammation caused by microbes and mediated by the host's immune response. This leads to the loss of the structures that support the teeth (Tonetti et al., 2018). Some typical characteristics of periodontitis include gingival inflammation, clinical attachment loss, radiographic evidence of alveolar bone loss, sites with deep probing depths, tooth mobility, bleeding upon probing, and pathologic migration of teeth (Kwon et al., 2024) Moreover, it has been associated with multiple chronic disorders, including cardiovascular disease, Alzheimer's disease, Diabetes mellitus, rheumatoid arthritis, nonalcoholic fatty liver disease and even certain cancers (Hajishengallis & Chavakis, 2021; Winning & Linden, 2015). Early and accurate diagnosis is crucial for effective management and prevention of disease progression. Traditional diagnostic methods, such as clinical examination and radiographic analysis, while effective, have limitations including inter-examiner variability and potential for subjective bias.

In general, diagnostic errors tend to reflect a sizable portion of all medical errors (Mamede et al., 2007) Fortunately, the rise of Artificial Intelligence (AI) and machine learning models has shown great promise in various medical fields, including dentistry. AI-based models can analyze complex datasets, identify patterns, and make predictions with a high degree of accuracy Employing such models has the potential to allow dental professionals to rise to this challenge by enhancing diagnostic accuracy, reducing human error, and providing consistent and reproducible results. Further, due to the relatively quick output of these systems, using optimised AI diagnostic aids could allow more patients to receive the medical attention they require especially in regions where caseloads on the resident dental professionals are heavy. This is especially true in low-income countries where only 1.4% of the total number of dentists work (World Health Organization, 2022).

The scalability of these models also appears promising as they can be deployed across various clinical settings while providing uniform diagnostic support and further improving the standard of care in diverse healthcare environments. Finally, a thorough investigation to determine the presence of the disease would require the identification and measurement of PD, CAL, bleeding on probing (BOP), plaque index, furcation involvement, suppuration, mobility, occlusal trauma, open contact areas, and radiographic interpretation of bone levels (Shaddox & Walker, 2010). Once again, due to its ability to integrate multiple data sources, the utility of AI diagnostic models cannot be ignored. The capability of AI models to provide accurate and rapid diagnoses, the ability to learn continuously and integrate various aspects of patient data and its scalability positions them as a highly efficient and critical tool in the future of the diagnostic process.

Convolutional Neural Networks and Support Vector Models are commonly used machine learning models in periodontitis diagnostics (Scott et al., 2023). CNNs are used when the primary task involves pattern and image recognition (O'Shea & Nash, 2015). As evidence of bone loss identified via radiographs is one of the primary prerequisites for a periodontitis diagnosis, it stands to reason that using a CNN-based diagnostic model would be a foremost choice. Support Vector Machine algorithms in conjunction with patient data were also used to aid the diagnoses of periodontitis.

The last 5 years have seen a growing body of research on the diagnostic accuracy of AI-based models in periodontitis, however, the results remain unclear This must be addressed by a systematic evaluation of the available evidence to determine the effectiveness and reliability of these models. This comprehensive review aims to provide valuable insights into the current state of Artificial Intelligence in diagnosing periodontitis, summarize the existing models that have undergone testing, identify gaps in the literature, and offer guidance for future research.

2. Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Checklist were followed in the reporting of this review (Page et al., 2021).

Review Questions

1. What is the current landscape of artificial intelligence applications in diagnosing periodontitis?
2. What are the input parameters (variables) commonly used in AI models for the detection, staging, and prognosis of periodontitis?
3. How are Convolutional Neural Network (CNN)-based models utilized in diagnosing periodontitis, either directly or indirectly through the identification of proxy indicators?

Inclusion criteria

Diagnostic accuracy studies, including cross-sectional, case-control, and cohort studies that assess the performance of AI models in diagnosing periodontitis are included in this review. Participants with Chronic Generalised or Chronic Localised Periodontitis were included.

Exclusion criteria

Cases of aggressive Periodontitis have been excluded. Studies that were not in English or those without English translations were excluded from this review. Records that were not primary research articles, including reviews, letters, personal opinions, book chapters, conference abstracts, and those with unavailable full texts were excluded.

Information sources and search strategy

Only those studies published in English in peer-reviewed journals are included in this view. A systematic literature search was performed in PubMed, Web of Science, CINAHL, Embase, Cochrane Library, and clinicaltrial.gov through December 2019. Two independent authors conducted a literature search using a structured search strategy. To begin with, free text searching of the keywords and their synonyms was carried out using appropriate truncation, and proximity searching. A second search was conducted for key concepts using corresponding subject headings in each database. The final search was carried out where the individual search results were combined using appropriate Boolean operators. Supplementarily, the references of selected articles were screened to find additional records that may have failed to appear in the database searches.

Study selection

All the citations, along with the title and abstract that had been retrieved using the search strategy, were imported to a specified endnote library, and a final list of studies was screened for inclusion in the study. Following this, the citations were imported into Rayyan and de-duplication was carried out. Two independent reviewers screened the records by assessing the title and abstract to shortlist the studies likely to satisfy the review's inclusion criteria. Any disagreement between the reviewers over the eligibility of particular studies was resolved through discussion with a third reviewer. Full-text articles for all the shortlisted studies were sought out and a meticulous screening was done adhering to the proposed inclusion and exclusion criteria. Studies not satisfying inclusion criteria and those with unavailable full-texts were excluded at this step.

Data collection process

Data extraction was facilitated using a data extraction form created apriori. Data items extracted included general study identification information and study characteristics. The focal data items were collected according to the Population, Intervention, Comparator and Outcome (PICO) format:

Population:

- Dataset
- Data source
- Inclusion and exclusion criteria

Intervention:

- Input parameters
- Artificial Intelligence model
- A brief description of the AI model provided by each paper
- Annotation methods used if relevant

Comparator/Ground truth

Outcome

- Study outcome
- Evaluation metrics
- Comments on model performance

Risk of Bias of Individual Studies

PROBAST, a prediction model study Risk Of Bias Assessment Tool, was used as the primary tool for quality and risk of bias assessment, considering the focus of this review and the nature of studies included (Moons et al., 2019). Two independent reviewers rated each study as per the requirements of the scale after which the ratings were compared. Discrepancies in ratings between the two reviewers were settled after a discussion.

Data synthesis

Due to the heterogeneous nature of the evaluation metrics used across studies and the high variation of artificial intelligence models used, it was not possible to conduct a meta-analysis. Therefore, the results of the review are presented as a narrative synthesis.

3. Results

Study selection

4173 records were retrieved from six databases on 17th February 2024. After de-duplication, 3553 records remained. These records were screened by title and abstract and as a result, 3535 records were excluded and the remaining 18 were sought for retrieval. All but one of these studies were successfully retrieved and underwent full-text screening. 12 studies met the inclusion criteria and were included for review. An additional 11 studies were handpicked from bibliography screening resulting in a total of 23 included studies.

The flow diagram outlining the process of study selection is presented in Fig 1 .

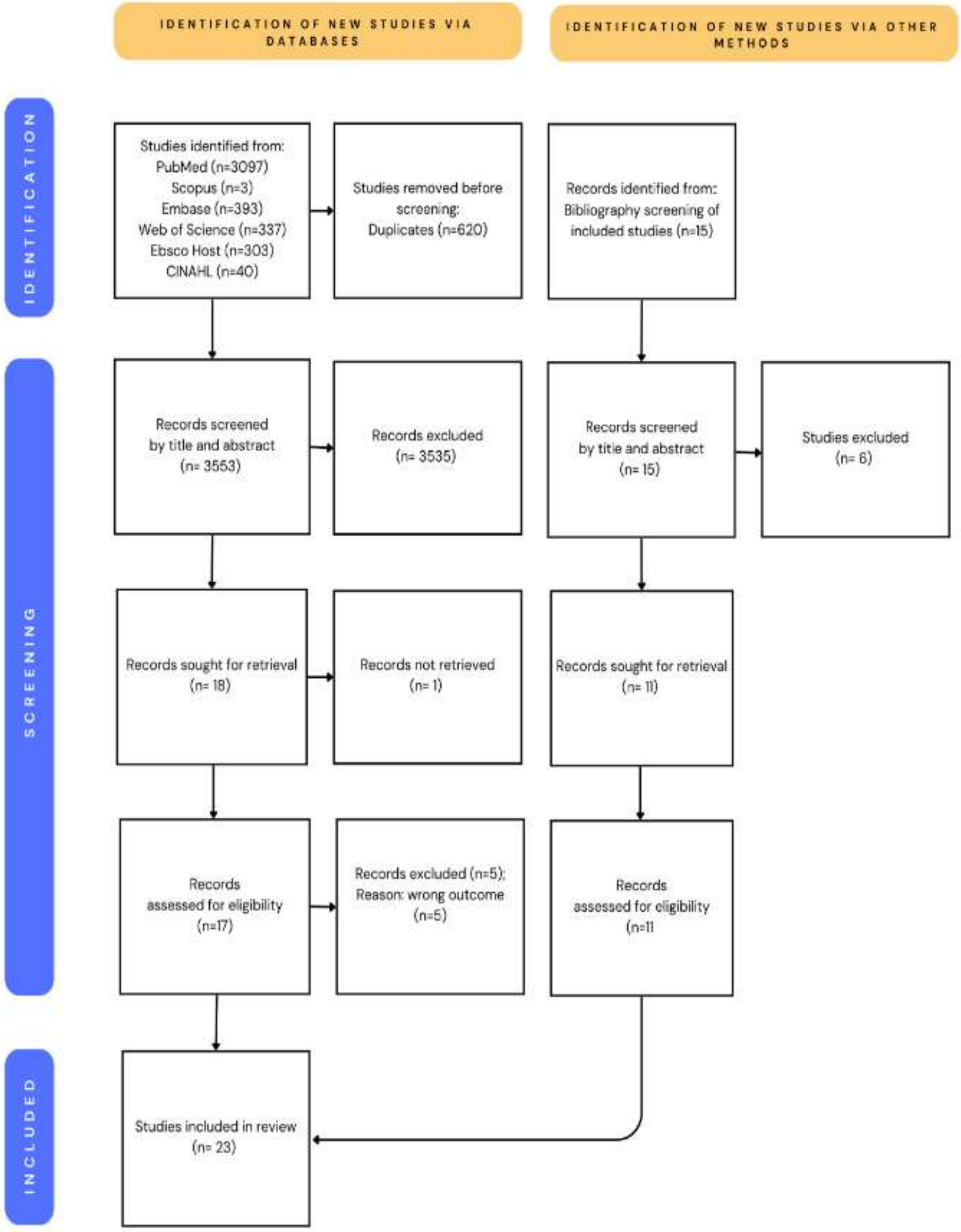
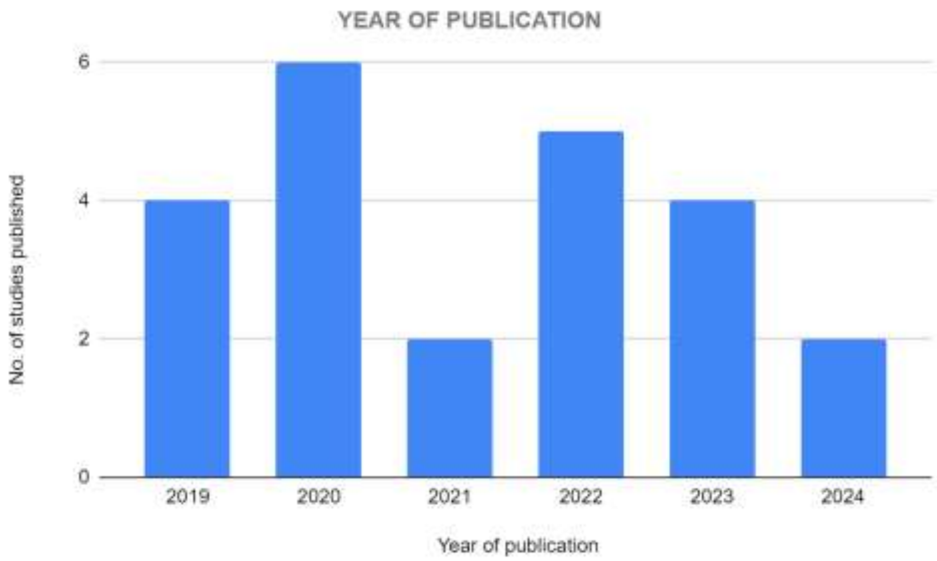


Fig 1 PRISMA Flow diagram outlining the process of study identification and selection.

Study characteristics

All the selected studies were published between 2018 and 2024. These studies took place across several countries - USA, UK, China, Republic of Korea, Indonesia, Chile, Turkey, Denmark, Iran and Poland. Each study tested the diagnostic ability of Artificial Intelligence systems in the case of Periodontitis with a few studies focusing on stage classification of periodontitis. Most studies relied on Panoramic radiographs as the input variable for diagnosis and stage classification. These datasets ranged from 20 to 12,179 PARS.



Risk of bias within studies

The risk of bias was assessed using the Prediction model study Risk of Bias Assessment Tool (PROBAST). 16 studies were rated to have a low risk of bias, four studies showed an unclear risk of bias and one study showed potential for a high risk of bias due to the lack of internal validation. It should be noted that for item "3.3 Were predictors excluded from the outcome definition?". Further, the item "Describe missing data on predictors and outcomes as well as methods used for missing data" from Domain 4, studies that utilised complete retrospective datasets carried the response "Not applicable". Concern for Applicability was evaluated as being "low" for all studies. Refer Table 1 for overall appraisal of the Risk of Bias evaluation.

Table 1 represents the domain wise PROBAST risk of bias evaluations for the included studies

Kim_2019	LOW	LOW	LOW	LOW	LOW	LOW
Krois_2019	LOW	LOW	LOW	LOW	LOW	LOW
Joo_2019	LOW	LOW	LOW	LOW	LOW	LOW
Huang_2019	LOW	LOW	LOW	UNCLEAR	UNCLEAR	LOW
Chang_2020	LOW	LOW	LOW	LOW	LOW	LOW
Bayrakdar_2020	LOW	LOW	LOW	LOW	LOW	LOW
Li_2020	LOW	LOW	LOW	LOW	LOW	LOW
Farhadian_2020	LOW	LOW	LOW	LOW	LOW	LOW
Moran_2020	LOW	LOW	LOW	LOW	LOW	LOW
Eun-Hye Kim_2020	LOW	LOW	LOW	LOW	LOW	LOW
Danks_2021	LOW	LOW	LOW	LOW	LOW	LOW
Kabir_2021	LOW	LOW	LOW	LOW	LOW	LOW

Selviani_2022	LOW	LOW	LOW	LOW	LOW	LOW
Chun-Teh Lee_2022	LOW	LOW	LOW	LOW	LOW	LOW
Jiang_2022	LOW	LOW	LOW	LOW	UNCLEAR	LOW
Chang_2022	LOW	LOW	LOW	LOW	HIGH	LOW
Shon_2022	LOW	LOW	LOW	LOW	LOW	LOW
Kong_2023	LOW	LOW	LOW	LOW	UNCLEAR	LOW
Liu_2023	LOW	LOW	LOW	LOW	UNCLEAR	LOW
Zhu_2023	LOW	LOW	LOW	LOW	LOW	LOW
Enevold_2023	LOW	LOW	LOW	LOW	LOW	LOW
Ayyildiz_2024	LOW	LOW	LOW	LOW	LOW	LOW
Mardini_2024	LOW	LOW	LOW	LOW	LOW	LOW

Summary of Evidence

Table 2 represents the **Summary of Evidence** gathered from the included studies.

Table 2- Summary of Evidence

Study ID	Year of publication	Study design	Complete Dataset	Dataset for Training	Dataset for Validation	Dataset for Testing	Inclusion criteria	Exclusion criteria	Artificial Intelligence Model	Input variables	Comparator	Outcome
Kim_2019	2019	diagnostic accuracy study	12,179 PARS	11189 PARS	190 PARS	800 PARS			DCNN (named DeNTNet)	panoramic radiographs	assessments of dental clinicians	PBL Detection
Krois_2019	2019	diagnostic accuracy study	2001 manually cropped image segments, each focusing on one particular tooth, from 85 randomly chosen digital panoramic dental radiographs	NA	NA	NA	Only radiographs from dentate individuals were included	Those images that heavily overlapped with the vertebrae and did not allow any kind of assessment	CNN-based model	panoramic radiographs	assessments by six dental practitioners (specialists in operative dentistry and periodontology (n=1), specialized in endodontics (n=1) or general dentists (n=4))	PBL Detection
Joo_2019	2019	diagnostic accuracy study	1400 images of periodontal tissue were taken; and in order to produce the output "not periodontal image" 594 pieces of image data not related to periodontal are used	1643 images	200 images	NA	NA	NA	CNN-based model	photographed periodontal images	assessment by expert?	Periodontitis diagnosis
Huang_2019	2019	comparative diagnostic accuracy study	Twenty-five (25) severe generalized periodontitis patients	NA	NA	NA	NA	History of hypertension, diabetes mellitus, coronary disease	Support Vector Machine, Random Forest, k-Nearest Neighbor, Linear Discriminant Analysis, andClassification and Regression Trees.	Antibody arrays obtained from GCF samples	diagnosis of professionals?	Prediction Of Severe Periodontal Disease
Chang_2020	2020	diagnostic accuracy study	340 PARS (To evaluate the classification performance forstaging the periodontitis, we used ten panoramic radiographs not used for detection.)	90% of dataset	NA	10% of dataset	NA	images of patients with primary or mixed dentition	modified CNN from the Mask R-CNN based on a feature pyramid network (FPN) and a ResNet101 backbone	panoramic radiographs	assessments by radiologists	PBL Detection AND Periodontitis Staging
Bayrakdar_2020	2020	diagnostic accuracy study	2276 panoramic radiography images	1856 images	n = 210	n = 210	NA	Panoramic radiographs of patients with a large number of teeth missing (i.e., patients with less than 20 teeth), patients younger than 18 years, and patients with excessive crown-damaged teeth as well as images with artifacts or partial/severe distortion were excluded from the dataset	CNN-based model	panoramic radiographs	assessment by an oral and maxillofacial radiologist and a periodontologist (with at least 5 years of professional experience)	Alveolar Bone Loss Detection (however conclusions mention PBL?)
Li_2020	2020	diagnostic accuracy study	302 digitized panoramic radiographs and 67 high resolution panoramic radiographs	270 images		28 images and 67 high resolution panoramic radiographs	NA	NA	Mask R-CNN based model (called Deetal-Perio)	panoramic radiographs	assessments by 3 professional dentists	Alveolar Bone Loss detection

Study ID	Year of publication	Study design	Complete Dataset	Dataset for Training	Dataset for Validation	Dataset for Testing	Inclusion criteria	Exclusion criteria	Artificial Intelligence Model	Input variables	Comparator	Outcome
Farhadian_2020	2020	diagnostic accuracy study	records of 300 patients	NA	NA	NA	NA	NA	SVM based model	Medical records	assessment by professional	Classification of gingivitis, localized periodontitis and generalized periodontitis
Moran_2020	2020	diagnostic accuracy study	6179 interproximal regions extracted from 467 different periapical radiographs	3022 images composed the training and validation sets					CNN- based models	Interproximal surfaces from X-rays	assessments by experienced dentists, and one of them is a specialist in oral radiology	Classification of periodontal bone destruction
Eun-Hye Kim_2020	2020	comparative diagnostic accuracy study	692 subjects (144 periodontally healthy controls and 548 generalized chronic periodontitis patients)	NA	45 additional mouthwash samples from Pusan National University Dental Hospital	NA	NA	-Those who received periodontal treatment within the past 6 months; -Women who were pregnant or breastfeeding; -Those who refused to sign the informed consent form	Four multivariate machine learning models- feed-forward neural network, random forest, support vector machines with linear kernel, and regularized logistic regression in R caret package (version 6.0-84)	Salivary Bacterial Copy Number (using mouthwash samples and DNA extraction)	assessments by two experienced periodontists	Prediction of Chronic Periodontitis Severity
Danks_2021	2021	diagnostic accuracy study	340 fully (anonymised) periapical radiographs	NA	NA	NA	NA	NA	symmetric hourglass network	periapical radiographs	baseline ResNet-based regression model without the proposed ISM model additions (i.e. no pixel non- maturation, ISM or transfer learning), a symmetric hourglass without additions, a network with an asymmetric hourglass architecture from with and without model additions and a stacked hourglass network	PBL Detection and Measurement
Kabir_2021	2021	diagnostic accuracy study	700 periapical X-rays	70%	10%	20%			HYNETS- the U-Net architecture and FCN architecture	periapical radiographs	assessment by three independent examiners: aboard-certified periodontist professor, a board-certified clinicalperiodontist, and one resident in the periodontics program	Periodontitis Staging based on RBL percentage and Periodontitis Diagnosis
Sehrawi_2022	2022	diagnostic accuracy study	23 Medical records	NA	NA	NA	NA	NA	Expert Decision Model	Medical records	assessment by doctors	Periodontitis diagnosis
Chun-Teh Lee_2022	2022	diagnostic accuracy study	693 periapical radiographic images from randomly selected 37 periodontitis patients further evaluated on 644 additional periapical imagesfrom randomly selected 40 cases	70%	10%	20%	NA	NA	U-Net with CNN, ResNet-34, and ResNet-50 encoder	periapical radiographs	assessment by three examiners	alveolar bone level measurement, RBL stage assignment, and preliminary periodontal diagnosis

Study ID	Year of publication	Study design	Complete Dataset	Dataset for Training	Dataset for Validation	Dataset for Testing	Inclusion criteria	Exclusion criteria	Artificial Intelligence Model	Input variables	Comparator	Outcome
Jiang_2022	2022	diagnostic accuracy study	640 panoramic radiographs	80%	NA	20%	Misplaced, overlapping teeth in thedataset were also included after careful identification andlabelling	Images of patients with primary or mixed dentition	Other network+CSPDarknet +spatial pyramid pooling module+path aggregation network+YOLO Head	panoramic radiographs	3 general dentists, all working in theAffiliated Stomatology Hospital, Zhejiang University School of Medicine for at least 3 years	detection and staging of radiographic periodontalalveolar bone loss
Chang_2022	2022	diagnostic accuracy study	8316 interproximal surfaces from 1836 X-rays were included	NA	NA	NA	-Patients with electronic dental records from the school between 01/01/2010 and 12/31/2020 were screened. -Only patients with full mouth standardized radiographs from the school were included in the study	- Radiographs that were not inthe standard format were excluded from the study -Teeth were excluded if the radiograph was lacking diagnostic quality: (1) improper angulationcausing severe elongation, foreshortening, or overlapping of anatomic landmarks inhibiting RBL classification; or (2) anatomic landmarks not captured in the image, such as root apex of a tooth.	CNN-based model	interproximal surfaces from X-rays	3 calibrated periodontists	Measurement and Classification of RBL
Shen_2022	2022	diagnostic accuracy study	977 PARS – CBMUH dataset; 4010 PARS – AIRUB dataset	U-Net model trained on 70% of CBMUH dataset YOLOv5 model trained on 70% of AIRub	YOLOv5-model- 30% of AIRub dataset	U-Net model tested with 30% of CBMUH dataset		Those with missing values and invalid imageproperties	Integrated framework of U-Net and YOLOv5.	panoramic radiographs	assessment by dental specialists	Periodontitis Staging
Kong_2023	2023	diagnostic accuracy study	1747 panoramic images	NA	NA	NA	NA	NA	CNN-based model (called PDCNN)	panoramic radiographs	benchmark networks – Retinanet, YOLO-v4, CenterNet,and Faster R-CNN	PBL Detection
Liu_2023	2023	diagnostic accuracy study	1275 PARS	1276 PARs		272 PARs	- Individuals were adults (>18 years old); - Individuals had information from clinical examinations and independent risk factors for periodontitis, including age, sex, hereditary periodontitis disease (HPD), smoking, diabetes, and hypertension; and - Individuals had secondary dentition throughout their whole mouth	- PARs with many metal artifacts, severely blur and noise, and those in which the affected area could not be selected accurately because of PBL determination. - (If the three periodontists' diagnoses were notcompletely consistent, the individual was excluded. When there were discrepancies in the diagnostic results after reading the PARs and clinical examination information, the individual was also excluded.)	CNN-based model (called PARS-CNN)	panoramic radiographs	Three blinded periodontists from the Department ofStomatology of the Second Affiliated Hospital with >5 yearsof experience independently assigned a diagnosis usingPARs and previous clinical examination information	

Study ID	Year of publication	Study design	Complete Dataset	Dataset for Training	Dataset for Validation	Dataset for Testing	Inclusion criteria	Exclusion criteria	Artificial Intelligence Model	Input variables	Comparator	Outcome
Zhu_2023	2023	comparative diagnostic accuracy study	602 qualified oral dental X-ray images	322 images (horizontal flipping and rotation were used in this study to expand the oral dental X-ray images in the training set to 1955 images.)		140 images	NA		PSPNet, U-Net, and Dense U-Net	oral dental X-ray images	assessments of six periodontists with medical practitioner qualifications with 8 years of experience	
[illegible]_2023	2023	diagnostic accuracy study	1476 CAMB participants 3580 DANHES participants	1034 participants	DANHES dtaset was used for validation	442 participants		Participants with less than two teeth were excluded	Prediction models constructed using the xgBoost, random forests, and partial least squares algorithms using the 'Caret' package (Kuhn, 2008) in RStudio (version 1.4.1717, RStudio Team) running R (version 4.1.0, R Foundation for Statistical Computing Vienna, Austria).	available demographic data and the subjects' questionnaire responses	CAMB —> Oral examinations carried out by two trained dental hygienists, and participants had a full-mouth periodontal examination including PPD, bleeding on probing and indirect measurement of CAL on six sites in every tooth. DANHES —> Oral examinations were conducted by three dental hygienists trained by two experienced clinical examiner (half-mouth registrations of PPD, bleeding on probing and indirect measurements of CAL for six sites on each tooth in one randomly selected upper quadrant and one randomly selected lower quadrant.	Prediction of Stage III/IV Periodontitis
Ayyildiz_2024	2024	comparative diagnostic accuracy study	The test datasets consisted of an equal number of images from healthy (168), Stage1/2 (168), and Stage3/4(300) patients.	The train datasets included images from healthy patients (1412), Stage1/2 (1474), and Stage3/4 (1502)patients.	NA	The test datasets consisted of an equal number of images from healthy (168), Stage1/2 (168), and Stage3/4(168) patients	NA	NA	T1-based CNN model	panoramic radiographs	2 calibrated professionals - a 3-year periodontology specialty student and a periodontist with 12 years of experience.	Periodontitis Diagnosis
[illegible]_2024	2024	diagnostic accuracy study	1005 PANS (giving rise to 2010 images)	1970 images		40 images	No other demographic characteristics were considered or recorded from the population (i.e. age, gender, ethnicity or socio-economic status).	Radiographs of edentulous patients and those from patients with temporary/mixed dentures.	Statistical Inference + DCNN + PBL algorithm	panoramic radiographs	diagnosis of dental professionals	PBL Detection and Measurement

Result of Data Synthesis

The studies selected for this review primarily focus on detecting proxy indicators, such as periodontal bone loss (PBL), as a means to assess and diagnose periodontitis. Given the scarcity of AI models with the capability to directly diagnose periodontitis, researchers have instead concentrated on detecting and analyzing radiographic evidence of Periodontal bone loss, Radiographic bone loss and Alveolar bone loss and bone level.

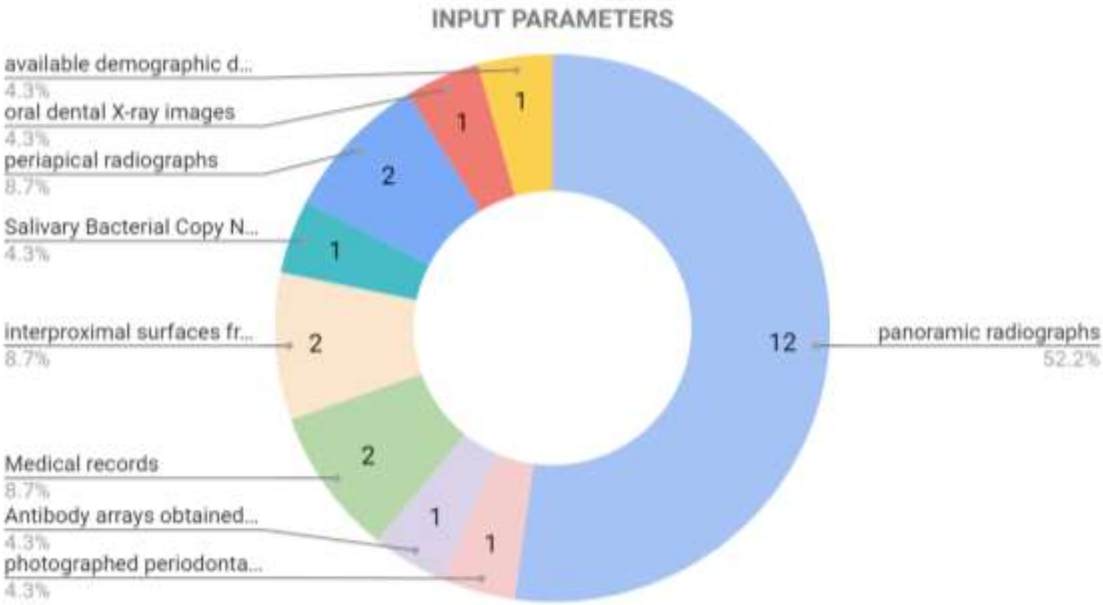
Inclusion and Exclusion criteria

Common reasons for exclusion included:

- patients having primary or mixed dentitions.
- Severe distortion in the radiograph resulting in poor readability (caused by metal artefacts, excessive blur or noise etc.)

Input Parameters

The most common input parameter was panoramic radiographs (n=12). The other input parameters included oral dental X-rays (n=1), interproximal surfaces from X-rays (n=2), patient medical records (n=2), periapical radiographs (n=2) demographic data and subjects' questionnaire responses (n=1), photographed periodontal tissue (n=1), salivary bacterial copy numbers (n=1) and antibody arrays (n=1).



Datasets

The datasets in all the included studies were collected and analysed retrospectively. In general, datasets were collected either from existing hospital records (after necessary steps were taken to anonymise the data /obtain informed consent) Institute archives or from open platform databases. Complete datasets ranged from as large as 12,179 PARS (Kim) to 23 Medical records (Selviani). However, in most studies involving PARS, the images were subject to multiple augmentation efforts, thus expanding their final datasets (cite all studies that underwent augmentation).

Annotation Methods and Annotators

All studies involving panoramic radiographs underwent annotation carried out by various individuals with credentials ranging from undergraduate (Mardini_2024) and postgraduate students with a specialisation in periodontology (Danks_2021) to experienced dental clinicians (Kim_2019), doctors (Kong_2023) and oral and maxillofacial (OMF) radiologists as well as periodontists(Zhu_2023)(Chun-Teh Lee_2022)(Jiang_2022). Common tools used for annotation include LabelBox (Labelbox Inc, CA) (Chang_2020) (Zhu_2023), LabelMe (Chang_2022) VGG Image Annotator (VIA) (Danks_2021)(Kong_2023), Label-Studio® (Mardini_2024). A single study used the Computer Vision Annotation Tool (CVAT) (Chun-Teh Lee_2022).

Performance of AI Models in Diagnosing Periodontitis

(I) Convolutional Neural Networks based models (n=9)

Six studies investigated the accuracy of CNNs and CNN-based models in diagnosing periodontitis. All 5 studies report relatively high accuracy of the models in diagnosing periodontal bone loss. The exception was the model developed by Kong_2023 the function of which was to identify and stage Radiographic Bone Loss. The PDCNN developed a two-stage CNN-based periodontitis detector and achieved an RBL classification accuracy of **0.762**. The detector was compared to and reportedly outperformed 4 other detectors, RetinaNet, YOLO-v4, CenterNet and Faster R-CNN, in the domains of localization accuracy, classification accuracy, and speed.

In terms of the detection of PBL, (Kim_2019) reports that the performance of the model DeNTNet surpassed that of dental clinicians. A modified CNN (from the Mask R-CNN) based on a feature pyramid network (FPN) and a ResNet101 Backbone was developed and tested by (Chang_2020). The ICC value computed i.e**. 0.91 (p<0.01)**, between the model and radiologists' diagnosis, showed high reliability in the diagnosis of periodontal bone loss. Similar positive results were reported in the case of the multi-tasking Inception V3 (Chang_2022) which demonstrated an average accuracy of **87±0.01**. In a comparative study by (Ayyildiz_2024), a DenseNet121 + GAP + mRMR-based SVM model had the highest performance value with an accuracy of **0.907**. The PAR-CNN model developed by Liu_2023 was used in conjunction with gradient-weighted class activation mapping (Grad-CAM) to output a prediction score of 0 to 1 for each PAR. This score showed a significant correlation with the severity of periodontitis. Bayrakdar_2020 Despite the positive results reported, the study by Krois_2019 reports that the mean accuracy of the model (which was developed using the TensorFlow framework and Keras) was not significantly higher than that of the examiners.

CNNs Specialized for Object Detection and Image Segmentation:

Danks_2021 proposed a model to localise dental landmarks and automatically calculate PBL and disease severity stages. The model employed an adjusted symmetric hourglass with ISM model additions with a peak **PCK of 88.9%** for single root teeth outperforming the benchmark networks. However, this was not the case for double and triple-rooted teeth (possibly owing to the smaller respective datasets). Shon_2023 integrated two deep learning algorithms, U-Net and YOLOv5, to develop an approach to detect PBL and CEJ boundaries (by U-Net model) and tooth numbering and length detection (by YOLOv5). The integrated model demonstrated an accuracy of **0.929**. Zhu et al. (2021) also investigated the capabilities of the U-Net, alongside PSPNet and Dense U-Net models, in the context of alveolar bone edge line extraction. The study revealed that the U-Net and PSPNet algorithms may not be well suited for this task whereas the Dense U-Net was capable of accurately segmenting the alveolar bone from oral dental X-rays. Although the difference in accuracy between the model and periodontal experts was insignificant, the model's accuracy evaluation (0.800) was higher than that of general dental practitioners (0.693). The study also highlighted that the time required to read the PARs by the model was significantly shorter than the time taken by both the periodontal experts and the general dental practitioners.

(2) Other models (n=3)

Selviani_2022 developed an expert decision model that inputs patient data including CAL values, tooth loss, bone loss and age. The model uses eight rules to decide the staging and the grade of periodontitis. The study reported an output accuracy of **86%**. Subject demographic data and subject responses to a questionnaire were utilized by (Enevold_2023) to predict Stage 3/4 periodontitis. In this case, xgBoost, partial least squares and random forests algorithms were used to develop three different models. Each used a combination of demographic parameters and questionnaire responses. The model employing both demographic data and questionnaire response performed better than the models using solely demographic or self-reported questionnaire data. However, the study concluded that these models had **limited capabilities** for predicting Stage 3/4 periodontitis.

Mardini_2024 adopted a model which stacked Statistical Inference, DCNN and a PBL algorithm to detect PBL. The study concluded that although the model was **satisfactory** in identifying light to moderate PBL, it **failed to detect severe PBL** from panoramic radiographs.

4. Discussion

This review outlines the current state and progress of Artificial Intelligence models in diagnosing periodontitis. From the synthesis of results, it can be inferred that these studies have demonstrated that AI models have the potential to be powerful tools in the diagnostic process. Not only do they offer the possibility of standardized, reliable and accurate results, but they also significantly reduce the time taken to conduct thorough assessments. The main obstacle faced by AI models, particularly CNN-based models, is that images imputed require significant pre-processing and instances of poor-quality images severely inhibit the performance of the model. A general limitation across models is that the dataset necessary for training must be vast as well as balanced to allow for higher accuracy in diagnosis.

The diagnosis of periodontitis relies on multiple diagnostic criteria such as bleeding on probing, probing pocket depth estimation, determination of furcation involvement, recession, assessment of clinical attachment level and radiographic assessment of bone loss. Therefore, the future direction we propose exploring is one that successfully stacks existing methods that complement clinical examination to directly diagnose the disease rather than rely on a single indicator. This would be a vital development as the diagnosis of this disease requires a multivariable approach. It is possible to get a glimpse of what this may look like in studies such as (Farhadian et al., 2020). Further, it would also be beneficial to pursue the development of models that allow for clear interpretation by experts such as those achieved by (Lee et al., 2022) and (Li et al., 2020).

Convolutional Neural Networks are a subset of neural networks specialised for tasks that require the analysis of visual imagery such as object recognition and image classification. This technique was advanced by Yann LeCun and colleagues. They are most known for successfully developing and training a CNN model (LeNet) for handwriting recognition (LeCun et al., 1998). Some of the most well-known CNN architectures include AlexNet, VGGNet, GoogLeNet/Inception, ResNet and DenseNet. In medicine, CNNs have been tested for tasks such as medical image classification, segmentation, detection and localization, often with positive results when compared with expert performance (Sarvamangala & Kulkarni, 2022). More specifically, CNNs have demonstrated the capacity to significantly impact the field of dental radiography as they may be trained to "read" a variety of intra- and extra-oral radiographs (Brahmi et al., 2024). As seen by the results of this review, this is of high utility in the case of detecting and determining the extent of periodontal bone loss for the diagnosis of periodontitis. Existing models have already proven to reach and even surpass the accuracy of experts and it can be expected that this approach will become more refined with time.

Clinical practice implications and Integration into dental care:

Apart from the obvious advantages of automated diagnosis in terms of speed of diagnosis and objectivity, another implication is efficiency in situations where there is a lack of manpower such as in developing regions. Integration of automated systems also has the potential to lower both patient waiting time and the cost of treatment. Artificial Intelligence diagnostic systems can aid in clinical settings from early detection to implementation of intervention and treatment. Further, AI models can also be harnessed to assist dental professionals in developing treatment plans.

The strength of this review is that, to the best of our knowledge, this is the first of its kind to comprehensively synthesize data concerning the current state of AI models in diagnosing periodontitis. It also sheds light on the popularity of CNN-based models used to identify and stage radiographic bone loss. The review also highlighted gaps in current research and points to future avenues for research.

This review is limited in that the data extracted from the studies could not be subjected to a meta-analysis due to the high heterogeneity in both evaluation metrics used as well as the models used for diagnosis. We aimed to offset this by conducting a thorough although narrative synthesis of the results. There is also a possibility of relevant studies in languages other than English being excluded due to the nature of the inclusion criteria of this review.

5. Conclusions

This review explores the potential of Artificial Intelligence, particularly CNN-based models, in diagnosing periodontitis with accuracy, efficiency, and standardization. While challenges such as image quality, data requirements, and model interpretability persist, AI shows promise in integrating multivariable approaches for comprehensive diagnosis. The findings highlight the transformative role of AI in dental care, addressing manpower shortages and improving patient outcomes while emphasizing the need for further research to refine models and enhance clinical applicability.

PROSPERO Registration

The systematic review was registered as Diagnostic accuracy of Artificial Intelligence- based models in Periodontitis: A Systematic Review Meta-Analysis at PROSPERO vide CRD4202451679.

Ethical Consideration

As this is a review, it is exempted from Ethical clearance.

Conflict of Interest

All authors, who have contributed to this work declare that there is no conflict of interest.

Financial disclosures

None.

References

[1] Brahmi, W., Jdey, I., & Drira, F. (2024). Exploring the Role of Convolutional Neural Networks (CNN) in Dental Radiography Segmentation: A Comprehensive Systematic Literature Review (arXiv:2401.09190). arXiv. http://arxiv.org/abs/2401.09190

[2] Cerda Mardini, D., Cerda Mardini, P., Vicuña Iturriaga, D. P., & Ortuño Borroto, D. R. (2024). "Determining the efficacy of a machine learning model for measuring periodontal bone loss." BMC Oral Health, 24(1), 100. https://doi.org/10.1186/s12903-023-03819-w

[3] Chang, H.-J., Lee, S.-J., Yong, T.-H., Shin, N.-Y., Jang, B.-G., Kim, J.-E., Huh, K.-H., Lee, S.-S., Heo, M.-S., Choi, S.-C., Kim, T.-I., & Yi, W.-J. (2020). Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis. Scientific Reports, 10(1), 7531. https://doi.org/10.1038/s41598-020-64509-z

[4] Chang, J., Chang, M.-F., Angelov, N., Hsu, C.-Y., Meng, H.-W., Sheng, S., Glick, A., Chang, K., He, Y.-R., Lin, Y.-B., Wang, B.-Y., & Ayilavarapu, S. (2022). Application of deep machine learning for the radiographic diagnosis of periodontitis. Clinical Oral Investigations, 26(11), 6629–6637. https://doi.org/10.1007/s00784-022-04617-4

[5] Chen, I. D. S., Yang, C.-M., Chen, M.-J., Chen, M.-C., Weng, R.-M., & Yeh, C.-H. (2023). Deep Learning-Based Recognition of Periodontitis and Dental Caries in Dental X-ray Images. Bioengineering, 10(8), 911. https://doi.org/10.3390/bioengineering10080911

[6] Danks, R. P., Bano, S., Orishko, A., Tan, H. J., Moreno Sancho, F., D'Aiuto, F., & Stoyanov, D. (2021). Automating Periodontal bone loss measurement via dental landmark localisation. International Journal of Computer Assisted Radiology and Surgery, 16(7), 1189–1199. https://doi.org/10.1007/s11548-021-02431-z

[7] Duong, D. Q., Nguyen, K.-C. T., Kaipatur, N. R., Lou, E. H. M., Noga, M., Major, P. W., Punithakumar, K., & Le, L. H. (2019). Fully Automated Segmentation of Alveolar Bone Using Deep Convolutional Neural Networks from Intraoral Ultrasound Images. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 6632–6635. https://doi.org/10.1109/EMBC.2019.8857060

[8] Enevold, C., Nielsen, C. H., Christensen, L. B., Kongstad, J., Fiehn, N. E., Hansen, P. R., Holmstrup, P., Havemose-Poulsen, A., & Damgaard, C. (2023). Suitability of machine learning models for prediction of clinically defined Stage III / IV periodontitis from questionnaires and demographic data in Danish cohorts. Journal of Clinical Periodontology, jcpe.13874. https://doi.org/10.1111/jcpe.13874

[9] Farhadian, M., Shokouhi, P., & Torkzaban, P. (2020). A decision support system based on support vector machine for diagnosis of periodontal disease. BMC Research Notes, 13(1), 337. https://doi.org/10.1186/s13104-020-05180-5

[10] Guler Ayyildiz, B., Karakis, R., Terzioglu, B., & Ozdemir, D. (2024). Comparison of deep learning methods for the radiographic detection of patients with different periodontitis stages. Dentomaxillofacial Radiology, 53(1), 32–42. https://doi.org/10.1093/dmfr/twad003

[11] Hajishengallis, G., & Chavakis, T. (2021). Local and systemic mechanisms linking periodontal disease and inflammatory comorbidities. Nature Reviews Immunology, 21(7), 426–440. https://doi.org/10.1038/s41577-020-00488-6

[12] Huang, W., Wu, J., Mao, Y., Zhu, S., Huang, G. F., Petritis, B., & Huang, R. (2020). Developing a periodontal disease antibody array for the prediction of severe periodontal disease using machine learning classifiers. Journal of Periodontology, 91(2), 232–243. https://doi.org/10.1002/JPER.19-0173

[13] Jain, N., Dutt, U., Radenkov, I., & Jain, S. (2024). WHO's global oral health status report 2022: Actions, discussion and implementation. Oral Diseases, 30(2), 73–79. https://doi.org/10.1111/odi.14516

[14] Jiang, L., Chen, D., Cao, Z., Wu, F., Zhu, H., & Zhu, F. (2022). A two-stage deep learning architecture for radiographic staging of periodontal bone loss. BMC Oral Health, 22(1), 106. https://doi.org/10.1186/s12903-022-02119-z

[15] Joo, J., Jeong, S., Jin, H., Lee, U., Yoon, J. Y., & Kim, S. C. (2019). Periodontal Disease Detection Using Convolutional Neural Networks. 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 360–362. https://doi.org/10.1109/ICAIIC.2019.8669021

[16] Kabir, T., Lee, C.-T., Nelson, J., Sheng, S., Meng, H.-W., Chen, L., Walji, M. F., Jiang, X., & Shams, S. (2021). An End-to-end Entangled Segmentation and Classification Convolutional Neural Network for Periodontitis Stage Grading from Periapical Radiographic Images. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 1370–1375. https://doi.org/10.1109/BIBM52615.2021.9669422

[17] Kim, E.-H., Kim, S., Kim, H.-J., Jeong, H., Lee, J., Jang, J., Joo, J.-Y., Shin, Y., Kang, J., Park, A. K., Lee, J.-Y., & Lee, S. (2020). Prediction of Chronic Periodontitis Severity Using Machine Learning Models Based On Salivary Bacterial Copy Number. Frontiers in Cellular and Infection Microbiology, 10, 571515. https://doi.org/10.3389/fcimb.2020.571515

[18] Kim, J., Lee, H.-S., Song, I.-S., & Jung, K.-H. (2019). DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. Scientific Reports, 9(1), 17615. https://doi.org/10.1038/s41598-019-53758-2

[19] Kong, Z., Ouyang, H., Cao, Y., Huang, T., Ahn, E., Zhang, M., & Liu, H. (2023). Automated periodontitis bone loss diagnosis in panoramic radiographs using a bespoke two-stage detector. Computers in Biology and Medicine, 152, 106374. https://doi.org/10.1016/j.compbiomed.2022.106374

[20] Krois, J., Ekert, T., Meinhold, L., Golla, T., Kharbot, B., Wittemeier, A., Dörfer, C., & Schwendicke, F. (2019). Deep Learning for the Radiographic Detection of Periodontal Bone Loss. Scientific Reports, 9(1), 8495. https://doi.org/10.1038/s41598-019-44839-3

[21] Kwon, T., Lamster, I. B., & Levin, L. (n.d.). Current concepts in the management of periodontitis. International Dental Journal, n/a(n/a). https://doi.org/10.1111/idj.12630

[22] LeCun, Y., Bottou, L., Bengio, Y., & Ha, P. (1998). Gradient-Based Learning Applied to Document Recognition.

[23] Lee, C., Kabir, T., Nelson, J., Sheng, S., Meng, H., Van Dyke, T. E., Walji, M. F., Jiang, X., & Shams, S. (2022). Use of the deep learning approach to measure alveolar bone level. Journal of Clinical Periodontology, 49(3), 260–269. https://doi.org/10.1111/jcpe.13574

[24] Lee, J.-H., Kim, D., Jeong, S.-N., & Choi, S.-H. (2018). Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. Journal of Periodontal & Implant Science, 48(2), 114. https://doi.org/10.5051/jpis.2018.48.2.114

[25] Li, H., Zhou, J., Zhou, Y., Chen, J., Gao, F., Xu, Y., & Gao, X. (2020). Automatic and Interpretable Model for Periodontitis Diagnosis in Panoramic Radiographs. In A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, & L. Joskowicz (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020 (Vol. 12262, pp. 454–463). Springer International Publishing. https://doi.org/10.1007/978-3-030-59713-9_44

[26] Liu, Q., Dai, F., Zhu, H., Yang, H., Huang, Y., Jiang, L., Tang, X., Deng, L., & Song, L. (2023). Deep learning for the early identification of periodontitis: A retrospective, multicentre study. Clinical Radiology, 78(12), e985–e992. https://doi.org/10.1016/j.crad.2023.08.017

[27] Mamede, S., Schmidt, H. G., & Rikers, R. (2007). Diagnostic errors and reflective practice in medicine. Journal of Evaluation in Clinical Practice, 13(1), 138–145. https://doi.org/10.1111/j.1365-2753.2006.00638.x

[28] Moons, K. G. M., Wolff, R. F., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Annals of Internal Medicine, 170(1), W1. https://doi.org/10.7326/M18-1377

[29] O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks (arXiv:1511.08458). arXiv. http://arxiv.org/abs/1511.08458

[30] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

[31] Patil, S., Albogami, S., Hosmani, J., Mujoo, S., Kamil, M. A., Mansour, M. A., Abdul, H. N., Bhandi, S., & Ahmed, S. S. S. J. (2022). Artificial Intelligence in the Diagnosis of Oral Diseases: Applications and Pitfalls. Diagnostics, 12(5), 1029. https://doi.org/10.3390/diagnostics12051029

[32] Sarvamangala, D. R., & Kulkarni, R. V. (2022). Convolutional neural networks in medical image understanding: A survey. Evolutionary Intelligence, 15(1), 1–22. https://doi.org/10.1007/s12065-020-00540-3

[33] Schwendicke, F., Samek, W., & Krois, J. (2020). Artificial Intelligence in Dentistry: Chances and Challenges. Journal of Dental Research, 99(7), 769–774. https://doi.org/10.1177/0022034520915714

[34] Scott, J., Biancardi, A. M., Jones, O., & Andrew, D. (2023). Artificial Intelligence in Periodontology: A Scoping Review. Dentistry Journal, 11(2), 43. https://doi.org/10.3390/dj11020043

[35] Selviani, Y., Oktawati, S., Samad, R., Muslihi, M. T., Gani, A., Lawi, A., & Paramma, Z. I. (2022). A Novel Proposed Expert System In Diagnosing Periodontal Disease. 20(15).

[36] Shaddox, L. M., & Walker, C. B. (2010). Treating chronic periodontitis: Current status, challenges, and future directions. Clinical, Cosmetic and Investigational Dentistry, 2, 79–91. https://doi.org/10.2147/ccide.s12160348

[37] Shon, H. S., Kong, V., Park, J. S., Jang, W., Cha, E. J., Kim, S.-Y., Lee, E.-Y., Kang, T.-G., & Kim, K. A. (2022). Deep Learning Model for Classifying Periodontitis Stages on Dental Panoramic Radiography. Applied Sciences, 12(17), 8500. https://doi.org/10.3390/app12178500

[38] Tonetti, M. S., Greenwell, H., & Kendler, K. S. (n.d.). Staging and grading of periodontitis: Framework and proposal of a new classification and case definition—Tonetti—2018—Journal of Clinical Periodontology—Wiley Online Library. Retrieved July 27, 2024, from https://onlinelibrary.wiley.com/doi/10.1111/jcpe.12945

[39] Zhu, X., Wang, F., Xie, Y., Li, W., Wang, X., & Liu, J. (2023). Periodontitis Auxiliary Diagnosis Based on Deep Learning and Oral Dental X-ray Images. Russian Journal of Nondestructive Testing, 59(4), 487–500. https://doi.org/10.1134/S1061830923600144