International Research Journal of Multidisciplinary Scope (IRJMS), 2026;7(4):1912-1925 IRJMS

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2026.v07i04.013703

Development and Validation of an Artificial Intelligence-based Mobile Application for Periodontal Diagnosis

Deepa Subramanian[1], Bhuminathan S[2], Vinod G[3], Harikaran Jeyakkodi[4], Bhavana Ganji[5], Sasmitha C[6], Jeevika S[7,8], Charumathi Dhanushkodi[9], Suganya P[9], Sruthi KG[10]*

[1]Department of Dental Science, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India, [2]Department of Prosthodontics and Crown and Bridge, Sree Balaji Dental College and Hospital, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India, [3]Biomedical Informatics Unit, Christian Medical College, Vellore, Tamil Nadu, India, [4]Department of Conservative Dentistry and Endodontics, Vinayaka Mission's Sankaracharya Dental College, Salem, Tamil Nadu, India, [5]Independent Research Consultant, Pune, Maharashtra, India, [6]Department of Conservative Dentistry and Endodontics, Nandha Dental College and Hospital, Erode, Tamil Nadu, India, [7]Facet Automation Pvt. Ltd., Chennai, Tamil Nadu, India, [8]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India, [9]Department of Public Health Dentistry, Sree Balaji Dental College and Hospital, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India, [10]Healing Fields Foundation, Hyderabad, Telangana, India. *Corresponding Author's Email: kgsruthi@gmail.com

Abstract

Periodontal diseases affect more than half of the global adult population, yet their clinical diagnosis remains operator-dependent and subject to inter-examiner variability. This retrospective diagnostic accuracy study aimed to develop and internally validate PerioScan AI, an artificial intelligence-based mobile application for periodontal diagnosis across seven diagnostic categories. Clinical case sheets from 745 patients attending a tertiary dental institution in Chennai, India, were pre-processed using custom Python pipelines. After removal of exact-match duplicate records, 435 unique records were retained and partitioned into a training set (n = 348, 80%) and a held-out test set (n = 87, 20%) by stratified random splitting. Seven classes were represented: periodontally healthy, localised and generalised chronic gingivitis, localised and Generalised Chronic Periodontitis (GCP), and localised and Generalised Aggressive Periodontitis (GAP). A total of 964 clinical features were extracted and one-hot encoded, yielding 3,957 predictor variables, and a Random Forest Classifier was trained and validated against the consensus diagnosis of two calibrated periodontists. Outcomes were overall accuracy, class-level sensitivity, specificity, area under the receiver operating characteristic curve and Cohen's kappa. On the held-out test set the classifier achieved an overall accuracy of 82.2% (95% confidence interval 72.9% to 88.8%) and a macro-averaged area under the curve of 0.780, marginally below the prespecified threshold of 0.80; class-level sensitivity ranged from 75.3% to 82.2%. Agreement between application-generated management recommendations and expert consensus was substantial (Cohen's kappa 0.80, 95% confidence interval 0.74 to 0.82). PerioScan AI shows promise as a first-generation decision support aid, but prospective multicentre validation incorporating 2017 staging and grading parameters is required before clinical deployment.

Keywords: Artificial Intelligence, Computer-assisted Diagnosis, Machine Learning, Periodontal Diseases, Periodontitis.

Introduction

Periodontal diseases, encompassing gingivitis and the more destructive condition of periodontitis, represent a substantial global health burden. The Global Burden of Disease (GBD) Study identified severe periodontitis as the sixth most prevalent human disease, affecting approximately 11.2% of the world population, while gingivitis approaches near-universal prevalence among dentate adults (1). The GBD 2017 analysis further estimated that oral conditions collectively affect approximately 3.5 billion people worldwide, with periodontal disease contributing a large and growing share of oral disability in low- and middle-income countries, where the burden is compounded by limited access to specialist care (2). In India, a systematic review and meta-analysis of thirty cross-sectional studies estimated the pooled prevalence of periodontal disease among adults at approximately 51%, with higher estimates among older age groups and socioeconomically disadvantaged populations (3). Periodontal disease is consequently a significant public health priority for India, where dental specialists are concentrated in urban centres and a large

This is an Open Access article distributed under the terms of the Creative Commons Attribution CC BY license (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

(Received 01st July 2026; Accepted 15th September 2026; Published 30th October 2026)

proportion of the population first presents to primary care providers without specialist periodontal training. Despite this epidemiological burden, the diagnosis of periodontal conditions in routine clinical practice remains substantially subjective. The 2017 World Workshop classification, developed jointly by the American Academy of Periodontology (AAP) and the European Federation of Periodontology (EFP), introduced a multidimensional framework encompassing a revised definition of periodontal health and gingival diseases (4, 5), an objective gingivitis case definition anchored to bleeding on probing thresholds (6), and a staging (I to IV) and grading (A to C) system for periodontitis (7). Accurate application of this system requires meticulous interpretation of multiple clinical parameters, including Probing Pocket Depth (PPD), Clinical Attachment Level (CAL), Bleeding On Probing (BOP), furcation involvement, radiographic bone loss and systemic modifiers, across individual tooth surfaces (7). Periodontal probing and charting are technique-sensitive procedures, and the resulting inter-examiner and intra-examiner variability constitutes a recognised limitation of examination-based periodontal diagnosis, particularly outside specialist settings (8). A diagnostic aid capable of integrating the full multiparametric examination dataset in a reproducible manner therefore holds clear clinical appeal.

Artificial intelligence (AI), and specifically supervised Machine Learning (ML), has emerged as a transformative paradigm for pattern recognition in complex, multivariate clinical datasets, with clinically validated applications now established across radiology, ophthalmology, cardiology and pathology (9, 10). Within dentistry, validated AI applications span caries detection, periapical pathology identification, orthodontic analysis and root morphology classification (11), and a substantial body of deep learning research has now been mapped systematically in periodontology and oral implantology (12). Systematic reviews of AI models for gingivitis and periodontal disease report encouraging but heterogeneous diagnostic performance across study designs and reference standards (13). In periodontology, prior investigations have applied Convolutional Neural Network (CNN) architectures to periapical and panoramic radiographs to diagnose periodontally compromised teeth (14), to detect periodontal bone loss (15), and to stage periodontitis radiographically (16, 17). Beyond imaging, comparative work using national survey data has demonstrated that ML algorithms can predict periodontitis from non-imaging clinical and demographic variables, although performance in that setting was constrained by the breadth of the available predictor set (18). Collectively, these studies have characteristically focused on single imaging biomarkers, narrow binary classification tasks or limited predictor sets, without integrating the full multiparametric clinical dataset required for a complete periodontal diagnosis as defined by the 2017 classification framework. Furthermore, the translation of such models into deployable, clinician-facing tools, particularly mobile applications suitable for resource-limited settings in which teledentistry and mobile health approaches have documented benefit (19), has remained an underexplored domain.

Three specific gaps in the existing evidence motivated the present study. First, to our knowledge, no published model performs patient-level, seven-category periodontal diagnostic classification, spanning health, localised and generalised gingivitis, and localised and generalised chronic and aggressive periodontitis, from structured clinical examination data alone. Second, existing periodontal AI research is dominated by radiograph-based, largely binary tasks that do not reproduce the multiparametric reasoning of a complete periodontal examination. Third, few studies couple a diagnostic output with a guideline-anchored management recommenda-tion, or package the model as a clinician-facing mobile application appropriate for primary dental care in India. PerioScan AI was developed to address this translational gap: the application integrates structured examination parameters derived from comprehensive periodontal case sheets to generate a diagnostic classification across seven categories together with a corresponding management recommendation. The originality of this work therefore lies in the combination of a seven-class, patient-level diagnostic task; a large structured clinical feature set of 964 variables in place of imaging data alone; a rule-based, guideline-anchored treatment

planning layer; and delivery through a mobile application designed for resource-limited practice. This study accordingly addressed the following objectives:

a) to develop a Random Forest Classifier (RFC)-based diagnostic model for seven periodontal disease categories using pre-processed clinical case sheet data from a South Indian tertiary dental institution;
b) to internally validate the diagnostic performance of the model against a blinded expert periodontist consensus reference standard, with the prespecified hypothesis that the model would achieve a macro-averaged area under the receiver operating characteristic curve (AUC) greater than 0.80; and
c) to quantify the agreement between application-generated management recommendations and expert consensus recommendations.

Methodology

Ethical Approval and Reporting Standard

This study received approval from the Institutional Ethics Committee (IEC) of Bharath Institute of Higher Education and Research (IEC Approval No. 12/2024). Informed written consent was obtained from all participants prior to data utilisation. The study was conducted in accordance with the Declaration of Helsinki (2013 revision) and was reported in compliance with the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 guidelines (20); model development and reporting additionally considered the TRIPOD statement for multivariable prediction models (21).

Study Design and Setting

A retrospective diagnostic accuracy study design was employed. Clinical case sheets were collected from the Department of Periodontology, Bharath Institute of Higher Education and Research Dental College and Hospital, Chennai, Tamil Nadu, India (approximate GPS coordinates 12.9428° N, 80.2057° E). The institution serves a diverse socioeconomic and demographic patient population representative of urban and peri-urban South India. Case sheets spanning June 2025 to January 2026 were included.

Dataset

A total of 745 patient clinical case sheets were received, organised across seven folders corresponding to the diagnosed periodontal conditions: periodontally healthy (n = 25), Localised Chronic Gingivitis (LCG; n = 100), Generalised Chronic Gingivitis (GCG; n = 100), Localised Chronic Periodontitis (LCP; n = 172), Generalised Chronic Periodontitis (GCP; n = 48), Localised Aggressive Periodontitis (LAP; n = 150) and Generalised Aggressive Periodontitis (GAP; n = 150). Diagnostic categories reflect the treating periodontist's provisional clinical diagnosis at the time of examination. Each case sheet was provided in spreadsheet format and consisted of three structural sections: row-wise patient demographic and history fields, extending from patient identifiers through family history and provisional diagnosis; column-wise tooth-level clinical observations, comprising PPD, CAL, BOP and furcation involvement across all sextants; and a combined row-and-column section encompassing personal history, oral hygiene practices, systemic examination and intraoral mucosal findings.

Inclusion and Exclusion Criteria

Case sheets were included if they represented a complete periodontal clinical examination performed by a qualified periodontist and corresponded to one of the seven defined diagnostic categories. Records with more than 20% missing values across core clinical fields, and exact-match duplicate records across all clinical fields, were excluded prior to model development.

Data Preprocessing and Feature Engineering

All preprocessing was implemented in Python using two custom-developed packages: data-preparer, for structural conversion, and data-preprocess, for feature selection and encoding. First-level preprocessing converted each case sheet from its heterogeneous three-section format into a uniform column-wise tabular representation, yielding 985 columns per record, inclusive of the filename and class label fields. Second-level preprocessing expanded multi-valued fields, such as personal history (smoking, alcohol), oral hygiene practices and temporomandibular joint findings, by suffixing numerical identifiers to each value position, producing 999 columns per record. The chief complaint field, which accommodated up to six distinct values, was split into six dedicated

binary indicator columns, and a unique case sheet identifier was prepended as the first column.

Class-specific preprocessed records were combined into a single multiclass dataset. Following structured review with the consultant periodontist, 35 administratively redundant or clinically non-informative fields were removed, reducing the feature space to 964 clinical variables. Of the 745 case sheets received, 310 (41.6%) were excluded as exact-match duplicate records across all clinical fields. No additional records exceeded the prespecified missing-data threshold of more than 20% missing values across core clinical fields, and no records were excluded for any other reason. The final analytical dataset therefore comprised 435 unique patient records (58.4% of the records received). Categorical variables were transformed using one-hot encoding (OHE), yielding 3,957 binary indicator features. The class label, encoded across the seven diagnostic categories, served as the dependent variable, and the final modelling dataset therefore comprised 435 records and 3,958 columns.

Machine Learning Model

An RFC was selected as the primary algorithm and implemented using the scikit-learn library in Python. The RFC is an ensemble learning method that constructs multiple decision trees on bootstrapped training subsets and aggregates predictions by majority vote, conferring robustness to overfitting in high-dimensional feature spaces (22). The dataset was partitioned into a training set (n = 348, 80%) and a held-out test set (n = 87, 20%) using stratified random splitting in order to preserve class proportions across both partitions. Hyperparameter optimisation was performed using stratified k-fold cross-validation within the training set, with a grid search across the number of estimators, maximum tree depth and minimum samples per leaf. Optimal parameters were selected on the basis of the macro-averaged F1-score in order to mitigate the effect of class imbalance.

Reference Standard

The reference standard for diagnostic accuracy assessment was the consensus diagnosis independently assigned by two calibrated specialist periodontists, each with a minimum of 10 years of clinical experience, in accordance with the 2017 AAP and EFP classification (4, 5, 7). Inter-rater agreement was quantified using Cohen's kappa before consensus determination, and discordant cases were adjudicated by a third senior periodontist. Both periodontists were blinded to model outputs during reference standard adjudication.

Treatment Plan Generation

PerioScan AI incorporated a rule-based treatment planning module that mapped each diagnostic output to a structured management recommendation in one of four categories: Oral Hygiene Instruction (OHI) alone, for healthy and gingivitis classifications; Non-Surgical Periodontal Therapy (NSPT), for Stage I and Stage II equivalent periodontitis; surgical referral, for Stage III and Stage IV equivalent periodontitis; and periodontal maintenance, or Supportive Periodontal Therapy (SPT). The algorithmic logic was anchored to the 2020 EFP S3-level clinical practice guideline (23).

Statistical Analysis

Model performance on the held-out test set was evaluated using overall accuracy, class-level sensitivity, specificity, Positive Predictive Value (PPV), Negative Predictive value (NPV) and AUC computed using a one-versus-rest (OvR) strategy. Ninety-five percent confidence intervals (CIs) for AUC were estimated using the method of DeLong. Cohen's kappa and its 95% CI were calculated for agreement between application-generated management recommendations and expert consensus. Aggregate metrics were reported both as macro-averages, computed as the unweighted arithmetic mean of the seven class-level values, and as support-weighted averages, in order to appropriately handle class imbalance, a recognised challenge in the development and evaluation of clinical classifiers (24); the weighted-average recall is numerically equivalent to overall accuracy. The 95% CI for overall accuracy was computed using the Wilson score interval. Statistical significance was set at $p < 0.05$. All analyses were performed in Python using the scikit-learn, SciPy and NumPy libraries.

Data Protection and Responsible Technology Safeguards

All case sheets were deidentified prior to preprocessing: patient names, registration numbers and contact details were removed and replaced with a study-specific case sheet identifier, and the deidentified dataset was stored on access-restricted institutional systems. PerioScan AI was designed strictly as a clinician-facing decision

support aid. Diagnostic outputs and management recommendations are presented to a qualified dental practitioner for verification against the clinical examination; they are not delivered to patients directly, and the application does not initiate, modify or discontinue treatment autonomously. The application does not transmit patient-identifiable data, and model predictions are accompanied by the underlying feature importance summary to support clinician scrutiny rather than uncritical acceptance. These design constraints were informed by international guidance on the ethics and governance of AI for health, which emphasises human oversight, transparency, data protection and accountability (25, 26). Potential avenues of misuse, and the corresponding mitigations, are considered explicitly in the Discussion.

Results

Study Population

The final analytical dataset comprised 435 unique patient records across seven diagnostic categories following deduplication from the initial 745 case sheets. The flow of records from retrieval through exclusion, feature engineering and partitioning to analysis is summarised in Figure 1. Of the 745 case sheets retrieved, 310 were excluded as exact-match duplicate records across all clinical fields, and no record was excluded for exceeding the missing-data threshold or for any other reason. The 435 unique records that remained were reduced to 964 clinical variables, expanded to 3,957 predictor variables by one-hot encoding, and partitioned by stratified random splitting into 348 training and 87 held-out test records, all 87 of which were analysed for diagnostic accuracy. The demographic and clinical characteristics of the study population are summarised in Table 1. Table 1 sets out the age and sex distribution of the cohort, the prevalence of smoking and of diabetes mellitus, the self-reported oral hygiene practice and the leading chief complaint, each expressed as a number with the corresponding percentage. The lower block of the table reports the class-wise distribution of the 745 case sheets received together with the composition of the analytical, training and test partitions and the size of the feature space at each pre-processing stage. The dataset exhibited marked class imbalance, with periodontally healthy patients constituting the smallest class and localised chronic periodontitis the largest. The training set contained 348 records and the test set 87 records, each proportionately stratified by class.

Table 1: Demographic and Clinical Characteristics of the Study Population (N = 435 after Deduplication from 745 Case Sheets Received)

Characteristic	Value
Age, mean (SD), years	42.6 (13.4)
Sex, n (%)	
Male	198 (45.5)
Female	237 (54.5)
Smoking (current or ever), n (%)	79 (18.2)
Diabetes mellitus, n (%)	54 (12.4)
Oral hygiene practice, n (%)	
Brushes once daily	314 (72.2)
Brushes twice daily	121 (27.8)
Chief complaint (leading), n (%)	
Gingival bleeding	228 (52.4)
Tooth mobility	109 (25.1)
Dental check-up	98 (22.5)
Periodontal diagnosis, case sheets received, n	
Periodontally healthy	25
Localised chronic gingivitis	100
Generalised chronic gingivitis	100
Localised chronic periodontitis	172
Generalised chronic periodontitis	48
Localised aggressive periodontitis	150
Generalised aggressive periodontitis	150
Total case sheets received	745

Records after deduplication (analytical dataset), n	435
Training set (80%), n	348
Test set (20%), n	87
Clinical features after preprocessing, n	964
Predictor variables after one-hot encoding, n	3,957

Note: SD = Standard Deviation. Diagnostic categories reflect the provisional clinical diagnosis terminology employed at the source institution prior to adoption of the 2017 classification. Class counts in the upper block describe the 745 case sheets received prior to deduplication.

Of the 745 case sheets retrieved from the departmental archive between June 2025 and January 2026, 310 were excluded as exact-match duplicate records across all clinical fields, leaving 435 unique records, as shown in Figure 1. These were reduced to 964 clinical variables and expanded to 3,957 predictor variables by one-hot encoding before stratified partitioning into a training set (n = 348) and a held-out test set (n = 87). The index test was the PerioScan AI classification across seven diagnostic categories and the reference standard was the consensus diagnosis of two calibrated periodontists, with discordant cases adjudicated by a third senior periodontist; all 87 held-out records were analysed for diagnostic accuracy. The figure was prepared from the study data using Python (Matplotlib).

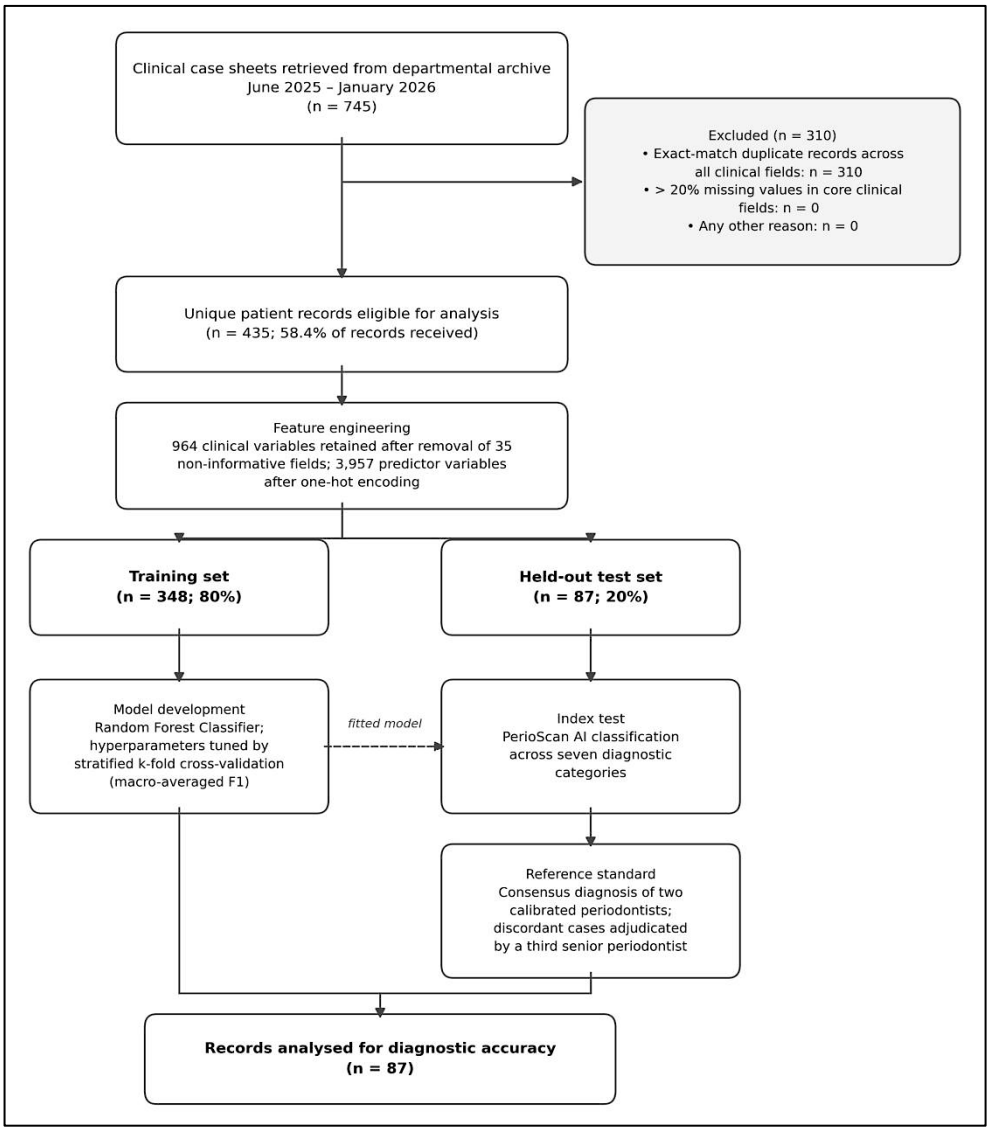


Figure 1: Flow of Clinical Case Sheets through the Study, Reported in Accordance with STARD 2015

Diagnostic Performance of Perioscan AI

On the held-out test set (n = 87), the RFC achieved an overall accuracy of 82.2% (95% CI 72.9-88.8). Class-level diagnostic performance metrics are presented in Table 2. Table 2 reports the sensitivity, specificity, positive predictive value, negative predictive value and area under the receiver operating characteristic curve, each with its 95% confidence interval, for every one of the seven diagnostic categories. The final two rows of the table give the macro-averaged and support-weighted summary values, the macro-average being the unweighted mean across classes and the weighted average being adjusted for class support. The macro-averaged AUC across all seven classes was 0.780 and the weighted-average AUC was 0.847. Class-level AUC exceeded the prespecified threshold of 0.80 for Localised Chronic Gingivitis [0.832] and Localised Aggressive Periodontitis [0.813], and ranged from 0.745 to 0.769 for the remaining five categories. Per-class sensitivity was highest for generalised chronic gingivitis (82.2%) and localised chronic gingivitis (80.1%), and lowest for generalised aggressive periodontitis (75.3%) and generalised chronic periodontitis (78.1%); the latter finding is likely attributable to the limited training representation of that class. The confusion matrix computed on the held-out test set indicated that the majority of misclassifications occurred between adjacent disease severity categories, consistent with biological overlap in clinical presentations.

Table 2: Diagnostic Performance of PerioScan AI Against the Expert Reference Standard (Held-out Test Set, n = 87)

Diagnostic Category	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	AUC (95% CI)
Periodontally healthy	79.2 (67.3-89.2)	74.3 (64.2-79.9)	9.9 (3.0-27.8)	99.0 (92.6-99.9)	0.768 (0.500-0.964)
Localised chronic gingivitis	80.1 (74.3-89.4)	75.3 (67.4-89.3)	34.2 (19.6-52.8)	95.9 (87.2-98.7)	0.832 (0.685-0.979)
Generalised chronic gingivitis	82.2 (71.9-85.3)	71.2 (59.5-81.3)	31.4 (18.2-49.4)	96.2 (86.4-98.5)	0.767 (0.603-0.931)
Localised chronic periodontitis	78.2 (68.3-86.3)	75.5 (67.4-89.3)	45.4 (29.4-62.5)	93.0 (83.2-97.3)	0.768 (0.632-0.904)
Generalised chronic periodontitis	78.1 (64.6-80.6)	75.8 (67.4-89.3)	19.3 (8.4-39.0)	97.9 (90.1-99.4)	0.769 (0.542-0.996)
Localised aggressive periodontitis	79.2 (69.3-86.5)	71.6 (59.5-81.3)	42.1 (27.0-58.5)	93.0 (83.1-97.5)	0.813 (0.686-0.940)
Generalised aggressive periodontitis	75.3 (67.4-89.3)	73.8 (67.4-89.3)	42.8 (27.0-59.4)	92.0 (82.6-97.1)	0.745 (0.605-0.885)
Macro-average	78.9	73.9	32.2	95.3	0.780
Weighted average	82.2	74.3	37.7	91.2	0.847

Note: AUC = Area Under the Receiver Operating Characteristic Curve, computed using a one-versus-rest strategy, CI = Confidence Interval, estimated using the method of DeLong, NPV = Negative Predictive Value, PPV = Positive Predictive Value. AUC values are reported on a 0 to 1 scale. Macro-averages are the unweighted arithmetic means of the seven class-level values and weighted averages are support-weighted; confidence intervals are not reported for aggregate rows.

Agreement with Expert Consensus

Agreement between the management recommendations generated by PerioScan AI and the recommendations of the expert consensus panel yielded an overall Cohen's kappa of 0.80 (95% CI 0.74-0.82; $p < 0.001$), indicating substantial agreement. Agreement was highest for the oral hygiene instruction category (kappa 0.84) and lowest for surgical referral (kappa 0.76), reflecting the downstream effect of accurate periodontitis staging on the fidelity of the management recommendation.

Comparative Performance and Feature Importance

A comparison of PerioScan AI with published artificial intelligence models for periodontal diagnosis is presented in Table 3. Table 3 sets out, for each comparator study, the model architecture, the composition of the sample and the reported accuracy, sensitivity and specificity, alongside the corresponding values for the present study. Reported accuracy in the four comparator studies ranged from 76.7% to 87%, and each of those

studies addressed a radiographic and largely binary classification task. The present model, by contrast, addressed a seven-class, patient-level classification derived entirely from structured clinical examination data. Feature importance analysis derived from the RFC identified probing depth measurements across the posterior sextants, clinical attachment level recordings, bleeding on probing indices and smoking status as the most discriminative predictor categories. Figure 2 summarises the study as a whole. It traces the pipeline from the 745 case sheets retrieved at the source institution, through pre-processing and feature engineering, model development and validation against the expert reference standard, to the dual output of the application, namely a seven-category diagnosis coupled to a guideline-anchored management recommendation. The lower band of the figure records the principal findings on the held-out test set and the intended scope of clinical use of the tool.

Table 3: Comparison of PerioScan AI with Published Artificial Intelligence Models for Periodontal Diagnosis

Year	Model or algorithm	Sample	Accuracy, %	Sensitivity, %	Specificity, %	Key finding	References
2018	CNN (pre-trained and self-trained; Keras)	1,740 periapical images (1,044 training, 348 validation, 348 test)	81.0 premolars; 76.7 molars	Not reported by class	Not reported by class	Tooth-level diagnosis and extraction prediction for periodontally compromised teeth.	(14)
2019	Seven-layer feed-forward CNN	2,001 image segments from panoramic radiographs	81 (SD 2)	81 (SD 4)	81 (SD 5)	CNN accuracy did not differ significantly from six dentist examiners (p = 0.067).	(15)
2020	Hybrid deep learning and computer-aided detection	2,000 panoramic radiographs	Not reported	Not reported by class	Not reported by class	Intraclass correlation coefficient 0.91 between automatic and radiologist periodontal staging.	(16)
2022	Multitasking InceptionV3, 5-fold cross-validation	236 patients, full-mouth periapical radiographs	87 (SD 1)	86 (SD 3)	88 (SD 3)	Binary radiographic bone loss classification (mild, less than 15% bone loss, versus severe, 15% or more).	(17)
2026	Random Forest Classifier (scikit-learn), one-hot encoded clinical features	435 unique case sheets (348 training, 87 test)	82.2	82.2	74.3	Seven-class periodontal diagnosis from 964 structured clinical parameters (3,957 binary features after one-hot encoding).	Present study

Note: CNN = convolutional neural network, SD = standard deviation. Values shown in parentheses after an accuracy, sensitivity or specificity figure are the standard deviations reported by the source study. Sensitivity and specificity for the present study are the support-weighted averages from Table 2.

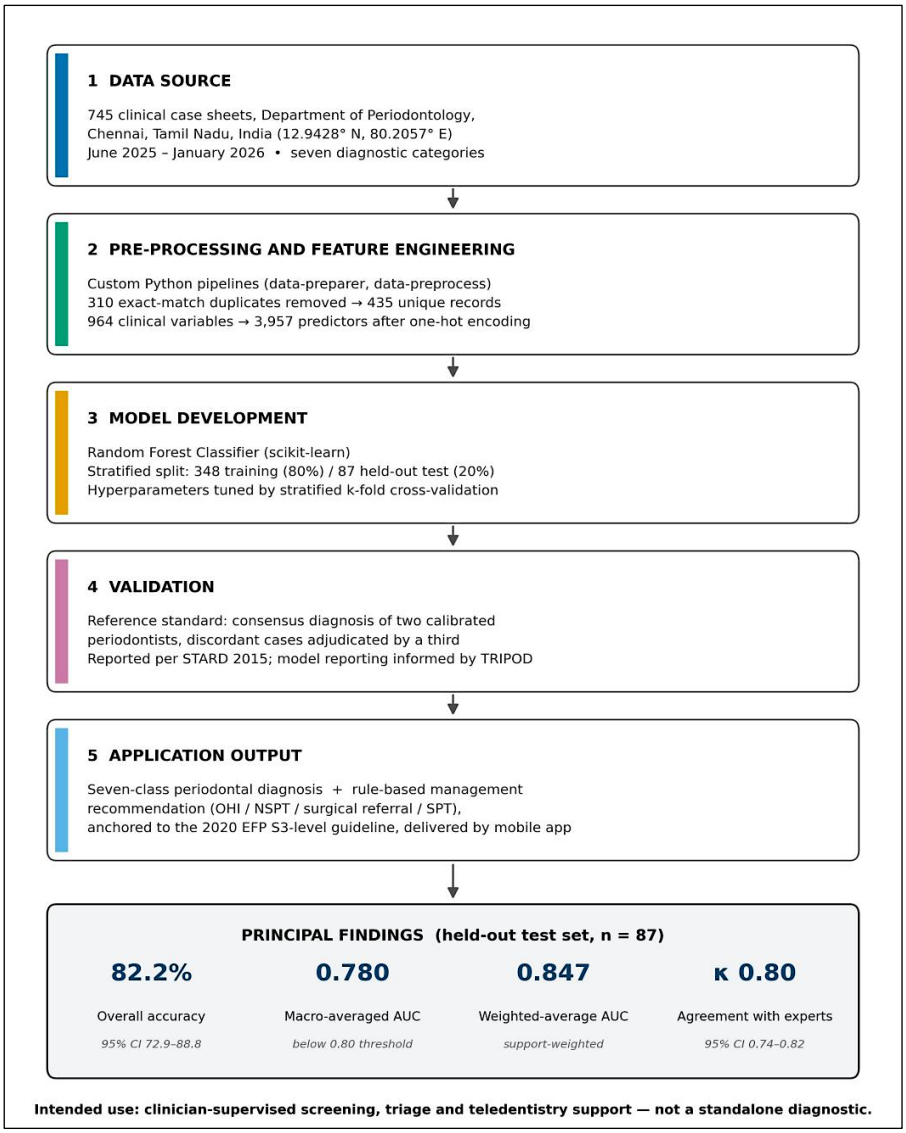


Figure 2: Overview of the Study

Note: AUC = Area under the Receiver Operating Characteristic Curve; EFP: European Federation of Periodontology; NSPT: non-surgical Periodontal Therapy; OHI: Oral Hygiene Instruction; SPT: Supportive Periodontal Therapy; STARD: Standards for Reporting of Diagnostic Accuracy Studies; TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.

In Figure 2, Stages 1 to 5 show the data source, the pre-processing and feature engineering pipeline, model development, validation against the reference standard, and the diagnostic and management output of the application. The lower band reports the principal findings on the held-out test set (n = 87) and the intended scope of clinical use. The figure was prepared from the study data using Python (Matplotlib).

Discussion

The present study describes the development and internal validation of PerioScan AI, an AI-based diagnostic application for periodontal disease classification across seven categories using structured multiparametric clinical examination data. To our knowledge, this is among the first studies to apply an RFC to a seven-class periodontal diagnostic problem,

including separate localised and generalised subclassifications of gingivitis, chronic periodontitis and aggressive periodontitis, using real-world case sheet data from a South Indian tertiary dental institution. The model correctly classified 82.2% of held-out records across seven categories. The macro-averaged AUC of 0.780 fell marginally below the prespecified threshold of 0.80, so the study hypothesis was not met at the aggregate level, although class-level discrimination exceeded that threshold for localised chronic gingivitis and localised aggressive periodontitis. Ensemble ML methods therefore show promise for periodontal diagnostic classification from comprehensive examination data, but the present model does not yet reach the discrimination required for standalone clinical use.

When contextualised within the published literature, the present findings align with and extend prior artificial intelligence investigations in periodontology. A convolutional neural network was applied to 1,740 periapical radiographic images in order to diagnose periodontally compromised teeth, and diagnostic accuracies of 81.0% for premolars and 76.7% for molars were achieved; that task was tooth-level and radiographic rather than a patient-level diagnostic classification (14). A seven-layer feed-forward convolutional neural network was applied to 2,001 image segments synthesised from panoramic radiographs in a German population, and a mean accuracy of 81% was reported for the binary detection of periodontal bone loss, which did not differ significantly from the accuracy achieved by six dentist examiners (15). A hybrid deep learning and computer-aided detection approach was applied to 2,000 panoramic radiographs in order to diagnose periodontal bone loss and to stage periodontitis automatically, and an intraclass correlation coefficient of 0.91 was reported between automated and radiologist staging, although class-level sensitivity and specificity were not reported (16). In a subsequent investigation, a multitasking InceptionV3 architecture was applied to full-mouth periapical radiographs obtained from 236 patients, and an accuracy of 87% was achieved for a binary radiographic bone loss classification, without full disease classification or management output (17). In a non-imaging context, multiple machine learning algorithms were systematically compared for the prediction of periodontitis using national survey data from Taiwanese and United States cohorts, and larger samples, richer predictor sets and robust external validation were concluded to be prerequisites for clinically useful models (18). The present study advances this body of evidence by classifying seven distinct periodontal categories, by incorporating 964 structured clinical features rather than imaging data alone, and by generating structured management recommendations, thereby approximating the complete clinical periodontal workflow.

The selection of the RFC was justified by the characteristics of the present dataset. Following one-hot encoding, the feature space comprised 3,957 binary predictors derived from 435 samples, a high-dimensional and low-sample-count regime in which ensemble methods with inherent feature selection and variance reduction properties are preferable to high-capacity single-kernel algorithms susceptible to overfitting (22). The RFC also provides native feature importance rankings, conferring a degree of interpretability that supports clinical adoption and addresses concerns about opaque decision-making in AI-assisted diagnostics (11).

Class imbalance represented the principal methodological challenge of this investigation (24). The periodontally healthy category was substantially underrepresented relative to the disease categories, and generalised chronic periodontitis was likewise a minority class. Stratified partitioning preserved class proportions in the training and test sets, and macro-averaged metrics ensured that minority class performance was not masked by majority class accuracy. Notwithstanding these measures, generalised chronic periodontitis yielded among the lowest per-class diagnostic performance, an expected finding given the scarcity of training exemplars. Targeted prospective enrichment of this category in future data collection is a research priority.

The agreement of management recommendations with expert consensus, quantified by a Cohen's kappa of 0.80, represents a clinically relevant outcome beyond the diagnostic accuracy metrics themselves. An application that generates accurate diagnoses but inappropriate management recommendations would provide limited clinical value. The treatment planning algorithm was explicitly anchored to the published EFP S3-level

guideline (23), and the observed kappa values suggest that the recommendations generated by the application approximated expert clinical practice at a level acceptable for a first-generation tool. The lower agreement observed for surgical referral (kappa 0.76) reflects the complex, multifactorial judgement underpinning surgical candidacy decisions (27), particularly in relation to systemic modifiers, patient compliance and local anatomical factors that are not fully captured within the present feature set.

Safeguards against misuse of the technology also merit explicit discussion. Three risks are pertinent to a tool of this kind. First, use of the application as a standalone diagnostic instrument by personnel without periodontal training could propagate misclassification into treatment decisions, a risk amplified by automation bias, the documented tendency of users to accept algorithmic outputs uncritically. PerioScan AI therefore enforces a clinician-in-the-loop workflow in which every output must be verified against the clinical examination before any management action, and the interface presents the diagnostic output together with its feature importance summary rather than as an unexplained verdict. Second, deployment of the model in populations that differ materially from the development population could yield systematically degraded or biased performance; accordingly, the application is restricted to research and supervised pilot use pending external validation, and post-deployment performance monitoring, model versioning and audit trails are planned prior to any clinical release. Third, diagnostic health data are sensitive: the application does not transmit patient-identifiable information, and all development data were deidentified and held on access-restricted systems. These mitigations follow international guidance on the ethics and governance of AI for health (25) and published ethical frameworks for the implementation of ML in health care (26).

Practical Implications and Recommendations

The findings support a defined and deliberately limited near-term role for PerioScan AI. In practical terms, the application is best positioned as a screening and triage aid in primary dental care and community outreach settings, where mid-level providers and general practitioners could use it to standardise the interpretation of a comprehensive periodontal examination and to prioritise referral of probable periodontitis cases to specialist care; the high negative predictive values observed across the seven diagnostic categories (92.0% to 99.0%) particularly support its use for ruling out advanced disease in screening workflows. The application can also be integrated into existing teledentistry pathways, in which structured examination data collected peripherally are reviewed centrally by a periodontist (19), and can serve as an educational adjunct for dental students and interns by making explicit the clinical parameters that drive a periodontal diagnosis. Conversely, the current evidence does not support standalone diagnostic use, use in place of radiographic assessment, or unsupervised deployment outside the development population. On the basis of these findings, we recommend: prospective data collection using case sheets aligned to the 2017 AAP and EFP classification, with radiographic bone loss recorded explicitly; external, multicentre validation across diverse Indian populations before any clinical deployment (28); targeted enrichment of underrepresented classes, particularly generalised chronic periodontitis and periodontal health; and periodic recalibration with performance auditing once the tool enters supervised pilot use.

Several limitations of this investigation require explicit acknowledgement. First, the dataset was derived from a single academic institution in Chennai, and the trained model may not generalise to populations with divergent demographic, systemic disease, dietary or oral hygiene profiles without retraining on representative data. Second, the diagnostic categories employed in the source case sheets reflect the pre-2018 nomenclature of chronic and aggressive periodontitis; a conceptual mapping to the 2017 AAP and EFP staging and grading framework was applied, but residual label inconsistency may constitute a source of noise. Critically, the current feature set does not include explicit radiographic bone level measurements, a required parameter for definitive periodontitis staging under the 2017 system (4, 7). Third, the class labels used for model training derive from the provisional diagnosis recorded at the time of examination, and the extent to which these were independently re-adjudicated by the reference standard panel constitutes a potential source of incorporation bias. Fourth, 310 of the 745 case

sheets received were exact-match duplicates across all clinical fields; while these were removed, the high duplicate rate raises the possibility of template-based data entry, and the independence of the retained 435 records cannot be fully verified. Fifth, the post-deduplication sample of 435 records across seven classes resulted in limited per-class test set sizes for minority conditions, with correspondingly wide confidence intervals. Finally, this represents an internal validation only; external validation on geographically and institutionally independent datasets, and prospective clinical testing in real practice settings, are essential prerequisites for clinical deployment (28).

Future directions include prospective data collection using a case sheet aligned to the 2017 AAP and EFP classification and incorporating radiographic bone loss measurements and explicit staging and grading criteria; multicentre validation across diverse Indian geographic and demographic populations; integration of intraoral image analysis using CNN modules to complement structured clinical data; clinical usability evaluation using validated human-computer interaction frameworks; and health-economic modelling of the diagnostic cost-benefit achievable through AI-assisted periodontal screening in community and primary care dental settings in India.

Conclusion

PerioScan AI classified seven periodontal disease categories with an overall accuracy of 82.2% using structured, multiparametric clinical examination data processed through a Random Forest Classifier trained on 348 patient records and validated on an independent held-out test set of 87 records. The macro-averaged AUC of 0.780 fell marginally short of the prespecified threshold of 0.80, indicating that discrimination across the minority classes requires improvement, although class-level discrimination exceeded the threshold for localised chronic gingivitis and localised aggressive periodontitis and negative predictive values were uniformly high. Agreement of the application-generated management recommendations with expert periodontists, quantified by a Cohen's kappa of 0.80, supports the potential clinical utility of the tool as a decision support aid rather than a standalone diagnostic. Within those boundaries, the study demonstrates that a complete, patient-level periodontal diagnostic classification, rather than a single radiographic biomarker, can be approximated from routinely recorded clinical examination data, and that such a model can be coupled to guideline-anchored management recommendations and delivered through a mobile application suited to resource-limited practice. The immediate value of the tool therefore lies in screening, triage, tele dentistry support and education under clinician oversight, with explicit technical and governance safeguards against standalone or out-of-population use. Prospective multicentre validation incorporating radiographic staging data aligned to the 2017 AAP and EFP classification is warranted in order to establish external validity and to support translation of PerioScan AI into routine clinical practice.

Abbreviations

CI: Confidence Interval, GAP: Generalised Aggressive Periodontitis, GCG: Generalised Chronic Gingivitis, NSPT: Non-Surgical Periodontal Therapy, OHE: One-Hot Encoding, RFC: Random Forest Classifier SD: Standard Deviation.

Acknowledgement

The authors thank the Department of Periodontology, Bharath Institute of Higher Education and Research Dental College and Hospital, Chennai, for access to the clinical case sheet archive.

Author Contributions

Deepa Subramanian: Conceptualisation, Methodology, Investigation, Data Collection, Writing - Original Draft, Writing - Review and Editing, Project Administration, Bhuminathan S: Conceptualisation, Validation, Clinical Supervision, Writing - Review and Editing, Supervision, Vinod G: Methodology, Software, Data Curation, Formal Analysis, Writing - Review and Editing, Harikaran Jeyakkodi: Investigation, Data Collection, Validation, Clinical Supervision, Writing - Review and Editing, Bhavana Ganji: Resources, Writing - Review and Editing, Sasmitha C: Investigation, Data Collection, Resources, Writing - Review and Editing, Jeevika S: Software, Data Curation, Formal Analysis, Investigation, Data Collection, Writing - Review and Editing, Charumathi Dhanushkodi: Writing - Review and Editing, Suganya P: Writing -

Review and Editing, Sruthi KG: Methodology, Formal Analysis, Writing - Original Draft, Writing - Review and Editing. All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Conflict of Interest

PerioScan AI was developed as an academic research tool; no commercial interest or industry sponsorship was received. Vinod G, one of the co-authors, is employed by Elsevier in a researcher-support role. This employment is unrelated to the present work and had no bearing on the design, conduct, analysis, interpretation or reporting of the study, and the employer had no role in the decision to submit the manuscript. The remaining authors declare that they have no competing interests.

Data Availability

The deidentified analytical dataset and the preprocessing and modelling scripts supporting the findings of this study are available from the corresponding author on reasonable request. The raw clinical case sheets cannot be shared publicly because they contain identifiable patient information.

Declaration of Artificial Intelligence (AI) Assistance

The authors acknowledge the use of generative AI to assist in the paraphrasing of certain parts of the manuscript. All content generated through AI was critically reviewed, edited and verified by the authors to ensure accuracy, originality and compliance with academic standards. The authors take full responsibility for the integrity and validity of the manuscript. No figure in this manuscript was created using artificial intelligence or generative artificial intelligence: Figures 1 and 2 were prepared by the authors from the study data using Python (Matplotlib), and no figure has been reproduced from any previously published source. Artificial intelligence methods were additionally the subject of the research itself and are described in full in the Methodology section.

Ethics Approval

This study was approved by the Institutional Ethics Committee of Bharath Institute of Higher Education and Research, Chennai, India (IEC Approval No. 12/2024). The study was conducted in accordance with the Declaration of Helsinki (2013 revision). Informed written consent was obtained from all participants prior to the use of their clinical records.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

1. Kassebaum NJ, Bernabé E, Dahiya M, *et al*. Global burden of severe periodontitis in 1990-2010: a systematic review and meta-regression. J Dent Res. 2014;93(11):1045-53. https://doi.org/10.1177/0022034514552491
2. Bernabé E, Marcenes W, Hernandez CR, *et al*. Global, regional, and national levels and trends in burden of oral conditions from 1990 to 2017: a systematic analysis for the Global Burden of Disease 2017 Study. J Dent Res. 2020;99(4):362-73. https://doi.org/10.1177/0022034520908533
3. Janakiram C, Mehta A, Venkitachalam R. Prevalence of periodontal disease among adults in India: a systematic review and meta-analysis. J Oral Biol Craniofac Res. 2020;10(4):800-6. https://doi.org/10.1016/j.jobcr.2020.10.016
4. Papapanou PN, Sanz M, Buduneli N, *et al*. Periodontitis: consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. J Periodontol. 2018;89(Suppl 1):S173-82. https://doi.org/10.1002/JPER.17-0721
5. Chapple ILC, Mealey BL, Van Dyke TE, *et al*. Periodontal health and gingival diseases and conditions on an intact and a reduced periodontium: consensus report of workgroup 1 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. J Periodontol. 2018;89(Suppl 1):S74-84. https://doi.org/10.1002/JPER.17-0719
6. Trombelli L, Farina R, Silva CO, *et al*. Plaque-induced gingivitis: case definition and diagnostic considerations. J Periodontol. 2018;89(Suppl 1):S46-73. https://doi.org/10.1002/JPER.17-0576
7. Tonetti MS, Greenwell H, Kornman KS. Staging and grading of periodontitis: framework and proposal of a new classification and case definition. J Periodontol. 2018;89(Suppl 1):S159-72. https://doi.org/10.1002/JPER.18-0006
8. Preshaw PM. Detection and diagnosis of periodontal conditions amenable to prevention. BMC Oral Health. 2015;15(Suppl 1):S5. https://doi.org/10.1186/1472-6831-15-S1-S5
9. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. https://doi.org/10.1038/s41591-018-0300-7
10. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14):1347-58. https://doi.org/10.1056/NEJMra1814259

11. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 2020;99(7):769-74. https://doi.org/10.1177/0022034520915714
12. Mohammad-Rahimi H, Motamedian SR, Pirayesh Z, *et al*. Deep learning in periodontology and oral implantology: a scoping review. J Periodontal Res. 2022;57(5):942-51. https://doi.org/10.1111/jre.13037
13. Revilla-León M, Gómez-Polo M, Barmak AB, *et al.* Artificial intelligence models for diagnosing gingivitis and periodontal disease: a systematic review. J Prosthet Dent. 2023;130(6):816-24. https://doi.org/10.1016/j.prosdent.2022.01.026
14. Lee JH, Kim DH, Jeong SN, *et al*. Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm. J Periodontal Implant Sci. 2018;48(2):114-23. https://doi.org/10.5051/jpis.2018.48.2.114
15. Krois J, Ekert T, Meinhold L, *et al*. Deep learning for the radiographic detection of periodontal bone loss. Sci Rep. 2019;9(1):8495. https://doi.org/10.1038/s41598-019-44839-3
16. Chang HJ, Lee SJ, Yong TH, *et al*. Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. Sci Rep. 2020;10(1):7531. https://doi.org/10.1038/s41598-020-64509-z
17. Chang J, Chang MF, Angelov N, *et al*. Application of deep machine learning for the radiographic diagnosis of periodontitis. Clin Oral Investig. 2022;26(11):6629-37. https://doi.org/10.1007/s00784-022-04617-4
18. Bashir NZ, Rahman Z, Chen SL. Systematic comparison of machine learning algorithms to develop and validate predictive models for periodontitis. J Clin Periodontol. 2022;49(10):958-69. https://doi.org/10.1111/jcpe.13692
19. Estai M, Kanagasingam Y, Tennant M, *et al*. A systematic review of the research evidence for the benefits of teledentistry. J Telemed Telecare. 2018;24(3):147-56. https://doi.org/10.1177/1357633X16689433
20. Bossuyt PM, Reitsma JB, Bruns DE, *et al*. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. https://doi.org/10.1136/bmj.h5527
21. Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55-63. https://doi.org/10.7326/M14-0697
22. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32. https://doi.org/10.1023/A:1010933404324
23. Sanz M, Herrera D, Kebschull M, *et al.* Treatment of stage I-III periodontitis: the EFP S3 level clinical practice guideline. J Clin Periodontol. 2020;47(Suppl 22):4-60. https://doi.org/10.1111/jcpe.13290
24. Chawla NV, Bowyer KW, Hall LO, *et al*. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321-57. https://doi.org/10.1613/jair.953
25. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. https://www.who.int/publications/i/item/9789240029200
26. Char DS, Shah NH, Magnus D. Implementing machine learning in health care: addressing ethical challenges. N Engl J Med. 2018;378(11):981-3. https://doi.org/10.1056/NEJMp1714229
27. Graziani F, Karapetsa D, Alonso B, *et al*. Nonsurgical and surgical treatment of periodontitis: how many options for one disease? Periodontol 2000. 2017;75(1):152-88. https://doi.org/10.1111/prd.12201
28. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245-7. https://doi.org/10.1016/j.jclinepi.2015.04.005

How to Cite: Subramanian D, Bhuminathan S, Vinod G, Jeyakkodi H, Ganji B, Sasmitha C, Jeevika S, Dhanushkodi C, Suganya P, Sruthi KG. Development and Validation of an Artificial Intelligence-based Mobile Application for Periodontal Diagnosis. Int Res J Multidiscip Scope. 2026;7(4):1912-1925.
DOI: 10.47857/irjms.2026.v07i04.013703