



Predictive performance of gradient boosting, random forest, and logistic regression for periodontitis in adults in the United States

Rendimiento predictivo de la potenciación del gradiente, el bosque aleatorio y la regresión logística para periodontitis en adultos estadounidenses

Maria Fernanda Escalante Freitez¹  

¹ Florida International University, USA.

Submitted: 12-03-2025 | Revised: 17-05-2025 | Accepted: 30-08-2025 | Published: 10-09-2025

Corresponding Author: Maria Fernanda Escalante Freitez 

ABSTRACT

Introduction: Periodontitis is a common chronic condition whose early identification supports prevention, yet it is unclear whether tree ensemble models outperform logistic regression for predicting it from routine sociodemographic and systemic data. **Methods:** We analyzed adults aged 30 years and older with a complete full-mouth periodontal examination in three cycles of the National Health and Nutrition Examination Survey (NHANES, 2009-2014). Total periodontitis was defined with the CDC/AAP case definitions. Logistic regression, random forest, and gradient boosting were trained on nine sociodemographic and systemic predictors and compared with stratified 5-fold cross-validation using the area under the receiver operating characteristic curve (AUROC) and pairwise DeLong tests. **Results:** The analytic sample comprised 10,642 adults, of whom 4,903 (46.07%) had periodontitis. Gradient boosting achieved the highest discrimination (AUROC 0.71; 95% CI 0.71 to 0.72), significantly higher than logistic regression (AUROC 0.70) and random forest (AUROC 0.68). Random forest did not outperform logistic regression. Age, poverty-income ratio, sex, race/ethnicity, and body mass index were the most influential predictors. **Conclusions:** Gradient boosting predicts periodontitis more accurately than logistic regression or random forest using readily available variables, whereas random forest offers no advantage over logistic regression.

Keywords: machine learning; periodontitis; NHANES; logistic regression; random forest; gradient boosting.

RESUMEN

Introducción: La periodontitis es una enfermedad crónica frecuente cuya identificación temprana favorece la prevención, aunque no está claro si los modelos de conjuntos de árboles superan a la regresión logística para predecirla a partir de datos sociodemográficos y sistémicos habituales. **Métodos:** Se analizaron adultos de 30 años o más con examen periodontal de boca completa en tres ciclos de la Encuesta Nacional de Examen de Salud y Nutrición (NHANES, 2009-2014). La periodontitis total se definió con los criterios CDC/AAP. Se entrenaron modelos de regresión logística, bosque aleatorio y potenciación del gradiente con nueve predictores sociodemográficos y sistémicos, y se compararon con validación cruzada estratificada de 5 pliegues mediante el área bajo la curva ROC (AUROC) y pruebas de DeLong pareadas. **Resultados:** La muestra analítica incluyó 10 642 adultos, de los cuales 4903 (46,07%) presentaban periodontitis. La potenciación del gradiente logró la mayor discriminación (AUROC 0,71; IC 95 % 0,71 a 0,72), significativamente superior a la regresión logística (AUROC 0,70) y al bosque aleatorio (AUROC 0,68). El bosque aleatorio no superó a la regresión logística. La edad, la razón pobreza-ingreso, el sexo, la raza/etnia y el índice de masa corporal fueron los predictores más influyentes. **Conclusiones:** La potenciación del gradiente predice la periodontitis con mayor exactitud que la regresión logística o el bosque aleatorio empleando variables de fácil acceso, mientras que el bosque aleatorio no aporta ventaja sobre la regresión logística.

Palabras clave: aprendizaje automático; periodontitis; NHANES; regresión logística; bosque aleatorio; potenciación del gradiente.

INTRODUCTION

Periodontitis is a chronic inflammatory disease of the tooth-supporting tissues that progresses through the destruction of the periodontal ligament and alveolar bone and, when untreated, results in attachment loss, tooth mobility, and tooth loss. It is among the most prevalent noncommunicable conditions of adulthood. A systematic review and meta-analysis of studies published between 2011 and 2020 estimated that periodontitis affects nearly two thirds of dentate adults worldwide, with severe disease concentrated in older age groups and in populations with limited access to care.(1) Estimates from the Global Burden of Disease study place oral disorders, of which periodontitis is a principal component, among the most common health conditions globally, affecting billions of people and generating substantial years lived with disability.(2) Because much of this burden is preventable when the disease is detected before irreversible tissue destruction has occurred, the accurate estimation of individual risk from information that is already available in routine care is a matter of practical consequence.

The epidemiology of periodontitis in the United States is well documented. National surveillance conducted within the National Health and Nutrition Examination Survey (NHANES) and interpreted with the surveillance case definitions of the Centers for Disease Control and Prevention and the American Academy of Periodontology (CDC/AAP) reported that almost half of adults aged 30 years and older had periodontitis in 2009 and 2010.(3) The condition does not occur in isolation: consensus documents link periodontitis bidirectionally with atherosclerotic cardiovascular disease, and it shares risk factors with diabetes and other metabolic disorders.(4) This comorbidity profile reinforces the value of risk prediction that can operate within medical as well as dental settings, where periodontal examination is not routinely performed.

The risk indicators of periodontitis are stable across populations and study designs. Older age, male sex, lower educational attainment, poverty, current smoking, and diabetes have been repeatedly associated with higher prevalence and greater severity in the NHANES adult population.(5) The clinical case for risk prediction is reinforced by the burden of the reference examination itself. A full-mouth periodontal examination is time-consuming and is not performed during routine medical encounters, so many adults with undiagnosed periodontitis are seen only by physicians and other non-dental providers.(3) A model that estimates periodontal risk from age, sex, socioeconomic indicators, and systemic conditions could extend case-finding into settings where periodontal probing is unavailable. This practical motivation makes the relative performance of statistical and machine learning methods a decision of consequence rather than a purely methodological exercise. These associations are conventionally quantified with multivariable logistic regression, which expresses each exposure as an additive and linear contribution to the log-odds of disease. The logistic model is transparent and familiar, but it cannot represent the nonlinearities and interactions that plausibly govern the relationship between continuous exposures such as age, body mass index, or the poverty-income ratio and periodontal risk unless those terms are specified by the analyst in advance.(6) Specifying the correct functional form is rarely possible when the underlying biology is incompletely understood.

Tree ensemble learners relax this requirement by construction. Random forests aggregate many de-correlated decision trees and thereby approximate nonlinear decision boundaries and interactions without explicit specification.(7) Gradient boosting assembles trees sequentially, fitting each new tree to the residuals of the preceding ensemble, a strategy that tends to produce the most accurate models for tabular clinical data.(8) The diffusion of these methods into periodontology has been rapid. Recent reviews describe a growing literature of artificial intelligence applied to periodontal diagnosis, staging, and risk prediction, spanning imaging, biological markers, and clinical records.(9-11) A systematic review and meta-analysis of patient-based studies concluded that artificial intelligence models achieve high accuracy for periodontitis diagnosis from non-invasive biological markers, while noting that the advantage over conventional statistical models is not uniform.(12) Whether machine learning actually outperforms logistic regression for periodontitis prediction from ordinary clinical variables remains unresolved. A systematic comparison of ten algorithms trained on Taiwanese and United States survey data found that gradient boosting and related ensembles reached the highest area under the receiver operating characteristic curve (AUROC), although the gain over logistic regression was modest and varied between datasets.(13) Models built on large electronic dental records and on the NHANES likewise reported competitive but not uniformly superior performance of machine learning relative to statistical models.(14,15) Longitudinal studies that predicted periodontitis progression with artificial intelligence also report ensemble methods ahead of, but not always statistically distinct from, simpler models.(16)

The resulting inconsistency leaves a specific question unanswered: given identical predictors,

identical validation, and paired statistical testing, do gradient boosting and random forest predict periodontitis more accurately than logistic regression in a nationally representative US sample? The objective of this study was to compare the predictive performance of gradient boosting, random forest, and logistic regression for total periodontitis defined by the CDC/AAP case definition in adults from NHANES 2009-2014, and to rank the most influential predictors. Three hypotheses were tested. H1: gradient boosting achieves a higher AUROC than logistic regression. H2: random forest achieves a higher AUROC than logistic regression. H3: gradient boosting and random forest achieve comparable discrimination.

METHODS

This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for cross-sectional studies(17)and the Transparent Reporting of a multi-variable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement.(6)

Study design and data source

We conducted a cross-sectional secondary analysis of publicly available data from three consecutive cycles of NHANES: 2009-2010, 2011-2012, and 2013-2014.(18)NHANES is a stratified, multistage probability survey designed to be representative of the civilian noninstitutionalized population of the United States. Each cycle collects household interviews, standardized physical examinations, and laboratory measures using a complex sampling design with oversampling of selected demographic groups. The de-identified public-use files and their documentation are distributed by the National Center for Health Statistics of the Centers for Disease Control and Prevention.(18)

Study population

Public-use files were merged across cycles by participant identifier. Before modeling we performed range and consistency checks on each variable, treated the survey reserved codes for refused, missing, or cannot-measure as missing, and verified the denominators of every derived variable against the raw component files. Continuous measurements outside physiologically plausible bounds were set to missing. The periodontal measurements were processed at the site level to compute per-tooth maximum pocket depth and attachment loss, from which the case-definition criteria were derived.

Eligible participants were adults aged 30 years and older who completed the full-mouth periodontal examination. In these cycles periodontal probing was performed on six sites per tooth across 28 teeth, excluding third molars, following a standardized protocol. At each site the examiner recorded pocket depth and clinical attachment loss to the nearest millimetre using a periodontal probe, with each measurement taken at the disto-facial, mid-facial, mesio-facial, disto-lingual, mid-lingual, and mesio-lingual positions. The examination was restricted to adults aged 30 years and older because periodontitis is rare at younger ages and the full-mouth protocol was not administered to children or adolescents. We retained records with a complete periodontal examination status and complete data on all covariates. Participants with missing values for any covariate were excluded in a complete-case analysis. The resulting analytic sample and the sequence of exclusions are reported in the participant flow diagram (Figure 1).

Outcome

The primary outcome was total periodontitis, defined as the presence of mild, moderate, or severe periodontitis according to the CDC/AAP case definitions for population-based surveillance.(19)These definitions, developed for national monitoring rather than clinical staging, combine clinical attachment loss (CAL) and probing depth (PD) at interproximal sites and are the reference standard for epidemiological estimates in the United States. They differ from the staging and grading framework proposed in 2018 for clinical classification, and the choice of definition materially influences the resulting prevalence estimate; studies applying the 2018 framework to the same NHANES data produce different figures.(20-22)We selected the CDC/AAP definitions because they are the standard against which US surveillance estimates are computed and because they are directly operationalizable from the public-use periodontal files.

Interproximal sites are used because they are the sites at which attachment loss first accumulates in periodontitis, and the definitions count qualifying sites on distinct teeth to avoid classifying disease from repeated measurements of a single severely affected tooth.

The definitions operate on the four interproximal sites per tooth: disto-facial, mesio-facial, disto-lingual, and mesio-lingual. Severe periodontitis required at least two interproximal sites with CAL of 6 mm or more on different teeth and at least one interproximal site with PD of 5 mm or more. Moderate periodontitis required at least two interproximal sites with CAL of 4 mm or more on different teeth, or at least two interproximal sites with PD of 5 mm or more on different teeth. Mild periodontitis required at least two interproximal sites with CAL of 3 mm or more and at least two interproximal sites with PD of 4 mm or more on different teeth, or at least one site with PD of 5 mm or more. A participant was classified as having total periodontitis if any of the three severity categories applied. Measurement values coded as missing, including the cannot-measure code and sub-millimeter placeholders, were treated as missing and contributed no qualifying sites.

Predictors

Nine predictors, selected a priori from established risk indicators, operationalized each hypothesis. Sociodemographic predictors were age in years, sex, race/ethnicity (Mexican American, other Hispanic, non-Hispanic White, non-Hispanic Black, or other/multiracial), educational attainment (less than high school, high school or equivalent, some college, or college graduate), and the poverty-income ratio, a continuous index of household income relative to the federal poverty threshold. Behavioral and systemic predictors were smoking status, body mass index in kilograms per square meter, self-reported diabetes, and self-reported hypertension. Smoking status was derived from two questionnaire items: participants who reported smoking fewer than 100 cigarettes in their lifetime were classified as never smokers, those who had smoked at least 100 cigarettes but did not currently smoke were classified as former smokers, and those who reported current smoking were classified as current smokers. Borderline diabetes was classified as non-diabetic. Continuous predictors were standardized to zero mean and unit variance, and categorical predictors were one-hot encoded with the first category as the reference, so that all three models shared the same final feature matrix.

Models and validation

Three models were specified before analysis. The logistic regression used the L-BFGS solver with a maximum of 3,000 iterations. The random forest comprised 500 trees. The gradient boosting model used 300 estimators, a learning rate of 0.05, and a maximum tree depth of 3. Hyperparameters were fixed rather than tuned so that every model shared the same feature representation and no model was advantaged by data-driven optimization; this choice prioritizes a fair comparison over the maximal performance that tuning might extract from each learner. Each model was trained and evaluated with stratified 5-fold cross-validation using a fixed random seed, with identical folds across models so that the out-of-fold predictions were paired at the level of the participant. Stratification preserved the proportion of periodontitis cases in every fold, which prevents the class ratio from varying across partitions and keeps each fold a fair evaluation set. The pooled out-of-fold predicted probabilities were used for all evaluations.

The primary metric of discrimination was the AUROC, estimated with the DeLong nonparametric method together with its 95% confidence interval.⁽²³⁾ The DeLong approach accounts for the correlation between predictions from models evaluated on the same participants, which is required for a valid paired comparison. To test each hypothesis, AUROC values were compared between models with the pairwise DeLong test at a two-sided significance level of 0.05. H1 was evaluated by the gradient boosting versus logistic regression comparison, H2 by the random forest versus logistic regression comparison, and H3 by the gradient boosting versus random forest comparison. As secondary measures, sensitivity, specificity, positive predictive value, negative predictive value, and the F1 score were computed at the threshold that maximized the Youden index, defined as the sum of sensitivity and specificity minus one. This threshold balances false positives and false negatives without assuming a particular cost ratio, and the resulting operating point is reported for each model. The contribution of each predictor was quantified with the Gini importance of the two tree ensembles.

All analyses were performed with Python 3 and the scikit-learn, pandas, NumPy, and SciPy libraries, using a fixed random seed of 42. The analysis script is self-contained and was executed twice to confirm that the reported metrics did not change between runs.

Ethical considerations

The data were de-identified, publicly available, and freely downloadable from the National Center for Health Statistics website. No institutional review board approval was required, and no

individual participants were contacted.

RESULTS

Study population and periodontitis prevalence

Of 14,071 adults aged 30 years and older across the three cycles, 11,753 had a complete full-mouth periodontal examination and 10,642 had complete data on all covariates and formed the analytic sample (Figure 1). Total periodontitis was present in 4,903 participants, corresponding to a prevalence of 46.07% (Table 1). Participants with periodontitis were older than those without (mean 55.60 versus 51.20 years), had a higher proportion of males, a lower mean poverty-income ratio, a higher proportion of current smokers, and a lower proportion of college graduates. Periodontitis was most prevalent among Mexican American and non-Hispanic Black participants and least prevalent among non-Hispanic White participants (Table 1). The full distribution of sociodemographic and systemic characteristics by periodontitis status is presented in Table 1.

Discrimination of the three models

Gradient boosting achieved the highest discrimination, with an AUROC of 0.71 (95% CI 0.71 to 0.72), followed by logistic regression (AUROC 0.70; 95% CI 0.69 to 0.71) and random forest (AUROC 0.68; 95% CI 0.67 to 0.69) (Table 2). Across the five cross-validation folds, the mean AUROC was 0.71 for gradient boosting, 0.70 for logistic regression, and 0.68 for random forest, indicating that the ordering was stable across partitions rather than driven by a single fold. All three models discriminated above chance, and the differences between them, although statistically significant, were small in absolute terms. The receiver operating characteristic curves are shown in Figure 2 (Figure 2).

Hypothesis tests

H1 was supported. Gradient boosting outperformed logistic regression (AUROC 0.71 versus 0.70; $z = 5.45$, $p < 0.001$). H2 was rejected. Random forest did not exceed logistic regression and was in fact lower (AUROC 0.68 versus 0.70; $z = -3.62$, $p < 0.001$). H3 was rejected. Gradient boosting and random forest did not achieve comparable discrimination, and gradient boosting was higher ($z = 9.27$, $p < 0.001$).

Performance at the optimal threshold

At the Youden-optimal threshold, logistic regression achieved sensitivity 0.76 and specificity 0.53, random forest achieved sensitivity 0.65 and specificity 0.62, and gradient boosting achieved sensitivity 0.62 and specificity 0.70 (Table 2). The positive predictive value was 0.63 for gradient boosting, 0.58 for logistic regression, and 0.59 for random forest; the corresponding negative predictive values were 0.68, 0.72, and 0.67. F1 scores followed the same ordering as the AUROC (Table 2). The operating points show a familiar trade-off: logistic regression favoured sensitivity, gradient boosting favoured specificity, and no model attained high values of both simultaneously, which reflects the modest separation implied by an AUROC near 0.70.

Feature importance

Age was the most influential predictor in both ensembles, followed by the poverty-income ratio and body mass index. Sex, race/ethnicity, educational attainment, and smoking contributed to a lesser extent (Table 3). In the gradient boosting model, age, poverty-income ratio, female sex, non-Hispanic White race/ethnicity, and body mass index were the five strongest predictors, together accounting for most of the total importance. Diabetes, hypertension, and smoking ranked lowest in the gradient boosting model, indicating that their marginal contribution was modest once age and socioeconomic position were accounted for. The ranking of the twelve most influential predictors is illustrated in Figure 3 (Figure 3).

Table 1. Characteristics of the analytic sample overall and by periodontitis status.

Characteristic	Overall	Periodontitis	No periodontitis
N	10642	4903	5739
Age, years, mean (SD)	53.2 (14.7)	55.6 (14.0)	51.2 (15.0)
BMI, kg/m2, mean (SD)	29.3 (6.8)	29.5 (6.7)	29.1 (6.8)
Poverty-income ratio, mean (SD)	2.6 (1.6)	2.3 (1.6)	2.8 (1.7)
Sex (Male / Female)	5268 (49.5%) / 5374 (50.5%)	2864 (58.4%) / 2039 (41.6%)	2404 (41.9%) / 3335 (58.1%)
Race/ethnicity (Mexican American / Other Hispanic / Non-Hispanic White / Non-Hispanic Black / Other/Multiracial)	1364 (12.8%) / 984 (9.2%) / 4838 (45.5%) / 2226 (20.9%) / 1230 (11.6%)	838 (17.1%) / 460 (9.4%) / 1848 (37.7%) / 1214 (24.8%) / 543 (11.1%)	526 (9.2%) / 524 (9.1%) / 2990 (52.1%) / 1012 (17.6%) / 687 (12.0%)
Education (<High school / High school/GED / Some college / College graduate)	2600 (24.4%) / 2339 (22.0%) / 2965 (27.9%) / 2738 (25.7%)	1508 (30.8%) / 1232 (25.1%) / 1293 (26.4%) / 870 (17.7%)	1092 (19.0%) / 1107 (19.3%) / 1672 (29.1%) / 1868 (32.5%)
Smoking (Never / Former / Current)	5703 (53.6%) / 2820 (26.5%) / 2119 (19.9%)	2309 (47.1%) / 1371 (28.0%) / 1223 (24.9%)	3394 (59.1%) / 1449 (25.2%) / 896 (15.6%)
Diabetes (No / Yes)	9193 (86.4%) / 1449 (13.6%)	4103 (83.7%) / 800 (16.3%)	5090 (88.7%) / 649 (11.3%)
Hypertension (No / Yes)	6428 (60.4%) / 4214 (39.6%)	2780 (56.7%) / 2123 (43.3%)	3648 (63.6%) / 2091 (36.4%)

Note: Values are mean (SD) or n (%). Periodontitis = total periodontitis according to the CDC/AAP case definition.

Table 2. Predictive performance of the three models for total periodontitis.

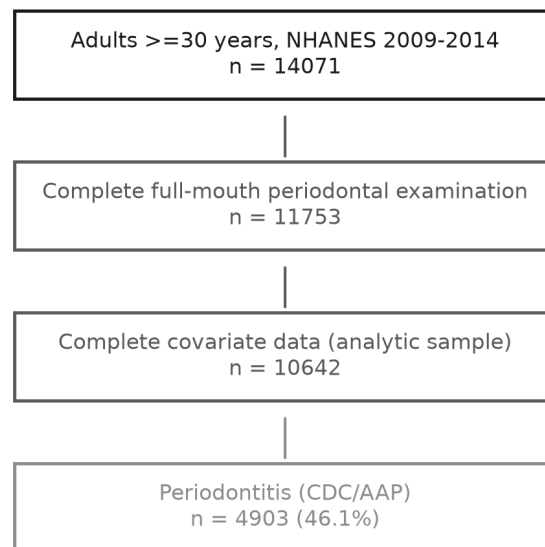
Model	AUROC	Lower 95% CI	Upper 95% CI	Sensitivity	Specificity	PPV	NPV	F1
Logistic regression	0.698	0.688	0.708	0.762	0.531	0.582	0.724	0.660
Random forest	0.683	0.673	0.693	0.650	0.618	0.592	0.674	0.620
Gradient boosting	0.715	0.705	0.724	0.616	0.696	0.633	0.679	0.624

Note: AUROC = area under the receiver operating characteristic curve; CI = confidence interval; PPV = positive predictive value; NPV = negative predictive value. Sensitivity, specificity, PPV, NPV and F1 are reported at the Youden-optimal threshold. Source: Out-of-fold predictions from stratified 5-fold cross-validation.

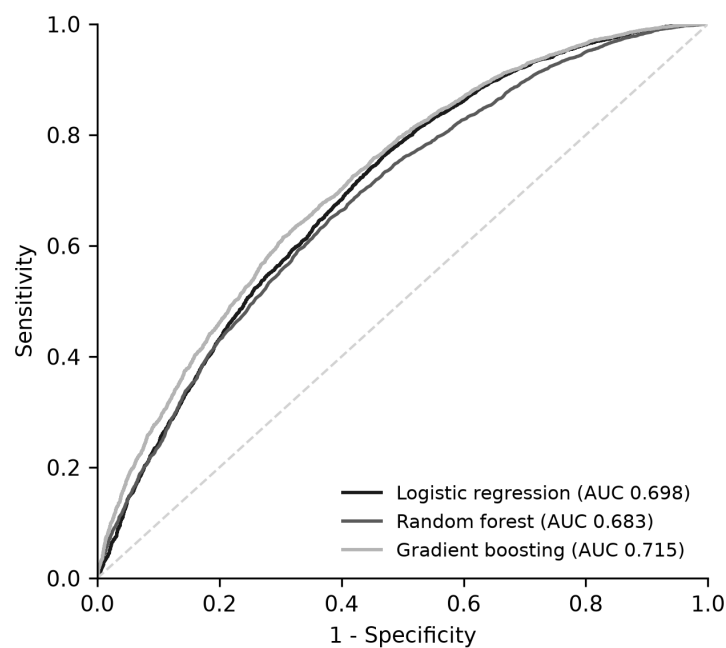
Table 3. Feature importance of the twelve most influential predictors.

Rank	Feature	Gradient boosting importance	Random forest importance
1	Age	0.315	0.239
2	Poverty-income ratio	0.170	0.234
3	Sex: female	0.128	0.033
4	Race/ethnicity: Non-Hispanic White	0.114	0.025
5	Body mass index	0.094	0.263
6	Education: College graduate	0.082	0.023
7	Smoking: current	0.048	0.022
8	Education: Some college	0.012	0.017
9	Diabetes	0.010	0.021
10	Race/ethnicity: Other Hispanic	0.008	0.012
11	Smoking: former	0.006	0.025
12	Hypertension	0.004	0.028

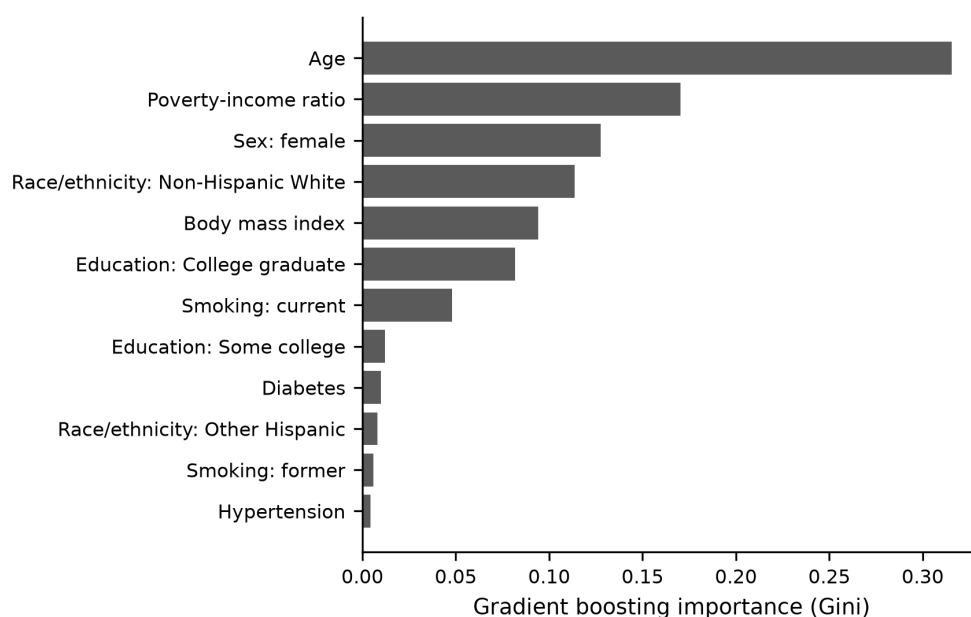
Note: Importance values are normalized to sum to 1 across all predictors. Ranked by gradient boosting importance. Source: Gini importance from the gradient boosting and random forest models.

Figure 1. Participant selection flow from NHANES 2009-2014 to the analytic sample.

Note: The analytic sample consisted of adults aged 30 years and older with a complete full-mouth periodontal examination and complete covariate data. Source: Compiled from NHANES 2009-2014.

Figure 2. Receiver operating characteristic curves of the three models.

Note: The diagonal line indicates discrimination no better than chance. Source: Out-of-fold predictions from stratified 5-fold cross-validation.

Figure 3. Feature importance of the gradient boosting model.

Note: The twelve most influential predictors are shown in descending order. Source: Gini importance from the gradient boosting model.

DISCUSSION

Gradient boosting predicted total periodontitis more accurately than logistic regression or random forest in this nationally representative sample of United States adults, whereas random forest did not improve on logistic regression. The findings resolve the three hypotheses in a consistent direction: boosting gained over the linear baseline, bagging did not, and the two tree ensembles were not equivalent. The results align with a broader pattern in which sequential boosting methods, but not bagging methods, extract a small and reliable advantage from a modest set of tabular clinical predictors.

H1 was supported. Gradient boosting exceeded logistic regression in discrimination, with an absolute difference of approximately two percentage points that was nonetheless statistically robust given the sample size. This direction agrees with prior work that identified gradient boosting or related boosting ensembles as the best-performing learners for periodontitis. In a systematic comparison of ten algorithms trained on Taiwanese and United States survey data, gradient boosting machines reached the highest AUROC, with a modest advantage over logistic regression that varied between datasets.⁽¹³⁾ Risk prediction models for severe periodontitis in a Thai population likewise found that tree-based machine learning matched or slightly exceeded statistical models after feature engineering.⁽²⁴⁾ The magnitude of the gain here is consistent with those reports and suggests that boosting offers an incremental, not a large, improvement over a well-specified logistic model for this outcome.

H2 was rejected: random forest performed below logistic regression. The result is not anomalous. Studies that compared several learners on periodontitis outcomes have reported that random forests do not consistently outperform logistic regression, particularly when the predictor set is small and the underlying associations are largely monotonic.^(25,26) A nomogram for identifying periodontitis cases in Denmark, for example, showed that a penalized linear model remained competitive with machine learning alternatives built on questionnaire data.⁽²⁶⁾ The mechanism is instructive. A random forest averages many deep trees grown on bootstrap samples, which reduces variance but does not add the sequential bias reduction that boosting provides.^(7,8) With only nine predominantly monotonic predictors, the extra flexibility of the forest offers little room for gain, and its tendency toward deeper, less regularized trees can slightly degrade discrimination on data of this kind.

H3 was also rejected: gradient boosting and random forest were not comparable, and boosting was clearly superior. The contrast between the two ensembles supports the interpretation

that the advantage of gradient boosting arises from its sequential fitting strategy rather than from tree-based flexibility alone. Comparable patterns appear across the periodontitis machine learning literature. Explainable deep and gradient-boosted models outperformed simpler ensembles for periodontitis risk screening on NHANES data.(15) Models for periodontal disease built on electronic dental records favored boosting-type learners over random forests.(14) Machine learning analyses of periodontitis with metabolic syndrome and with systemic health indicators likewise reported ensembles ahead of random forests for classification.(27,28) A model that predicted coronary heart disease risk within a periodontitis population using the same NHANES cycles also relied on ensemble methods.(29)

Beyond prediction of disease presence, machine learning has been applied to periodontal outcomes with direct clinical consequence. A model that predicted the clinical response to periodontal treatment one year after baseline trained a random forest on demographic, clinical, and microbiological features and found that baseline disease severity dominated the prediction.(30) A longitudinal cohort study of periodontitis progression likewise reported that artificial intelligence models built on routine clinical variables could anticipate worsening, with gradient-boosted ensembles among the stronger learners.(16) These applications reinforce the general conclusion that, for tabular periodontal data, the choice of learner matters less than the predictive content of the variables, and that boosting consistently ranks at or near the top across tasks.

The systematic literature on the question is consistent with these findings. A systematic review and meta-analysis of patient-based studies reported that artificial intelligence models for periodontitis diagnosis from non-invasive biological markers achieve high pooled accuracy, although the included studies varied widely in design and reporting, and few offered head-to-head comparisons against logistic regression.(12) A scoping review of artificial intelligence in periodontology likewise observed that most studies report a single algorithm rather than a controlled comparison, which limits the strength of conclusions about relative performance.(10) Our design, which holds predictors and validation constant across learners, addresses exactly that gap.

Age was the most influential predictor in both ensembles, followed by the poverty-income ratio and body mass index. This ranking mirrors established epidemiology. Age is the strongest single risk indicator of periodontitis in United States adults, and socioeconomic position, indexed here by the poverty-income ratio, tracks lifetime exposure and access to care.(5) Smoking and diabetes contributed less than their epidemiological prominence might predict, plausibly because part of their effect is captured through the age and socioeconomic gradients in this cross-sectional population. The prominence of sociodemographic over modifiable predictors is consistent with machine learning studies of periodontitis and metabolic syndrome, in which age and socioeconomic variables dominated the feature rankings.(27) Studies that incorporate biological markers or transcriptomic data have reported higher internal discrimination, but those inputs are not available in routine care, whereas the nine predictors used here are.(31) Analyses of the same NHANES periodontitis data through biological aging and osteoporosis lenses similarly confirm age and related measures as dominant.(32,33) The ranking also carries a practical message: the variables that best separate cases from non-cases are largely non-modifiable, which means that predictive models built on these inputs are better suited to risk stratification and targeted referral than to informing the specific behaviour changes that might prevent disease.

The absolute discrimination of all three models was moderate, with AUROC values near 0.70. This level is consistent with other periodontitis prediction models built from questionnaire and demographic data alone, which typically report AUROC values in the 0.60 to 0.75 range.(24,25) A value of this magnitude indicates that the models are informative for group-level risk stratification but are not sufficient for definitive individual diagnosis, which continues to require clinical examination. The value of the comparison therefore lies less in the absolute performance than in the consistent ordering, which shows that the small advantage of boosting is reproducible rather than an artefact of chance.

The choice of case definition deserves comment because it conditions every estimate in the study. The CDC/AAP definitions were designed for surveillance and are stricter in their thresholding than the 2018 staging and grading classification, so prevalence and model performance would shift if another definition were substituted.(20-22) We retained the CDC/AAP definitions to remain comparable with national surveillance, but the sensitivity of the models to the outcome definition is a recognized boundary of the findings.

The study has several strengths. It used a large, nationally representative sample with a gold-standard clinical outcome defined by the CDC/AAP case definition.(19) The three models shared the same predictors, the same cross-validation folds, and the same evaluation procedure, and differences in AUROC were tested with a paired DeLong test rather than compared informally.(23)

The analysis is deterministic and reproducible through a fixed seed and a self-contained script.

The fixed-hyperparameter design deserves emphasis. By refraining from tuning, the comparison measures the three learners under their standard defaults, which is the setting in which most clinical users would deploy them. It is possible that tuned random forests or tuned logistic models would close or reverse the observed differences, and the fixed design therefore answers a fairness question rather than a maximal-performance question. Studies that tune hyperparameters across a broad grid nevertheless reach the same ordering we observed, which supports the robustness of the conclusion.⁽¹³⁾

The limitations should temper interpretation. The design is cross-sectional, so the models estimate prevalent rather than incident periodontitis and cannot establish temporal precedence between the predictors and disease. Complete-case analysis excluded participants with missing covariate data, which may reduce representativeness if missingness was not random. Survey weights were not applied, so the sample should not be used to estimate national prevalence; the weighting decision reflects the model-comparison aim of the study rather than an inferential target. Diabetes, hypertension, and smoking were self-reported, introducing measurement error that would bias the corresponding coefficients downward. Hyperparameters were fixed rather than tuned, which is transparent but may underestimate the achievable performance of each learner. The models were validated internally by cross-validation and were not tested on an external population, so their transportability to other settings is unknown. Calibration was not formally assessed, and the reported sensitivity and specificity depend on a threshold selected from the same data, so calibration curves and decision-curve analyses on independent data would be required before deployment. Each limitation suggests a specific next step: prospective or linked longitudinal data, multiple imputation or inverse-probability weighting for missingness, tuned pipelines, external validation in independent cohorts, and calibration assessment against observed outcomes.

For clinical and public health practice, the results support gradient boosting as the preferred learner when the goal is to estimate individual periodontitis risk from variables already collected in routine care. However, the modest absolute improvement over logistic regression also implies that a well-specified logistic model remains a defensible and more interpretable alternative when transparency and ease of implementation are the priority. The dominance of age and socioeconomic predictors is itself a finding with implications: risk stratification for periodontitis is inseparable from the social determinants of oral health, and screening tools that ignore these variables will systematically miscalibrate risk across populations.⁽²⁾ The choice between the two kinds of model also carries a transparency cost. Logistic regression reports coefficients with confidence intervals that a clinician can read directly, whereas the gradient boosting model requires a feature-importance or explanation layer to be interpreted. When the accuracy gain is on the order of two percentage points of AUROC, institutions may reasonably prefer the interpretable model and reserve the boosted model for settings in which the additional discrimination justifies the added explanation burden. The practical contribution of the study is therefore not a specific model to deploy but a clear and reproducible answer to a recurring methodological question, together with a ranking of predictors that clinicians can use to judge individual risk.

CONCLUSIONS

This study compared gradient boosting, random forest, and logistic regression for predicting total periodontitis in United States adults using nine sociodemographic and systemic predictors from NHANES 2009-2014. Gradient boosting achieved the highest discrimination and significantly outperformed logistic regression and random forest. Random forest did not improve on logistic regression, and the two tree ensembles were not equivalent. Age, poverty-income ratio, sex, race/ethnicity, and body mass index were the most influential predictors. The findings indicate that boosting methods offer a small but reliable advantage for periodontitis risk prediction from routine data, while logistic regression remains a reasonable and interpretable baseline. The concentration of predictive information in age and socioeconomic position also shows that accurate risk stratification requires attention to the social determinants of oral health as much as to the choice of algorithm. Future work should extend the comparison to longitudinal outcomes, tuned pipelines, and external validation to establish whether the observed ordering persists when the models are used for incident disease and in populations beyond the United States.

REFERENCES

1. Trindade D, Carvalho R, Machado V, Chambrone L, Mendes JJ, Botelho J. Prevalence of periodontitis in dentate people between 2011 and 2020: A systematic review and meta-analysis of epidemiological studies. *J Clinic Periodontology*. 2023;50(5):604-626. doi: 10.1111/jcpe.13769
2. GBD 2017 Oral Disorders Collaborators, Bernabe E, Marcenes W, Hernandez CR, Bailey J, Abreu LG, 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):362-373. doi: 10.1177/0022034520908533
3. Eke PI, Dye BA, Wei L, Thornton-Evans GO, Genco RJ. Prevalence of Periodontitis in Adults in the United States: 2009 and 2010. *J Dent Res*. 2012;91(10):914-920. doi: 10.1177/0022034512457373
4. Tonetti MS, Van Dyke TE, working group 1 of the joint EFP/AAP workshop. Periodontitis and atherosclerotic cardiovascular disease: consensus report of the Joint EFP/AAP Workshop on Periodontitis and Systemic Diseases. *J Clinic Periodontology*. 2013;40(s14):S24-9. doi: 10.1111/jcpe.12089
5. Eke PI, Wei L, Thornton-Evans GO, Borrell LN, Borgnakke WS, Dye B, et al. Risk Indicators for Periodontitis in US Adults: NHANES 2009 to 2012. *Journal of Periodontology*. 2016;87(10):1174-1185. doi: 10.1902/jop.2016.160013
6. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162(1):55-63. doi: 10.7326/m14-0697
7. Breiman L. Random Forests. *Machine Learning*. 2001;45(1):5-32. doi: 10.1023/a:1010933404324
8. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Statist*. 2001;29(5):1189-1232. doi: 10.1214/aos/1013203451
9. Ding H, Wu J, Zhao W, Matinlinna JP, Burrow MF, Tsoi JKH. Artificial intelligence in dentistry—A review. *Front Dent Med*. 2023;4:1085251. doi: 10.3389/fdmed.2023.1085251
10. Scott J, Biancardi AM, Jones O, Andrew D. Artificial Intelligence in Periodontology: A Scoping Review. *Dentistry Journal*. 2023;11(2):43. doi: 10.3390/dj11020043
11. Pitchika V, Büttner M, Schwendicke F. Artificial intelligence and personalized diagnostics in periodontology: A narrative review. *Periodontology 2000*. 2024;95(1):220-231. doi: 10.1111/prd.12586
12. Jafer M, Ibraheem W, Dawood T, Abbas A, Hakami K, Khurayzi T, et al. The Application and Performance of Artificial Intelligence (AI) Models in the Diagnosis, Classification, and Prediction of Periodontal Diseases: A Systematic Review. *Diagnostics*. 2025;15(24):3247. doi: 10.3390/diagnostics15243247
13. Bashir NZ, Rahman Z, Chen SL. Systematic comparison of machine learning algorithms to develop and validate predictive models for periodontitis. *J Clinic Periodontology*. 2022;49(10):958-969. doi: 10.1111/jcpe.13692
14. Patel JS, Su C, Tellez M, Albandar JM, Rao R, Iyer V, et al. Developing and testing a prediction model for periodontal disease using machine learning and big electronic dental record data. *Front Artif Intell*. 2022;5:979525. doi: 10.3389/frai.2022.979525
15. Suh B, Yu H, Cha JK, Choi J, Kim JW. Explainable Deep Learning Approaches for Risk Screening of Periodontitis. *J Dent Res*. 2025;104(1):45-53. doi: 10.1177/00220345241286488
16. Furquim CP, Caruth L, Chandrasekaran G, Cucchiara A, Kallan MJ, Martin L, et al. Developing Predictive Models for Periodontitis Progression Using Artificial Intelligence: A Longitudinal Cohort Study. *J Clinic Periodontology*. 2025;52(10):1478-1490. doi: 10.1111/jcpe.14194
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*. 2007;370(9596):1453-1457. doi: 10.1016/s0140-6736(07)61602-x
18. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey (NHANES), 2009-2014 [Internet]. 2020. Available from: <https://wwwn.cdc.gov/nchs/nhanes/>
19. Eke PI, Page RC, Wei L, Thornton-Evans G, Genco RJ. Update of the Case Definitions for Population-Based Surveillance of Periodontitis. *Journal of Periodontology*. 2012;83(12):1449-1454. doi: 10.1902/jop.2012.110664
20. Ke L, Nogueira G, Thomson WM. Influence of case definitions on epidemiological estimates of periodontitis prevalence and its associations with smoking and OHRQoL. *Comm Dent Oral Epid*. 2023;51(2):194-200. doi: 10.1111/cdoe.12726
21. Tay JRH, Holtfreter B, Baumeister S, Peres MA, Nascimento GG. Application of the 2018 Periodontal Status Classification to Epidemiological Survey Data (ACES) Framework to Estimate the Periodontitis Prevalence in the United States. *J Clinic Periodontology*. 2025;52(7):1032-1043. doi: 10.1111/jcpe.14132
22. Botelho J, Machado V, Proença L, Mendes JJ. The 2018 periodontitis case definition improves accuracy performance of full-mouth partial diagnostic protocols. *Sci Rep*. 2020;10(1):7093. doi: 10.1038/s41598-020-63700-6
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. 1988;44(3):837. doi: 10.2307/2531595
24. Teza H, Pattanateepapon A, Lertpimonchai A, Vathesatogkit P, J McKay G, Attia J, et al. Development of Risk Prediction Models for Severe Periodontitis in a Thai Population: Statistical and Machine Learning Approaches. *JMIR Form Res*. 2023;7:e48351. doi: 10.2196/48351
25. Enevold C, Nielsen CH, Christensen LB, Kongstad J, Fiehn NE, Hansen PR, et al. Suitability of machine learning models for prediction of clinically defined Stage III/IV periodontitis from questionnaires and demographic data in Danish cohorts. *J Clinic Periodontology*. 2024;51(12):1561-1573. doi: 10.1111/jcpe.13874
26. Cassiano LS, Jensen AB, Pajaniaye JB, Leite FRM, Li H, Andersen A, et al. Development of a nomogram for identifying periodontitis cases in Denmark. *Sci Rep*. 2024;14(1):11280. doi: 10.1038/s41598-024-60624-3

27. Wilensky A, Frank N, Mizraji G, Tzur D, Goldstein C, Almozni G. Periodontitis and Metabolic Syndrome: Statistical and Machine Learning Analytics of a Nationwide Study. *Bioengineering*. 2023;10(12):1384. doi: 10.3390/bioengineering10121384
28. Yan Y, Sharma P, Suvan J, D'Aiuto F. The Association of Periodontal Inflammation and Systemic Health Indicators: A Machine Learning Approach. *J Clin Periodontology*. 2025;52(10):1466-1477. doi: 10.1111/jcpe.70000
29. Wang Y, Ni B, Xiao Y, Lin Y, Jiang Y, Zhang Y. Application of machine learning algorithms to construct and validate a prediction model for coronary heart disease risk in patients with periodontitis: a population-based study. *Front Cardiovasc Med*. 2023;10:1296405. doi: 10.3389/fcvm.2023.1296405
30. Feher B, de Souza Oliveira EH, Mendes Duarte P, Werdich AA, Giannobile WV, Feres M. Machine learning-assisted prediction of clinical responses to periodontal treatment. *Journal of Periodontology*. 2025;96(11):1199-1212. doi: 10.1002/jper.24-0737
31. Kim EH, Kim S, Kim HJ, Jeong HO, Lee J, Jang J, et al. Prediction of Chronic Periodontitis Severity Using Machine Learning Models Based On Salivary Bacterial Copy Number. *Front Cell Infect Microbiol*. 2020;10:571515. doi: 10.3389/fcimb.2020.571515
32. Li S, Wen C, Bai X, Yang D. Association between biological aging and periodontitis using NHANES 2009–2014 and mendelian randomization. *Sci Rep*. 2024;14(1):10089. doi: 10.1038/s41598-024-61002-9
33. Peng J, Chen J, Liu Y, Lyu J, Zhang B. Association between periodontitis and osteoporosis in United States adults from the National Health and Nutrition Examination Survey: a cross-sectional analysis. *BMC Oral Health*. 2023;23(1):254. doi: 10.1186/s12903-023-02990-4

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Maria Fernanda Escalante Freitez.
Writing–review and editing: Maria Fernanda Escalante Freitez.