Journal of Dental Research
1–11

DOI: 10.1177/00220345261473819
journals.sagepub.com/home/jdr

Comparative Diagnostic Performance of Oral Microbial Habitats for Periodontitis

I. Tomás[1,2], B. Suárez-Rodríguez[1], T. Blanco-Pintos[1], A. Sánchez-Barco[1], M. Alonso-Sampedro[3], M.J. Carreira[2,4], A. Regueira-Iglesias[1], and C. Balsa-Castro[1,2]

Abstract

Periodontitis affects more than 1 billion people worldwide but remains largely underdiagnosed. The oral microbiota offers a noninvasive diagnostic window, but studies based on 16S rRNA data have reported heterogeneous and often suboptimal performance. It remains unclear which oral habitat and modeling strategy best support clinical decision making. We analyzed 2,303 adult samples from 33 publicly available 16S rRNA V3-V4 Illumina projects encompassing saliva, supragingival, and subgingival habitats. Amplicon sequence variants (ASVs) were identified against the expanded Human Oral Microbiome Database. Compact microbial signatures were identified through multivariate techniques (sparse partial least squares discriminant analysis and a customized genetic algorithm). Five machine-learning algorithms were benchmarked under repeated cross-validation and test-set evaluation. Diagnostic performance was assessed using the area under the receiver-operating characteristic curve (AUC), and clinical net benefit was assessed using decision curve analysis (DCA). Generalized additive models (GAMs) provided the best discrimination with compact panels of 6 to 12 ASVs. Saliva and supragingival achieved the best discrimination (AUC = 0.928 and 0.903; sensitivity = 92.4% and 87.6%; specificity = 85.3% and 81.7%), significantly outperforming subgingival plaque (0.803, 78.2% and 75.1%, respectively; adjusted $P \leq 0.001$). DCA further demonstrated the highest and most stable net clinical benefit in saliva (AUC-DCA = 0.437), followed by supragingival (0.419) and subgingival (0.331). Forty-eight taxonomically annotated ASVs emerged as predictors, underscoring the models' biological interpretability. In subgingival sites, the most prevalent features were health-associated commensals, such as *Streptococcus oralis* subsp. *dentisani*. Supragingival plaque and saliva predominantly revealed periodontitis-associated taxa, including *Porphyromonas gingivalis* and *Filifactor alocis*. Thus, saliva and supragingival plaque better capture the microbial fingerprint of periodontitis-associated dysbiosis. Oral microbiome–based models across habitats showed consistent diagnostic performance for periodontitis. The GAM algorithm with compact, interpretable microbial signatures achieved strong discrimination and superior net benefit in saliva and supragingival samples, highlighting its potential for translation into scalable, noninvasive diagnostic tools for precision periodontics.

Keywords: periodontal diseases, microbiota, saliva, dental plaque, machine learning, biomarkers

Introduction

Periodontitis is a plaque-biofilm–associated, host-mediated inflammatory disease that results in the progressive destruction of the tooth-supporting apparatus (Papapanou et al 2018). It is among the most prevalent chronic inflammatory diseases worldwide. In 2021, more than 1 billion people were affected, and projections indicate a 44% increase by 2050 (Nascimento et al 2024). Its chronic nature and systemic implications—including diabetes, cardiovascular disease, and adverse pregnancy outcomes—underscore its impact beyond oral health (Bacali et al 2022).

The subgingival region is traditionally regarded as the primary ecological habitat for dysbiotic microbial communities in periodontitis. However, supragingival plaque and saliva also exhibit disease-related microbial alterations, reflecting broader ecological shifts across the oral cavity (Ji et al 2023; Wen et al 2025). These findings highlight the diagnostic potential of the oral microbiota.

Artificial intelligence (AI)–based machine learning (ML) can harness this potential by classifying samples, identifying disease-associated taxa beyond conventional abundance testing (Regueira-Iglesias et al 2024), and quantifying feature importance to improve interpretability (Wu et al 2021).

[1]Oral Sciences Research Group, Special Needs Unit, Department of Surgery and Medical-Surgical Specialties, School of Medicine and Dentistry, Universidade de Santiago de Compostela, Health Research Institute of Santiago (IDIS), Santiago de Compostela, Spain
[2]Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain
[3]Department of Internal Medicine and Clinical Epidemiology, Complejo Hospitalario Universitario, Health Research Institute of Santiago (IDIS), Santiago de Compostela, Spain
[4]Departamento de Electrónica e Computación, Escola Técnica Superior de Enxeñaría, Universidade de Santiago de Compostela, Health Research Institute of Santiago (IDIS), Santiago de Compostela, Spain

A supplemental appendix to this article is available online.

Corresponding Author:
I. Tomás, Alba Regueira Iglesias, School of Medicine and Dentistry, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain.
Email: inmaculada.tomas@usc.es

Despite this promise, diagnostic models for periodontitis based on 16S rRNA data have shown highly variable performance, using anywhere from 1 (Diao et al 2021) to more than 200 predictors (Na et al 2020). Reported areas under the curve (AUCs) range from 0.973 in saliva (Diao et al 2021) to 0.629 in supragingival plaque (Na et al 2020), while accuracy varies from 94.8% in subgingival plaque (Chen et al 2014) to 79.5% in saliva (Lundmark et al 2019). This heterogeneity reflects methodological limitations: small sample sizes, limited feature selection, and insufficient validation.

Clinically useful models, therefore, require rigorous methodology, heterogeneous populations, and robust validation to ensure generalizability (Steyerberg 2019; Collins et al 2024; Tsegaye et al 2025).

Only a few investigations have compared different ML algorithms (Chen et al 2018; Na et al 2020) or oral habitats (Chen et al 2015; Na et al 2020; Chew et al 2025), and just one has assessed both, limited to buccal mucosa and supragingival plaque (Na et al 2020). These studies rarely evaluate algorithms under optimized conditions, which may confound their conclusions. Consequently, the most informative type of oral sample and algorithmic configuration for clinically meaningful diagnosis remain unclear.

Genetic algorithms (GAs), inspired by Darwinian evolution, are well-suited to high-dimensional omics data because they simultaneously explore complex feature interactions. They have been successfully applied for feature selection and optimization in microbiome studies (Fu et al 2023; Meng et al 2023) but never across multiple oral habitats for diagnosing periodontitis. Beyond algorithmic optimization, however, clinical applicability must be considered, as diagnostic models are valuable only if they improve decision-making in practice. Decision curve analysis (DCA) addresses this gap by quantifying the net clinical benefit of model predictions across relevant probability thresholds (Vickers et al 2016). While DCA has been applied in intestinal microbiome models (Wang et al 2024), no previous studies have implemented it in oral microbiome–based ML for periodontitis, leaving a gap in the evaluation of clinical utility.

Therefore, this study aimed to identify the best-performing combination of ML algorithm and oral microbial habitat (subgingival, supragingival, and salivary) for diagnosing periodontitis using publicly available 16S rRNA data. We adhered to best-practice guidelines for meta-barcoding and diagnostic prediction modeling (Steyerberg 2019; Regueira-Iglesias et al 2023; Collins et al 2024; Tsegaye et al 2025) and implemented a customized GA for model optimization. Model performance was assessed using both conventional metrics and clinical net benefit analysis to provide a comprehensive framework for evaluating diagnostic value. Ultimately, this work seeks to bridge the gap between AI-driven microbiome research and clinical decision-making in oral health, highlighting saliva as a noninvasive and scalable diagnostic matrix for population-level screening.

Methods

The study workflow is shown in Figure 1. This study integrated publicly available 16S rRNA datasets from adults classified as periodontally healthy or as having periodontitis, covering 3 oral microbial habitats: subgingival plaque, supragingival plaque, and saliva.

All selected projects applied a clinical or combined clinical–radiographic diagnosis of periodontal status and used Illumina sequencing of the 16S rRNA gene V3–V4 region.

In this study, "saliva" refers to whole salivary samples, stimulated or unstimulated, as reported by the original authors. Dental plaque samples were likewise obtained using different procedures, such as curettes or paper points.

A total of 2,303 samples were analyzed, including 819 from healthy individuals and 1,484 from patients with periodontitis (151 + 714 subgingival, 200 + 457 supragingival, and 468 + 313 saliva). Details of the search strategy, inclusion criteria, quality assessment, study characteristics, and sample collection procedures are provided in the Appendix Materials.

Sequences were quality filtered and denoised using USEARCH (v10) (Edgar 2010) and processed with mothur (Schloss et al 2009) to infer amplicon sequence variants (ASVs). Taxonomy was assigned through the expanded Human Oral Microbiome Database developed by Escapa et al (2020). The analysis environment was implemented in R 4.4.1 (Bioconductor 3.19). After filtering rare features, abundance tables were centered log-ratio transformed using the mixOmics package (v6.28.0) (Rohart et al 2017).

Supervised classification by sparse partial least squares discriminant analysis (sPLS-DA) was applied independently to each oral habitat to reduce dimensionality and identify the most discriminative ASVs. This multivariate approach accounts for multicollinearity and identifies nonredundant microbial predictors in high-dimensional data.

A customized GA implemented with the GA package (v3.2.4) (Scrucca 2013) was used to refine the feature subsets selected by sPLS-DA. GA optimization evaluates multiple feature interactions simultaneously, favoring compact and high-performing subsets of ASVs. Each oral habitat was analyzed under GA-optimized configurations to ensure that performance differences reflected intrinsic diagnostic capacity rather than tuning bias.

Five supervised ML algorithms were benchmarked using the caret framework (Kuhn 2008): generalized additive model (GAM), generalized linear model (GLM), high-dimensional discriminant analysis (HDDA), naïve Bayes (NB), and random forest (RF). Data were stratified into training (two-thirds) and test (one-third) sets. Repeated cross-validation (20 × 3-fold) was applied within training sets to prevent overfitting. To address class imbalance, synthetic interpolation (CutMix) (Yun et al 2019) and bootstrap resampling were implemented. Final training and test sample sizes, along with the initial sample sizes per clinical group, are presented in Appendix Table S3.

Model selection combined conventional discrimination metrics with a penalty for model complexity and unbalanced sensitivity/specificity, prioritizing parsimonious models with stable diagnostic performance (≤5 percentage-point difference between sensitivity and specificity).

Receiver-operating characteristic (ROC) curves were generated for each algorithm, and pairwise comparisons were

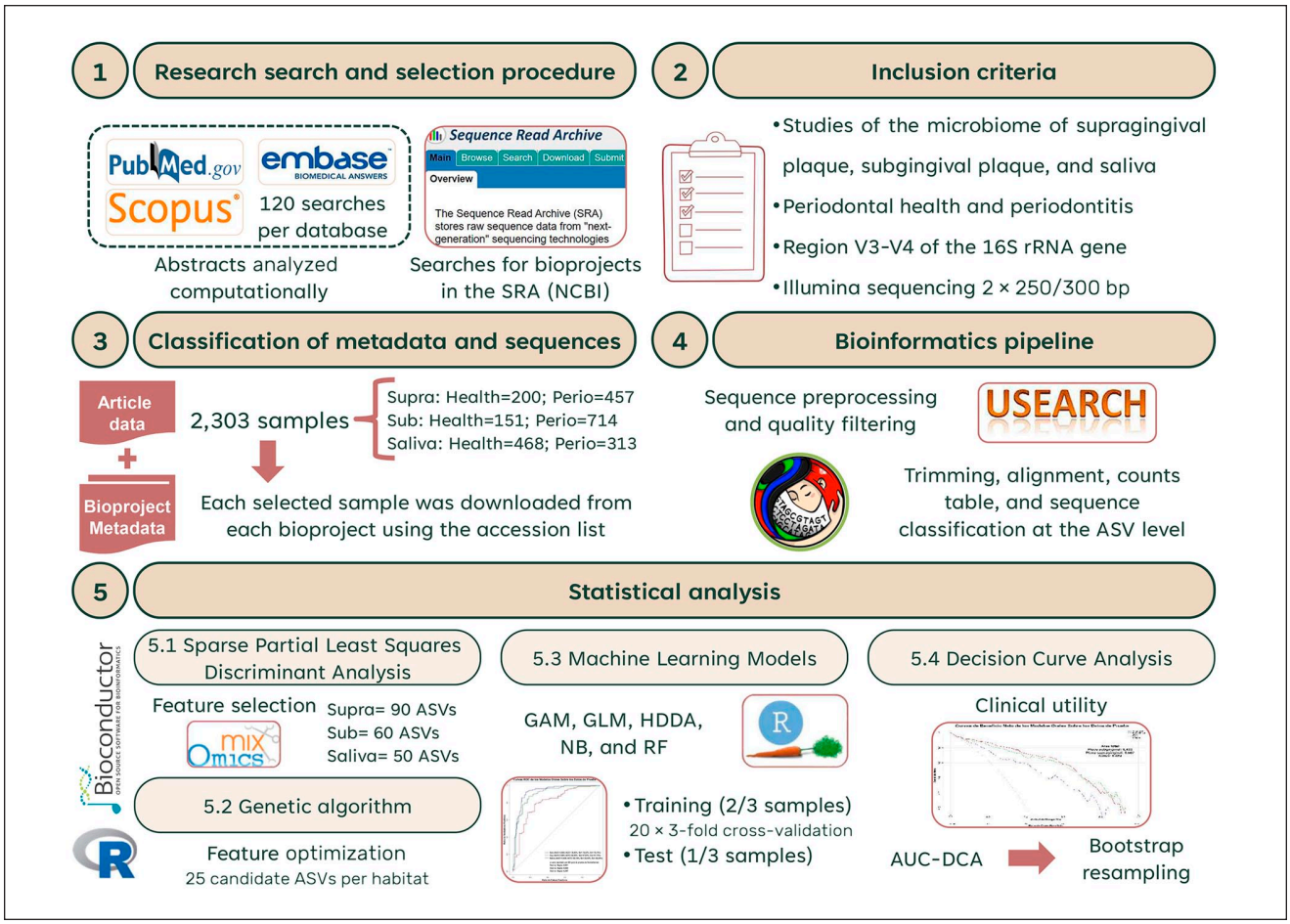

Figure 1. Schematic representation of the study workflow.

performed using the Venkatraman permutation test, which evaluates overall differences in ROC curve shape rather than AUC alone. Statistical significance was set at $P < 0.05$ after Benjamini–Hochberg correction for multiple testing.

DCA was used to estimate the net clinical benefit of each model across relevant probability thresholds, implemented with the rmda package (Brown 2018). The area under the decision curve (AUC-DCA) quantified the overall benefit relative to "assess-all" and "assess-none" strategies. Bootstrap resampling was used to estimate 95% confidence intervals for pairwise differences in AUC-DCA.

ASVs repeatedly selected by top-performing models across habitats were examined to confirm biological consistency with their relative abundances in health and disease, highlighting shared and habitat-specific microbial signatures of periodontitis. The bioinformatician responsible for preprocessing and model development was blinded to the clinical meaning of the verified outcome labels.

Detailed methodological procedures are provided in the Appendix Materials. The study followed TRIPOD+AI guidelines (Collins et al 2024), with the completed checklist in Appendix Figure S1.

Results

Characteristics of the Included Investigations

The selection process identified 33 bioprojects fulfilling the inclusion criteria (Appendix Fig S2). A detailed summary of diagnostic definitions, severity grading, and methodological quality for each bioproject is provided in Appendix Tables S1 and S2 and Appendix Figure S3.

Predictive Performance of ML Models in Oral Habitats

In subgingival plaque, models retained 7 to 12 ASVs and showed the weakest performance (test AUC 0.738–0.822), with sensitivities ranging from 55.5% (GLM) to 81.1% (NB) and specificities ranging from 70.7% (NB) to 83.9% (GLM). GAM demonstrated the highest accuracy (ACC, 76.5%) and the best balance between sensitivity and specificity (78.2% and 75.1%, respectively). Compared to the training set, RF suffered the sharpest decline (−13.2 pp AUC; −12.2 pp sensitivity), suggesting overfitting, and GLM underperformed, with sensitivity loss (−10.7 pp) (Appendix Table S4).

In supragingival plaque, models retained 7 to 8 ASVs (test AUC 0.847–0.903), with sensitivities ranging from 78.4% (RF) to 87.6% (GAM) and specificities ranging from 75.4% (GLM) to 87.4% (RF). GAM demonstrated the highest ACC (84.5%) and a strong balance between sensitivity (87.6%) and specificity (81.7%). In comparison with the training set, only RF overfitted, with a sensitivity loss (−13.7 pp) despite improved specificity (+5.8 pp) (Appendix Table S5).

In saliva, models retained 6 to 12 ASVs and achieved the best performance (test AUC 0.895–0.928), with sensitivities ranging from 78.1% (RF) to 92.4% (GAM) and specificities ranging from 74.4% (NB) to 87.8% (RF). GAM demonstrated the highest ACC value (88.1%) and a balance between sensitivity and specificity (92.4% and 85.3%, respectively), prioritizing the detection of positive cases while maintaining good performance in identifying negatives. In comparison with the training set, no model showed a change greater than 10 pp. The highest change was observed in NB, characterized by a loss of specificity (−7.0 pp) (Appendix Table S6).

Across the 3 habitats in the training set ROC curves, GAM was the best-performing model, showing significant differences relative to the other algorithms, except in the subgingival habitat, where GAM and RF exhibited similar curves. In the test set, ROC curves for the subgingival habitat confirmed that GLM was consistently inferior to HDDA and NB. In the supragingival habitat, ROC comparisons showed that GAM and NB significantly outperformed GLM. However, in saliva, no pairwise differences were detected between the algorithms (Figs 2A, B; 3A, B; 4A, B).

Cross-Habitat Comparisons

GAM was selected for direct comparison across habitats due to its consistent balance and net benefit. In the test set, saliva achieved the highest test AUC (0.928, −3.0 pp), followed by supragingival (0.903, −4.8 pp) and subgingival (0.803, −9.9 pp). Sensitivity and specificity were higher in saliva (92.4% and 85.3%) than in supragingival (87.6% and 81.7%) and subgingival (78.2% and 75.1%) (Appendix Table S7).

ROC comparisons confirmed that both saliva and supragingival plaque outperformed subgingival plaque (adjusted $P \leq 0.001$), with no significant difference between them (Fig. 5B). Additional performance metrics are provided in Appendix Tables S4 to S6.

Net Clinical Benefit of ML Models in Oral Habitats

In the test set of subgingival plaque, net benefit peaked at 0.872 (pt = 0.1) and remained positive at pt = 0.7 except for GLM; RF showed the lowest positive value (0.170). Across thresholds, AUC-DCA ranged from 0.274 (GLM) to 0.331 (GAM), while the assess-all strategy was much lower (0.187). All ML models significantly outperformed the assess-all approach; GAM was significantly better than GLM, while GLM was significantly inferior to HDDA, NB, and RF (Fig 2C, D).

In the test set of supragingival plaque, the net benefit ranged from 0.917 (pt = 0.1) to 0.307 (pt = 0.7), with AUC-DCA values of 0.366 (GLM) to 0.419 (GAM). All models significantly outperformed the assess-all strategy (0.188); GAM showed a significant advantage over GLM and HDDA, while GLM performed worse than NB (Fig 3C, D).

In the test set of saliva, the net benefit remained highest and most stable (0.915 at pt = 0.1; 0.330 at pt = 0.7). AUC-DCA values ranged from 0.357 (NB) to 0.437 (GAM), with the assess-all value again being significantly lower (0.143). GAM significantly outperformed HDDA, NB, and RF, while GLM outperformed NB and RF. HDDA was superior to NB (Fig 4C, D).

Cross-Habitat Comparisons (DCA)

Cross-habitat comparisons using GAM confirmed that saliva was the most clinically useful habitat (AUC-DCA 0.437), followed by supragingival (0.419) and subgingival (0.331) habitats. Across threshold probabilities of 0.1 to 0.7, net benefit values for saliva and supragingival sites remained numerically higher than those for subgingival sites. Saliva significantly outperformed the subgingival habitat (Fig 5C, D).

Predictive ASVs Associated with Periodontal Health and Disease across Oral Habitats

Across the 3 oral habitats, a total of 48 unique ASVs were identified as predictors of periodontal status. The table reports each ASV's taxonomic identity, habitat, number of selecting models, and relative abundance in the health and periodontitis groups (Appendix Table S8).

In the subgingival habitat, 14 exclusive ASVs were detected, with most (10/14) linked to periodontal health. Representative examples included *Haemophilus parainfluenzae*, *Gemella haemolysans*, and *Streptococcus oralis* subsp. *dentisani*, all consistently enriched in healthy samples. Fewer but relevant disease-associated taxa were also found; both *Desulfobulbus* sp. HMT041 and *Peptostreptococcaceae bacterium* HMT369 were selected by multiple models.

In supragingival plaque, 6 exclusive ASVs were identified, and all were associated with disease. Among them, *Treponema denticola*, *Streptococcus anginosus*, and *Fretibacterium fastidiosum* appeared consistently enriched in periodontitis samples.

The salivary habitat yielded 12 exclusive ASVs, of which 11 were associated with periodontitis. Notable examples included *Campylobacter rectus*, *Fusobacterium nucleatum* subsp. *vincentii*, *Porphyromonas endodontalis*, and *Prevotella* spp. The only salivary taxon linked to health was *S. oralis* subsp. *dentisani*.

Finally, 8 ASVs were shared across habitats, including robust biomarkers such as *F. nucleatum* subsp. *vincentii* and *Tannerella forsythia*, consistently associated with periodontitis

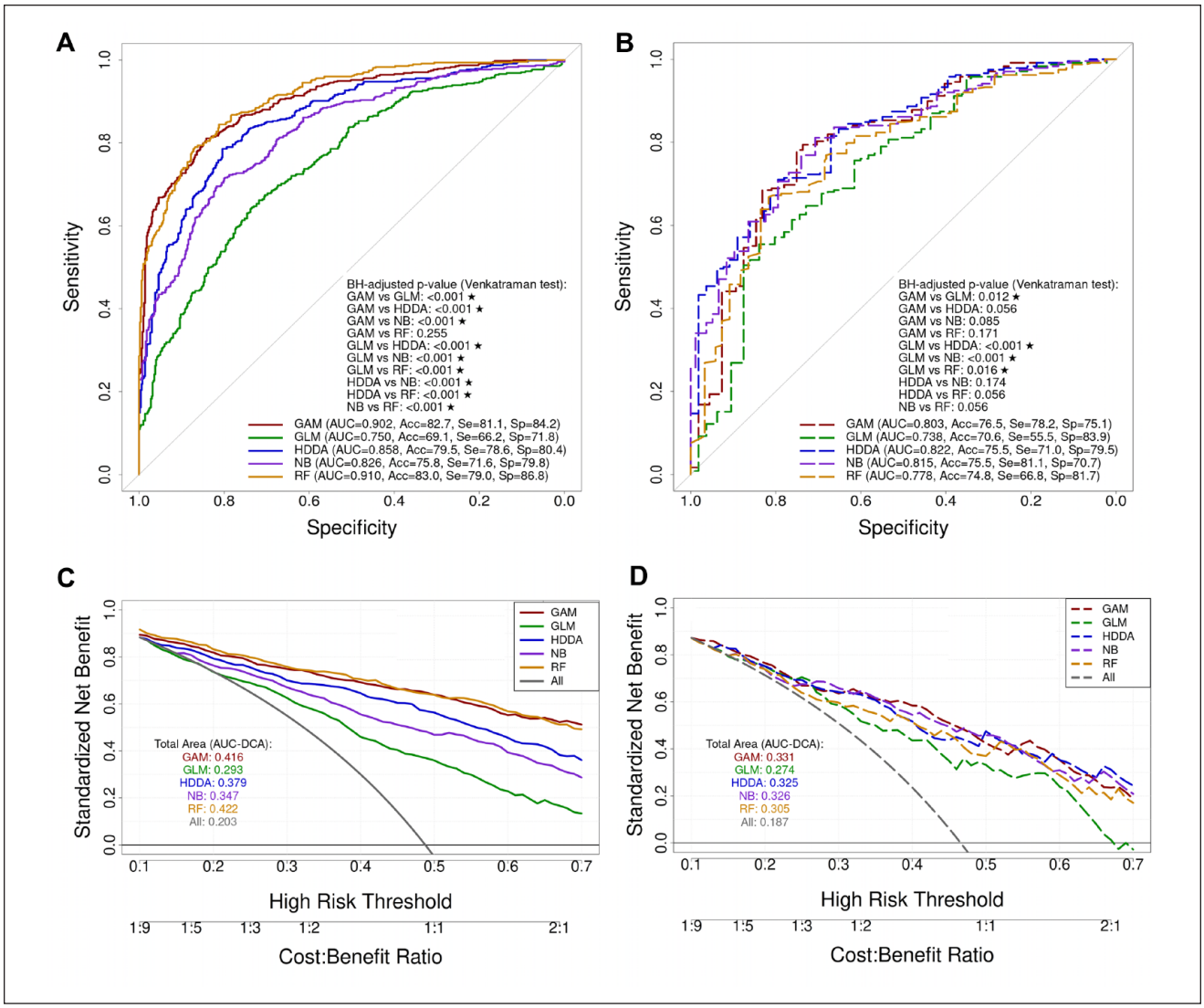

Figure 2. Receiver-operating characteristic (ROC) and decision curve analyses (DCAs) of machine-learning models based on the subgingival microbiota for periodontitis diagnosis. (**A**, **B**) ROC curves in the training and test sets, including performance parameters and pairwise comparisons between models. (**C**, **D**) DCAs for the training and test sets, showing the area under the decision curve (AUC–DCA) values."All" denotes the assess-all strategy, corresponding to comprehensive periodontal assessment of all individuals; the horizontal line at zero net benefit represents the assess-none strategy. Solid lines represent training sets, while dashed lines represent test sets. GAM, generalized additive model; GLM, generalized linear model; HDDA, high-dimensional discriminant analysis; NB, naïve Bayes; NPV, negative predictive value; PPV, positive predictive value; RF, random forest. Significant comparisons in the AUC-DCA values in the test set were found between assess-all and GAM (−0.178, −0.108), assess-all and GLM (−0.121, −0.059), assess-all and HDDA (−0.174, −0.106), assess-all and NB (−0.174, −0.108), and assess-all and RF (−0.151, −0.084). In addition, significant comparisons were observed between GAM and GLM (0.010, 0.089), GLM and HDDA (−0.077, −0.016), GLM and NB (−0.075, −0.014), and GLM and RF (−0.049, −0.001). All other pairwise comparisons were not statistically significant, as their bootstrap-derived confidence intervals included zero.

and identified in all 3 habitats. Other cross-habitat disease-associated taxa included *Porphyromonas gingivalis*, *Filifactor alocis*, and *Dialister invisus*, reinforcing their potential as generalizable indicators of disease across the oral microbiota.

Discussion

AI is not yet routine in dental diagnostics due to limited data availability, variable dataset quality, and methodological inconsistencies (Schwendicke et al 2020). Following best-practice recommendations for diagnostic prediction models (Collins et al 2024), we compiled 2,303 oral microbiota samples from 33 bioprojects. Feature selection combined sPLS-DA with a customized GA, a rarely implemented yet highly effective strategy for high-dimensional microbiome data (Fu et al 2023; Meng et al 2023). GA optimization favored compact, high-performing subsets, while held-out testing identified performance differences among subgingival, supragingival,

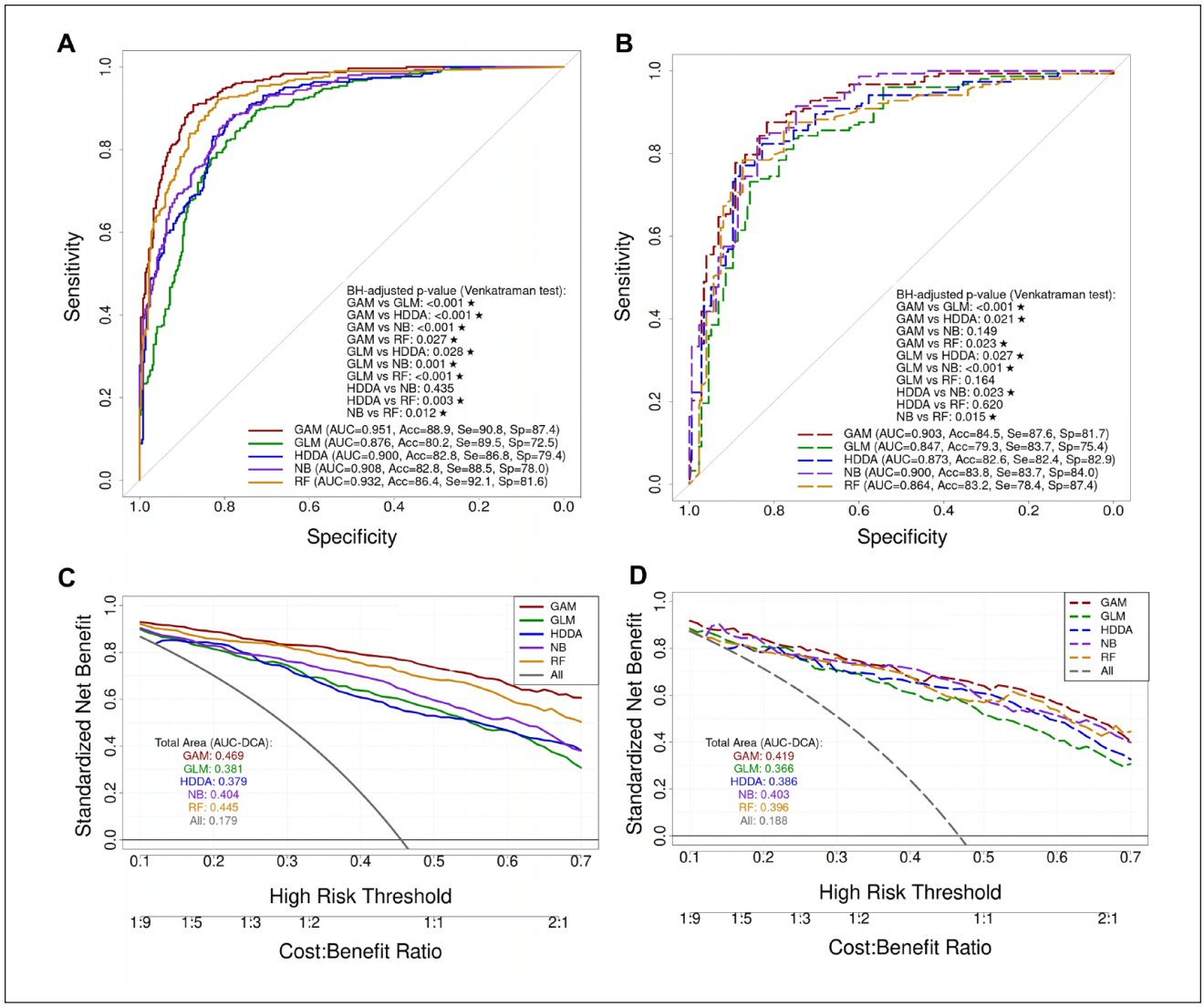

Figure 3. Receiver-operating characteristic (ROC) and decision curve analyses (DCAs) of machine-learning models based on the supragingival microbiota for periodontitis diagnosis. (**A**, **B**) ROC curves in the training and test sets, including performance parameters and pairwise comparisons between models. (C, D) DCAs for the training and test sets, showing the area under the decision curve (AUC–DCA) values."All" denotes the assess-all strategy, corresponding to comprehensive periodontal assessment of all individuals; the horizontal line at zero net benefit represents the assess-none strategy. Solid lines represent training sets, while dashed lines represent test sets. GAM, generalized additive model; GLM, generalized linear model; HDDA, high-dimensional discriminant analysis; NB, naïve Bayes; NPV, negative predictive value; PPV, positive predictive value; RF, random forest. Significant comparisons in the area under the curve (AUC)–DCA values in the test set were found between assess-all and GAM (−0.274, −0.186), assess-all and GLM (−0.220, −0.136), assess-all and HDDA (−0.244, −0.150), assess-all and NB (−0.258, −0.169), and assess-all and RF (−0.254, −0.168). In addition, significant comparisons were observed between GAM and GLM (0.010, 0.085), GAM and HDDA (0.003, 0.061), and GLM and NB (−0.063, −0.012). All other pairwise comparisons were not statistically significant, as their bootstrap-derived confidence intervals included zero.

and salivary habitats. To prevent overfitting, we used repeated cross-validation, test set evaluation, and class balancing (Gordon-Rodriguez et al 2022), ensuring transparent and reproducible model validation in accordance with TRIPOD+AI standards (Collins et al 2024).

Although several studies have applied ML to microbiome-based diagnosis of periodontitis, few have systematically compared multiple oral habitats or algorithms within a unified framework (Chen et al 2014, 2015, 2018; Na et al 2020; Chew et al 2025). Previous investigations remain limited by small cohorts, lack of independent validation, and methodological heterogeneity (Appendix Table S9). Our models achieved comparable or superior diagnostic performance while addressing these limitations through rigorous design and test-set evaluation (Chen et al 2014; Na et al 2020). A notable advantage is the use of compact predictive subsets (6 to 12 variables), among the smallest reported, enhancing interpretability and clinical applicability. In contrast, prior studies such as Chen et al (2014, 2015) and Chew et al (2025) used larger or unspecified feature sets, thereby increasing the risk of overfitting.

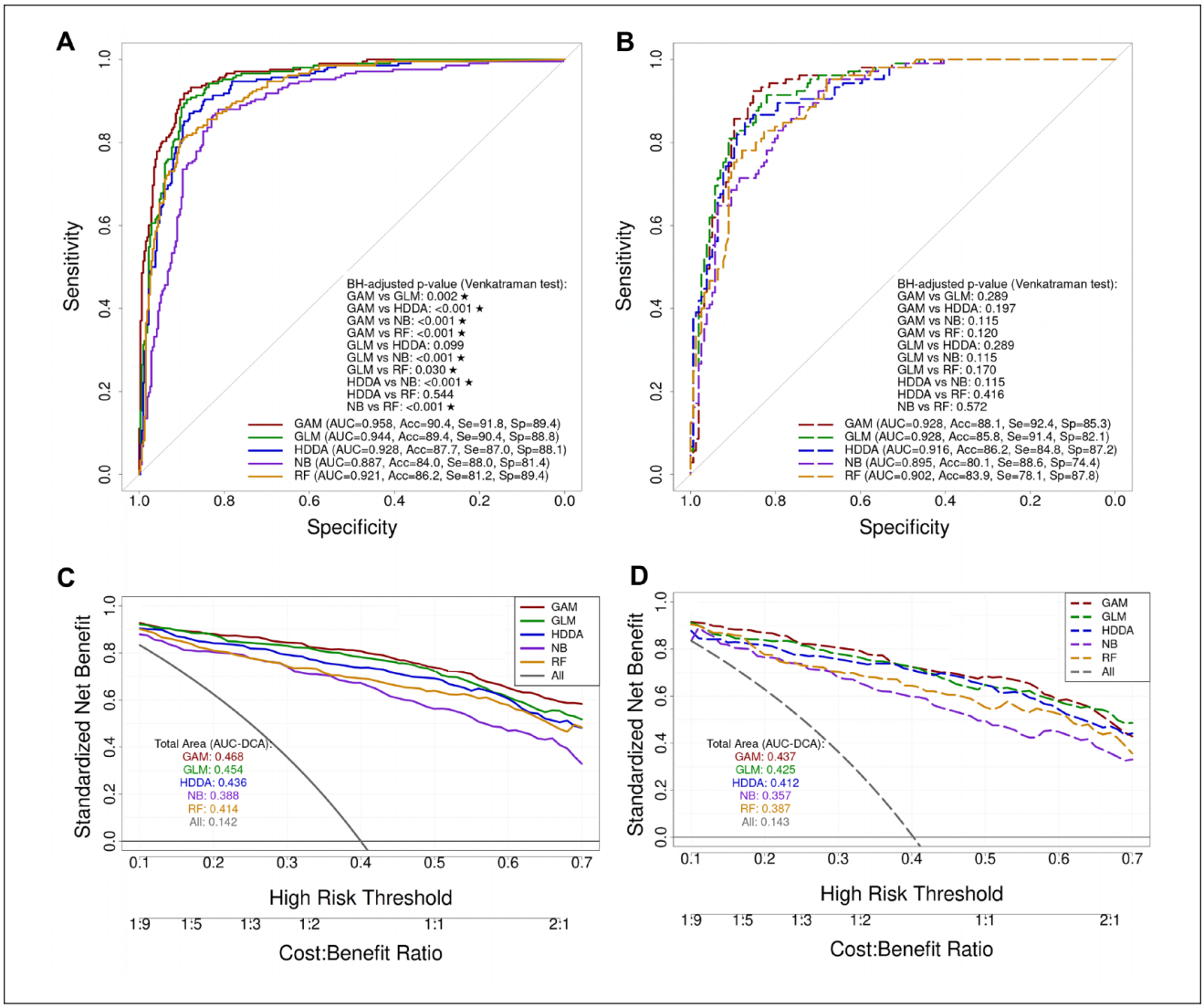

Figure 4. Receiver-operating characteristic (ROC) and decision curve analyses (DCAs) of machine-learning models based on the salivary microbiota for periodontitis diagnosis. (**A**, **B**) ROC curves in the training and test sets, including performance parameters and pairwise comparisons between models. (**C**, **D**) DCAs for the training and test sets, showing the area under the decision curve (AUC–DCA) values. "All" denotes the assess-all strategy, corresponding to comprehensive periodontal assessment of all individuals; the horizontal line at zero net benefit represents the assess-none strategy. Solid lines represent training sets, while dashed lines represent test sets. GAM, generalized additive model; GLM, generalized linear model; HDDA, high-dimensional discriminant analysis; NB, naïve Bayes; NPV, negative predictive value; PPV, positive predictive value; RF, random forest. Significant comparisons in the AUC-DCA values in the test set were found between assess-all and GAM (−0.346, −0.241), assess-all and GLM (−0.334, −0.233), assess-all and HDDA (−0.329, −0.216), assess-all and NB (−0.268, −0.160), and assess-all and RF (−0.297, −0.190). In addition, significant comparisons were observed between GAM and HDDA (0.000, 0.054), GAM and NB (0.042, 0.123), GAM and RF (0.013, 0.092), GLM and NB (0.030, 0.110), GLM and RF (0.001, 0.083), and HDDA and NB (0.028, 0.095). All other pairwise comparisons were not statistically significant, as their bootstrap-derived confidence intervals included zero.

Compared with earlier reports employing RF, SVM, NB, or deep learning (Chen et al 2014, 2018; Na et al 2020), our findings show that statistically significant differences between algorithms largely disappear on test sets, particularly in saliva. This attenuation confirms that apparent superiority during cross-validation often reflects model fitting rather than true predictive capacity (Collins et al 2024). Among the algorithms tested, the GAM provided the most consistent balance between accuracy and clinical benefit—a novel observation, as no previous studies compared this model across habitats. Conversely, RF, despite being widely regarded as a strong performer (Chen et al 2018), exhibited overfitting tendencies, consistent with its known sensitivity to local probability peaks and limited generalizability in large, diverse datasets (Barreñada et al 2024).

Our best-performing models reached excellent to outstanding discrimination according to Hosmer et al (2013). Salivary models achieved AUCs up to 0.928 (GAM, GLM) with sensitivities up to 92.4%, and specificities up to 87.8%, outperforming Chen et al (2015) (AUC 0.817) and Chew et al (2025) (AUC 0.882). This confirms saliva's superior diagnostic

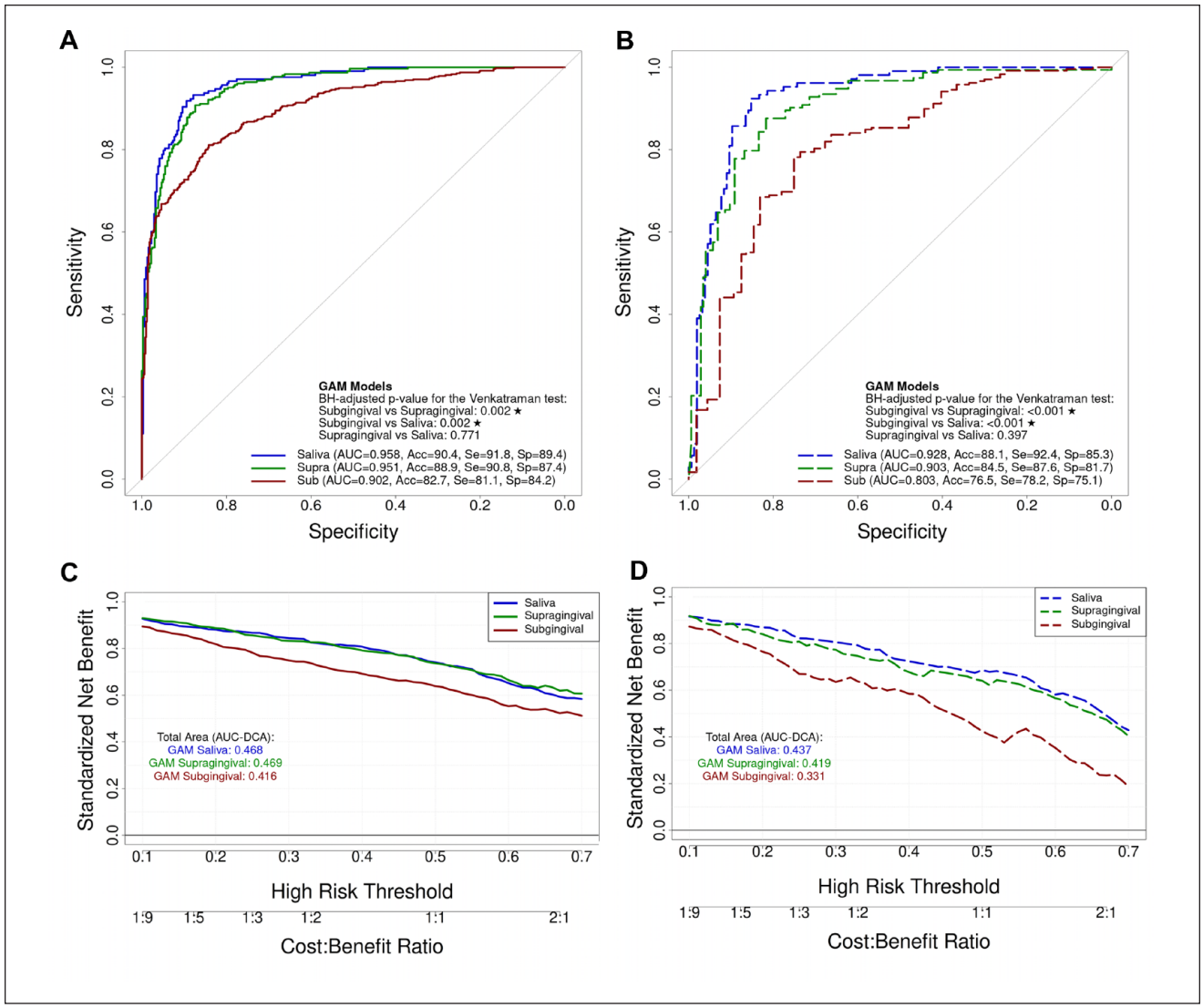

Figure 5. Receiver-operating characteristic (ROC) and decision curve analyses (DCAs) of generalized additive models (GAMs) based on subgingival, supragingival, and salivary microbiotas for periodontitis diagnosis. (**A**, **B**) ROC curves in the training and test sets, including performance parameters and pairwise comparisons between oral habitats. (**C**, **D**) DCAs for the training and test sets, showing the area under the decision curve (AUC–DCA) values for the oral habitats. The assess-all strategy is not shown; the horizontal line at zero net benefit represents the assess-none strategy. Solid lines represent training sets, while dashed lines represent test sets. Significant differences in AUC-DCA values in the test set were observed between saliva and subgingival (0.023, 0.200). All other pairwise comparisons were not statistically significant, as their bootstrap-derived confidence intervals included zero.

potential, likely due to its integration of microbial signals across the oral cavity (Jung et al 2024). To our knowledge, this is the first demonstration that saliva—and, to a lesser extent, supragingival plaque—outperforms subgingival plaque as a diagnostic matrix for periodontitis, as evidenced by significant differences in ROCs between habitats.

DCA evaluates the clinical benefit of predictive models beyond conventional performance metrics (Vickers et al 2016). Although ML has been used for oral microbiome–based diagnosis of periodontitis (Chen et al 2014, 2015, 2018; Na et al 2020; Chew et al 2025), none have assessed clinical utility. This study represents the first application of DCA to ML models of the oral microbiota for periodontitis, integrating AUC-DCA quantification and bootstrap-based comparisons (Zhang et al 2018). All algorithms outperformed the "assess-all" strategy, with GAM showing the highest net benefit—particularly in saliva—while GLM performed weakest in plaque-based models and NB in saliva. Salivary (AUC-DCA 0.437) and supragingival (0.419) models demonstrated meaningful clinical utility, both of which surpassed subgingival plaque (0.331). Only the difference between saliva and subgingival habitats was statistically significant (95% CI: 0.023–0.200), confirming saliva's superior applicability.

The identification of 48 taxonomically annotated ASVs highlights ML models' ability to extract biologically meaningful predictors from complex oral microbiota data with minimal variable sets. These findings support periodontitis as a dysbiotic condition, with predictive signals reflecting community-level microbial shifts rather than individual pathogens (Cui et al 2025).

In the subgingival habitat, most exclusive ASVs (10/14) were unexpectedly linked to periodontal health, including *Haemophilus parainfluenzae* and *Streptococcus oralis* subsp. *dentisani*; the latter has been investigated as an oral probiotic (Ferrer et al 2020). This likely reflects the ecological depletion of commensals in disease, making residual healthy taxa the most distinctive features detected by parsimonious models (Cui et al 2025). Conversely, supragingival and salivary habitats predominantly yielded disease-associated ASVs, providing clearer contrasts between health and periodontitis. Key pathobionts included *Treponema denticola*, *Fusobacterium nucleatum* subsp. *vincentii*, and *Porphyromonas endodontalis*, consistently implicated in periodontal pathology (Cui et al 2025). This divergence suggests that oral habitats distant from the periodontal pocket—particularly saliva—more effectively capture the integrated microbial fingerprint of oral dysbiosis than the pathobiont-enriched subgingival environment.

Clinically, the performance and net benefit of salivary and supragingival models support these habitats as minimally invasive matrices for microbiome-based periodontitis screening, particularly when comprehensive periodontal examination is not readily available (Ji et al 2023; Jung et al 2024; Regueira-Iglesias et al 2024). Supragingival plaque may serve as an alternative to saliva for site-specific biofilm assessment, whereas subgingival models may remain more relevant for advanced or site-oriented periodontal evaluation. Compact, interpretable microbial signatures (6 to 12 ASVs) may facilitate translation into clinical assays while maintaining high diagnostic accuracy (Steyerberg 2019; Collins et al 2024). Combined with clinical parameters, these models could support risk stratification and personalized decision-making, while DCA reinforces their role as adjunctive tools in precision periodontics across clinically relevant thresholds (Vickers et al 2016; Zhang et al 2018).

Within this translational framework, preanalytical and clinical variability should be considered as inherent sources of heterogeneity in integrated datasets (Knight et al 2018).

In conventional microbiome studies, differences in sampling procedures or clinical characteristics may affect taxonomic composition and biological interpretation (Camelo-Castillo et al 2015; Beyer et al 2017; Gomar-Vercher et al 2018). In contrast, ML-based predictive modeling prioritizes classification performance and may benefit from heterogeneous data when rigorous validation is applied (Steyerberg 2019; Collins et al 2024; Balendran et al 2025; Marconi and Cabitza 2025). In this context, although source bioprojects were unevenly distributed across habitats, the analytical datasets were relatively comparable in sample size (865 subgingival, 657 supragingival, and 781 salivary samples), and models were developed separately for each habitat. Therefore, the observed performance differences are unlikely to be explained by sample-size imbalance alone. Importantly, major technical sources of bias were controlled in this study by standardizing the targeted 16S rRNA region, sequencing platform, and bioinformatic processing (Regueira-Iglesias et al 2023). The strong performance observed in the test sets, particularly for saliva, supports the robustness of the predictive signals despite variability in sampling procedures and clinical conditions (Steyerberg 2019; de Hond et al 2023). Future studies with larger datasets and standardized metadata should further assess the impact of specific covariates, such as sampling technique, disease stage, smoking status, or systemic conditions (Collins et al 2024; Tsegaye et al 2025).

Notwithstanding these considerations, 2 limitations should be acknowledged. First, the 3 oral habitats were analyzed in different individuals, precluding intraindividual comparisons and making interindividual variability difficult to separate from habitat-specific differences (Hall et al 2017). Second, although disease-severity thresholds were broadly comparable across studies, the most recent standardized diagnostic criteria (Papapanou et al 2018) were applied in only one-third of them, potentially affecting disease classification and data harmonization.

Despite extensive internal validation, external replication is required to confirm generalizability and to assess model calibration, cost-effectiveness, and incremental value relative to clinical predictors (de Hond et al 2023).

Conclusions

This study presents the first systematic benchmarking of machine-learning algorithms across subgingival, supragingival, and salivary microbiotas for diagnosing periodontitis. Integrating a GA with ML and DCA provided a unified analytical framework that may be transferable to other diagnostic applications.

Oral microbiome–based models across habitats showed consistent diagnostic performance for periodontitis. GAMs achieved the best balance between discrimination and clinical utility. Parsimonious models using only 6 to 12 ASVs captured both periodontitis-associated pathobionts (*Fusobacterium nucleatum* subsp. *vincentii* and *Tannerella forsythia*) and health-associated commensal (*Streptococcus oralis* subsp. *dentisani*), supporting the biological interpretability of the models. Saliva and supragingival plaque showed higher diagnostic performance than subgingival plaque and emerged as promising habitats for noninvasive microbiome-based periodontitis diagnosis.

Author Contributions

I. Tomás, C. Balsa-Castro, contributed to conception and design, data analysis and interpretation, drafted and critically revised the manuscript; B. Suárez-Rodríguez, contributed to acquisition and interpretation, drafted the manuscript; T. Blanco-Pintos, A. Sánchez-Barco, M. Alonso-Sampedro, contributed to acquisition, critically revised the manuscript; M.J. Carreira, contributed to analysis and

interpretation, critically revised the manuscript; A. Regueira-Iglesias, contributed to acquisition and interpretation, drafted and critically revised the manuscript. All authors gave their final approval and agree to be accountable for all aspects of the work.

Ethics Statement

Not applicable. This study used only previously published and publicly available datasets obtained from international repositories (NCBI SRA, ENA, DDBJ). No new human or animal data were collected.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study has been funded by the Instituto de Salud Carlos III (ISCIII) via project number PI24/00222 and was co-funded by the European Union (EU); Xunta de Galicia - Consellería de Educación, Ciencia, Universidades y Formación Profesional (grant ED431B GPC2024/05); and Nueva Galimplant S.L.U. (project reference: 2025-CE011), which provided support for personnel costs. The funders had no role in the study's design, data collection and analysis, decision to publish, or manuscript preparation.

ORCID iDs

I. Tomás https://orcid.org/0000-0002-3317-0853
B. Suárez-Rodríguez https://orcid.org/0009-0003-6044-0606
T. Blanco-Pintos https://orcid.org/0000-0002-4186-7714
A. Sánchez-Barco https://orcid.org/0009-0001-2926-7572
M. Alonso-Sampedro https://orcid.org/0000-0002-0916-4672
M.J. Carreira https://orcid.org/0000-0003-0532-2351
A. Regueira-Iglesias https://orcid.org/0000-0002-6549-7738
C. Balsa-Castro https://orcid.org/0000-0001-7937-6673

Data Availability

All sequencing datasets analyzed in this study are publicly available through international repositories, including the NCBI Sequence Read Archive (SRA), the European Nucleotide Archive (ENA), and the DNA Data Bank of Japan (DDBJ). The complete list of BioProject accession numbers corresponding to the datasets used is provided in Appendix Table S1. These resources ensure that other researchers can independently reproduce the analyses presented in this work.

References

Bacali C et al. 2022. Oral microbiome: getting to know and befriend neighbors, a biological approach. Biomedicines. 10(3):671. https://doi.org/10.3390/biomedicines10030671

Balendran A et al. 2025. A scoping review of robustness concepts for machine learning in healthcare. NPJ Digit Med. 8(1):38. https://doi.org/10.1038/s41746-024-01420-1

Barreñada L, Dhiman P, Timmerman D, Boulesteix AL, Van Calster B. 2024. Understanding overfitting in random forest for probability estimation: a visualization and simulation study. Diagn Progn Res. 8(1):14. https://doi.org/10.1186/s41512-024-00177-1

Beyer K et al. 2017. Comparison of paperpoint and curette sampling of subgingival microbiome composition as analyzed by 16S rRNA gene amplicon sequencing. J Oral Microbiol. 9(Suppl 1):1325260. https://doi.org/10.1080/20002297.2017.1325260

Brown M. 2018. rmda: risk model decision analysis. R package version 1.6. R Foundation for Statistical Computing; [accessed 2025 Mar]. https://CRAN.R-project.org/package=rmda

Camelo-Castillo AJ et al. 2015. Subgingival microbiota in health compared to periodontitis and the influence of smoking. Front Microbiol. 6:119. https://doi.org/10.3389/fmicb.2015.00119

Chen H et al. 2015. A *Filifactor alocis*-centered co-occurrence group associates with periodontitis across different oral habitats. Sci Rep. 5:9053. https://doi.org/10.1038/srep09053

Chen W, Cheng YM, Zhang SW, Pan Q. 2014. Supervised method for periodontitis phenotypes prediction based on microbial composition using 16S rRNA sequences. Int J Comput Biol Drug Des. 7(2–3):214–224. https://doi.org/10.1504/IJCBDD.2014.061647

Chen WP et al. 2018. Composition analysis and feature selection of the oral microbiota associated with periodontal disease. Biomed Res Int. 2018:3130607. https://doi.org/10.1155/2018/3130607

Chew RJJ, Tan KS, Chen T, Al-Hebshi NN, Goh CE. 2025. Quantifying periodontitis-associated oral dysbiosis in tongue and saliva microbiomes—an integrated data analysis. J Periodontol. 96(1):55–66. https://doi.org/10.1002/JPER.24-0120

Collins GS et al. 2024. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 385:e078378. https://doi.org/10.1136/bmj-2023-078378

Cui Z, Wang P, Gao W. 2025. Microbial dysbiosis in periodontitis and peri-implantitis: pathogenesis, immune responses, and therapeutic. Front Cell Infect Microbiol. 15:1517154. https://doi.org/10.3389/fcimb.2025.1517154

de Hond AAH et al. 2023. Perspectives on validation of clinical predictive algorithms. NPJ Digit Med. 6(1):86. https://doi.org/10.1038/s41746-023-00832-9

Diao J et al. 2021. Potential roles of the free salivary microbiome dysbiosis in periodontal diseases. Front Cell Infect Microbiol. 11:711282. https://doi.org/10.3389/fcimb.2021.711282

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26(19):2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Escapa IF et al. 2020. Construction of habitat-specific training sets to achieve species-level assignment in 16S rRNA gene datasets. Microbiome. 8(1):65. https://doi.org/10.1186/s40168-020-00841-w

Ferrer MD et al. 2020. Topic application of the probiotic *Streptococcus dentisani* improves clinical and microbiological parameters associated with oral health. Front Cell Infect Microbiol. 10:465. https://doi.org/10.3389/fcimb.2020.00465

Fu G et al. 2023. Feature selection with a genetic algorithm can help improve the distinguishing power of microbiota information in monozygotic twins' identification. Front Microbiol. 14:1210638. https://doi.org/10.3389/fmicb.2023.1210638

Gomar-Vercher S, Simón-Soro A, Montiel-Company JM, Almerich-Silla JM, Mira A. 2018. Stimulated and unstimulated saliva samples have significantly different bacterial profiles. PLoS One. 13(6):e0198021. https://doi.org/10.1371/journal.pone.0198021

Gordon-Rodriguez E, Quinn TP, Cunningham JP. 2022. Data augmentation for compositional data: advancing predictive models of the microbiome. Adv Neural Inf Process Syst. 35:20551–20565.

Hall MW et al. 2017. Inter-personal diversity and temporal dynamics of dental, tongue, and salivary microbiota in the healthy oral cavity. NPJ Biofilms Microbiomes. 3:2. https://doi.org/10.1038/s41522-016-0011-0

Hosmer DW Jr, Lemeshow S, Sturdivant RX. 2013. Applied logistic regression. 3rd ed. John Wiley & Sons.

Ji S et al. 2023. Characteristics of the salivary microbiota in periodontal diseases and potential roles of individual bacterial species to predict severity of periodontal disease. Microbiol Spectr. 11(3):e0432722. https://doi.org/10.1128/spectrum.04327-22

Jung J-S et al. 2024. Salivary microbiota reflecting changes in subgingival microbiota. Microbiol Spectr. 12(11):e0103024. https://doi.org/10.1128/spectrum.01030-24

Knight R et al. 2018. Best practices for analysing microbiomes. Nat Rev Microbiol. 16(7):410–422. https://doi.org/10.1038/s41579-018-0029-9

Kuhn M. 2008. Building predictive models in R using the caret package. J Stat Softw. 28(5):1–26. https://doi.org/10.18637/jss.v028.i05

Lundmark A et al. 2019. Identification of salivary microbiota and its association with host inflammatory mediators in periodontitis. Front Cell Infect Microbiol. 9:216. https://doi.org/10.3389/fcimb.2019.00216

Marconi L, Cabitza F. 2025. Show and tell: a critical review on robustness and uncertainty for a more responsible medical AI. Int J Med Inform. 202:105970. https://doi.org/10.1016/j.ijmedinf.2025.105970

Meng Y, Duan Q, Jiao K, Xue J. 2023. A screened predictive model for esophageal squamous cell carcinoma based on salivary flora data. Math Biosci Eng. 20(10):18368–18385. https://doi.org/10.3934/mbe.2023816

Na HS et al. 2020. Identification of potential oral microbial biomarkers for the diagnosis of periodontitis. J Clin Med. 9(5):1549. https://doi.org/10.3390/jcm9051549

Nascimento GG, Alves-Costa S, Romandini M. 2024. Burden of severe periodontitis and edentulism in 2021, with projections up to 2050: the Global Burden of Disease 2021 study. J Periodontal Res. 59(5):823–867. https://doi.org/10.1111/jre.13337

Papapanou PN et al. 2018. Periodontitis: consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. J Clin Periodontol. 45(Suppl 20):S162–S170. https://doi.org/10.1111/jcpe.12946

Regueira-Iglesias A, Balsa-Castro C, Blanco-Pintos T, Tomás I. 2023. Critical review of 16S rRNA gene sequencing workflow in microbiome studies: from primer selection to advanced data analysis. Mol Oral Microbiol. 38(5):347–399. https://doi.org/10.1111/omi.12434

Regueira-Iglesias A et al. 2024. The salivary microbiome as a diagnostic biomarker of periodontitis: a 16S multi-batch study before and after the removal of batch effects. Front Cell Infect Microbiol. 14:1405699. https://doi.org/10.3389/fcimb.2024.1405699

Rohart F, Gautier B, Singh A, Lê Cao K-A. 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Comput Biol. 13(11):e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Schloss PD et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 75(23):7537–7541. https://doi.org/10.1128/AEM.01541-09

Schwendicke F, Samek W, Krois J. 2020. Artificial intelligence in dentistry: chances and challenges. J Dent Res. 99(7):769–774. https://doi.org/10.1177/0022034520915714

Scrucca L. 2013. GA: a package for genetic algorithms in R. J Stat Softw. 53(4):1–37. https://doi.org/10.18637/jss.v053.i04

Steyerberg EW. 2019. Clinical prediction models: a practical approach to development, validation, and updating. 2nd ed. Springer.

Tsegaye B et al. 2025. Larger sample sizes are needed when developing a clinical prediction model using machine learning in oncology: methodological systematic review. J Clin Epidemiol. 180:111675. https://doi.org/10.1016/j.jclinepi.2025.111675

Vickers AJ, Van Calster B, Steyerberg EW. 2016. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 352:i6. https://doi.org/10.1136/bmj.i6

Wang J et al. 2024. Gut microbiome-based noninvasive diagnostic model to predict acute coronary syndromes. Front Cell Infect Microbiol. 13:1305375. https://doi.org/10.3389/fcimb.2023.1305375

Wen CC et al. 2025. Full-length 16S rRNA sequencing reveals microbial characteristics in supragingival plaque of periodontitis patients. J Dent Sci. 20(3):1739–1748. https://doi.org/10.1016/j.jds.2025.03.040

Wu S et al. 2021. Towards multi-label classification: next step of machine learning for microbiome research. Comput Struct Biotechnol J. 19:2742–2749. https://doi.org/10.1016/j.csbj.2021.04.054

Yun S et al. 2019. CutMix: regularization strategy to train strong classifiers with localizable features [preprint]. arXiv. 1905.04899. https://doi.org/10.48550/arXiv.1905.04899

Zhang Z et al. 2018. Decision curve analysis: a technical note. Ann Transl Med. 6(15):308. https://doi.org/10.21037/atm.2018.07.02.