Predicting Periodontal Disease Using Interpretable Machine Learning Models

Joana M. A. Lopes[1], Ana C. Braga[1(✉)], and Marta M. M. Relvas[2]

[1] ALGORITMI Research Centre, LASI, University of Minho, Braga, Portugal
acb@dps.uminho.pt
[2] UNIPRO IUCS CESPU, Gandra, Portugal
marta.relvas@iucs.cespu.pt

Abstract. As well known that periodontal disease is a chronic inflammatory condition and constitutes a major cause of tooth loss globally. An early identification of individuals at risk remains challenging due to the complex interplay among clinical, behavioural, systemic, and biological factors. The present study evaluates supervised Machine Learning models for predicting periodontal disease by integrating multiple data types, with a particular focus on clinical interpretability.

A dataset with 189 participants containing variables as clinical measures of periodontal health, oral hygiene routines, smoking status, diabetes status, and salivary inflammatory marker concentrations was analysed. Several classification algorithms were assessed, including Naive Bayes, Decision Trees, Random Forests, Support Vector Machines, K-Nearest Neighbours, and Logistic Regression. Model performance was evaluated using accuracy, sensitivity, specificity, F1-score, and area under the receiver operating characteristic (ROC) curve.

From this study, Logistic Regression exhibited robust predictive performance and interpretability, supporting its potential utility in clinical decision-making. Some key predictors were find significant as periodontal indicators, behavioural risk factors, and inflammatory biomarkers, particularly IL-1β, IL-6, and IL-8. These results indicate that interpretable Machine Learning approaches may enhance the identification of periodontal risk and inform personalised prevention strategies.

Keywords: Periodontal Disease · Machine Learning · Risk Factors · Inflammatory Biomarkers · Clinical Decision Support

1 Introduction

Periodontal disease (PD) is a chronic inflammatory condition of the tooth-supporting structures and remains an important public health concern. Although advances in preventive care have been made, it is still a leading cause of tooth loss, affecting quality of life, and could be linked to systemic medical problems. So, early diagnosis and risk stratification are important, and traditional diagnostics often focus only on isolated factors and may not reflect the disease's complexity.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16759, pp. 119–132, 2027.
https://doi.org/10.1007/978-3-032-30518-3_8

Recent advances in data science and machine learning (ML) allow for integration among diverse clinical and biological data, affording new opportunities for healthcare prediction. In periodontology, ML is being investigated to improve diagnosis and risk-factor identification, but its broad clinical use is limited by challenges related to interpretability and transparency.

The main objective of our study is to identify machine learning models for predicting periodontal disease that demonstrate both robust predictive accuracy and practical clinical utility. By combining clinical parameters, behavioural factors, systemic conditions, and salivary inflammatory biomarkers, we expect to improve risk assessment and support more effective clinical decision-making.

This manuscript is organized into seven main sections. Section 1 introduces concepts relating to periodontal disease and highlights the importance of interpretable machine learning models for clinical risk stratification. Section 2 provides a brief review of the literature on aetiological factors, salivary biomarkers and supervised algorithms relevant to periodontology. Sections 3 and 4 describe the procedure, including the dataset of 189 individuals, data pre-processing and comparisons between six algorithms. Section 5 presents the model results and the contribution of features, demonstrating the superior performance of Logistic Regression. Finally, Sects. 6 and 7 discuss the clinical implications of the selected model, as well as the limitations and directions for future research.

2 Review of the Literature

As described in [28], periodontal disease is a chronic, multifactorial inflammatory condition that affects the tissues supporting the teeth: the gingiva, the periodontal ligament, the root cementum and the alveolar bone. Initiated by the accumulation of bacterial biofilm on tooth surfaces, this process triggers an immune-inflammatory response in the host. Its progression and severity are influenced by risk factors, including genetic susceptibility, smoking, systemic conditions such as diabetes, and local factors such as dental anatomy and defective restorations [28]. Thus, PD is best seen as a host-biofilm interaction, where the disease trajectory depends on the balance between microbial challenge and host response [9,21]. Rather than a single dominant pathogen, PD involves a dysbiotic shift in biofilm composition. The oral microbiome moves from mainly Gram-positive aerobic species to Gram-negative anaerobes, thereby altering community behaviour and sustaining chronic inflammation [14]. This microbiome represents a dynamic ecosystem in which hundreds of species interact and adapt to changes in nutrient levels, mechanical stress, and pH. Understanding this complexity is key to creating effective prevention and treatment strategies for periodontal diseases [12].

Oral healthcare services and public awareness of oral hygiene have improved in Europe over the past two decades. However, the prevalence of periodontitis has remained stable, which is concerning, as it is a leading cause of tooth loss and a significant public health issue [6]. According to the World Health Organisation, severe periodontal diseases affect about 19% of adults, more than one

billion cases worldwide. Dental conditions, including periodontitis, result in substantial direct treatment costs, estimated at 298 billion US dollars annually, or 4.6% of global health expenditure, alongside considerable indirect costs due to productivity loss [6,19]. Some authors report that in several European countries, oral healthcare can represent almost a third of total healthcare expenses. For example, in Germany, 30–40% of tooth extractions are still due to periodontal disease [6]. This ongoing trend contributes to additional economic burdens, as post-operative discomfort and pain following periodontal surgery negatively affect quality of life and productivity, a situation further exacerbated by lost working hours in countries such as Canada [6].

Economic evaluations show that preventing periodontitis by effectively managing gingivitis is cost-effective. Programs such as supervised toothbrushing, water fluoridation, and the provision of dental supplies in early childhood settings have yielded significant savings for health systems [6]. These authors noted that initiatives like Mouth Care Matters in the United Kingdom aim to increase healthcare professionals' awareness of the link between oral hygiene and general health, with projections indicating substantial direct and broader social benefits for each unit of currency spent [6]. Modelling studies tracking transitions between periodontal health states reveal that individuals with gingivitis have a significant chance of progressing to periodontitis during each follow-up cycle, and some progress even with treatment. These data highlight the importance of early, accurate diagnosis and draw attention to how inconsistent clinical definitions complicate timely disease identification and classification [6].

Recent literature has examined the role of inflammatory mediators in periodontal disease pathogenesis and monitoring, with particular focus on salivary biomarkers such as interleukins IL-1β, IL-6, and IL-8. These biomarkers reflect local immune-inflammatory activity and could aid in risk stratification and personalised care [2,3,7,8,15,17,22]. The biological complexity of periodontitis, characterised by heterogeneity and overlapping risk profiles, has sparked interest in computational methods that integrate clinical, behavioural, sociodemographic, and molecular data [4,24,27]. Within this framework, ML methods—such as Naive Bayes, K-Nearest Neighbours, Decision Trees, Random Forests, Support Vector Machines, and Logistic Regression—are increasingly used in periodontal and biomedical datasets. These approaches seek to improve disease prediction and severity characterisation, exploit nonlinear relationships, and help identify key predictors for clinical support [1,11,18,20,27]. Several studies find that models like Logistic Regression provide strong discriminative performance and interpretability, supporting easier clinical integration through risk scoring and early detection [10,25].

Further supporting these trends, recent studies have emphasised the potential of salivary inflammatory mediators as biomarkers for assessing periodontal disease severity and progression [3,7,16]. Incorporating these biomarkers into machine learning frameworks has demonstrated improved predictive performance. However, challenges persist regarding data preprocessing, feature selection, and clinical translation [1,4,10,11,24,27].

3 Materials and Methods

3.1 Dataset

This study analyzed clinical data from 189 individuals to determine the presence or absence of PD. Data collection followed process no. 22/2022 of the Health Ethics Committee of the Centro Hospitalar do Tâmega e Sousa and were recorded in an Excel database.

The database includes adult participants aged 19–72 years who underwent periodontal clinical examinations, as well as information from questionnaires on oral hygiene habits and tobacco use. Variables also included measurements of inflammatory mediators and cytokines. Saliva samples were collected according to approved ethical protocols, and cytokine levels were quantified using commercial kits.

For participants without diabetes mellitus (DM), salivary cytokine levels were measured in pg/mL using the *LEGENDplex^{TM}Human Macrophage/Microglia Panel (13-plex) Kit (Cat. 740503, BioLegend, San Diego, CA, USA)*. This flow cytometry-based kit allows simultaneous detection of 13 human cytokines and chemokines relevant to immune responses. The assay was performed according to the manufacturer's protocol using fluorescence-coded microbeads specific to each analyte. Readings were acquired by flow cytometry and analyzed with the *LEGENDplex^{TM}Data Analysis Software Suite*. The detection limit for the cytokines included in this study was 2.44 pg/mL [5].

In participants with DM, salivary cytokine concentrations were measured in pg/mL using the LEGENDplex^{TM}Human Essential Immune Response Panel (13-plex) Kit (ref. 740930, BioLegend, CA, USA), which quantifies 13 analytes simultaneously. The assay employed Luminex®xMAP technology, which utilises magnetised microbeads with distinct fluorescent codes to enable highly sensitive multiplex analysis. The experimental protocol adhered to the manufacturer's instructions, including incubation with capture and detection antibodies and measurement on a Luminex®-compatible system. Detection limits were 2.44 pg/mL for most cytokines, except IL-1β (2.69 pg/mL), MCP-1 (2.2 pg/mL), IL-17A (3.66 pg/mL), IL-6 (2.2 pg/mL), IL-8 (2.69 pg/mL), and TGF-β (5.37 pg/mL) [5].

3.2 Data Preprocessing

Data preprocessing comprised exploratory analysis, handling of missing values, normalisation of continuous variables, and logarithmic transformation of skewed biomarker distributions. Values below the detection limit for inflammatory mediators were treated using established imputation strategies. Fig. 1 illustrates the preprocessing flow applied to the dataset, from the original dataset containing 31 variables to the construction of the final dataset with 21 variables, properly processed and prepared for the application of ML algorithms. This workflow ensured the consistency and quality of the data used in predictive modeling.

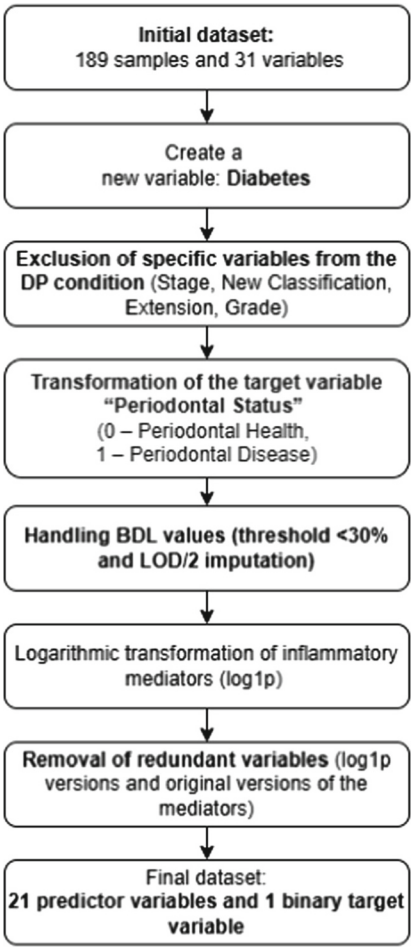

Fig. 1. Data preprocessing workflow.

3.3 Dataset Preparation, Model Selection and Clinical Translation Strategy

The dataset was split into training (80%) and testing (20%) subsets using stratified sampling, which preserved the original class proportions of "Periodontal Health" (38.3% training, 36.8% testing) and "Periodontal Disease" (61.7% training, 63.2% testing). This approach ensured class representativeness across both subsets, minimizing bias in model evaluation. The features were scaled using Standard Scaler to ensure algorithms sensitive to feature magnitudes, such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), performed optimally.

Six supervised ML classification algorithms were selected and trained:

1. Naive Bayes (NB)
2. KNN
3. Decision Tree (DT)
4. Random Forest (RF)
5. SVM
6. Logistic Regression (LR)

Model performance was assessed using standard evaluation metrics, including Accuracy, Precision, Recall, F1-Score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

To facilitate clinical applicability, a structured methodology for model selection was used. Six binary classification algorithms were initially trained to distinguish between periodontal health and periodontal disease. The top three models were then selected based on their performance. Feature importance analysis was conducted in two steps: first, to identify common predictors across the top models, and second, to determine algorithm-specific top features unique to each model. Three training scenarios were then applied to the selected models: (i) all 21 preprocessed variables, (ii) only the identified common features, and (iii) only the top features for each algorithm. This methodology aims to help dental practitioners by streamlining data collection while preserving predictive performance.

4 Results

Data analysis was carried out in Python 3.12.7 using the following libraries: `NumPy` for mathematical operations and array manipulation, `Pandas` for data processing and analysis, `Matplotlib` for graphical visualisation, and `Scikit-learn` for implementing machine learning models and evaluation metrics. The code for the practical section is available on a GitHub repository - (project in Portuguese available at: https://github.com/joanalopes7/Dissertacao-de-Mestrado-Doenca-Periodontal-e-Machine-Learning.git)

The exploratory data analysis reveals right-skewed distributions in key clinical variables, including `CALtotal`, with severe outliers at 7.42 mm and 9.73 mm. These outliers are consistent with Stage IV periodontitis. Inflammatory mediators also showed skewness but were retained due to their clinical relevance. Mediators with more than 30% of values below the detection limit (BDL) were excluded. The remaining mediators (IP-10, IL-1, MCP-1, IL-6, IL-8) were imputed using LOD/2 and log-transformed (`log1p`) to address skewness.

We use Mann-Whitney U tests to evaluate differences between two groups. There are statistically significant differences revealed between the Periodontal Health and Disease groups for all mediators ($p < 0.05$). IL-1β, IL-6, and IL-8 had the lowest p-values. After imputation, correlations preserved expected (Fig. 2) biological relationships, such as the IL-1β/IL-8 correlation ($r = 0.68$), supporting data quality. These results show the discriminative potential of the

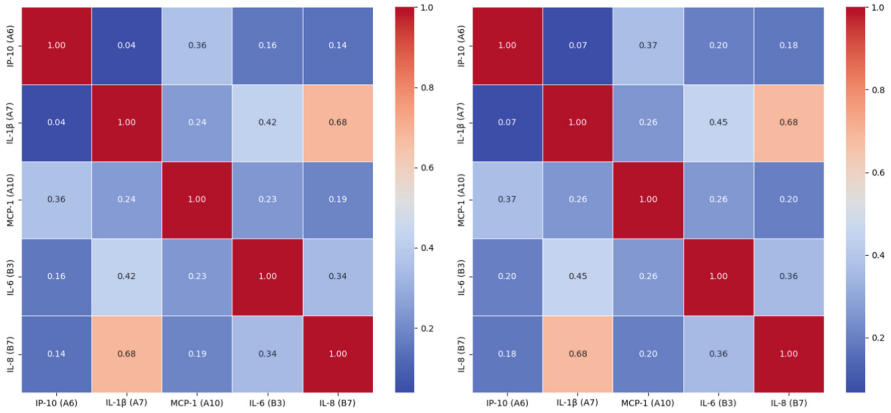

Fig. 2. *Heatmaps* showing the correlation between inflammatory mediators before (left) and after (right) the imputation of BDL values.

selected mediators as robust predictors for periodontal condition classification models.

Subsequent assessment indicated that tree-based models (DT and RF), SVM, and LR achieved high accuracy and area under the ROC curve, indicating strong

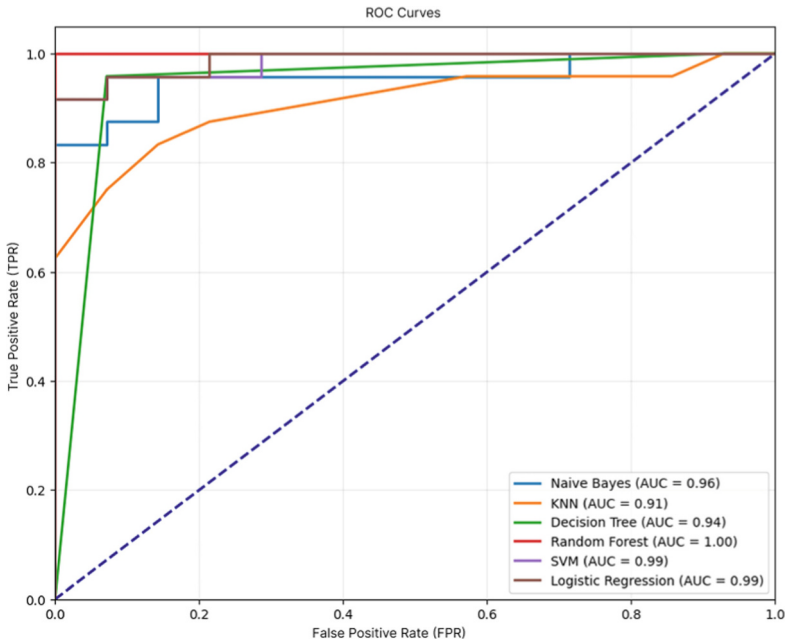

Fig. 3. ROC curves of the six classification models evaluated in this study.

performance in identifying periodontal disease. The Naive Bayes model had high precision, meaning its positive predictions were usually correct, but lower recall, so it missed some cases. The KNN model performed slightly worse, possibly because it is affected by the number of neighbours used and by the data itself.

These results are illustrated by ROC curves in Fig. 3.

Considering the observed performance and ROC curves, the models chosen for the subsequent stage of the methodology adopted in this study were Random Forest, Support Vector Machine, and Logistic Regression. Among these three evaluated models (Table 1), Logistic Regression demonstrated a strong balance between predictive performance (Accuracy=0.974, F1-score=0.979, AUC=1.000) and interpretability. Although ensemble models achieved comparable or slightly higher accuracy, Logistic Regression provided transparent coefficient estimates that facilitated clinical interpretation (Table 2).

Feature importance analysis (Table 3) revealed that periodontal clinical indicators, such as probing depth and plaque index, were among the strongest predictors.

Table 1. Model performance across different feature sets.

Model	Feature Set	Accuracy	Precision	Recall	F1-score	AUC
Random Forest	All Features	0.947	1.0	0.917	0.957	1.000
Random Forest	Shared Features	0.947	1.0	0.917	0.957	0.997
Random Forest	Top Specific Features	0.947	1.0	0.917	0.957	1.000
Logistic Regression	All Features	0.947	1.0	0.917	0.957	0.988
Logistic Regression	Shared Features	0.974	1.0	0.958	0.979	0.988
Logistic Regression	**Top Specific Features**	**0.974**	**1.0**	**0.958**	**0.979**	**1.000**
Support Vector Machine	All Features	0.947	1.0	0.917	0.957	0.985
Support Vector Machine	Shared Features	0.947	1.0	0.917	0.957	0.991
Support Vector Machine	Top Specific Features	0.974	1.0	0.958	0.979	1.000

Table 2. Coefficients, Odds Ratios, and clinical interpretation of the Logistic Regression model with top specific features.

Variable	Coefficient (β)	OR = e^{β}	Clinical Interpretation
BOP_total_percentage	7.4197	1670.7	Each 1% increase in BOP multiplies the odds by ∼1670x
PPD_total	1.4394	4.22	Each additional mm in probing depth increases the odds by 322%
PI_total_percentage	0.8359	2.31	Each 1% increase in PI approximately doubles the odds
log_IL-8 (B7)	0.1595	1.17	Each unit increase in IL-8 increases the odds by 17%
Cigarettes_per_day	0.1028	1.11	Each additional cigarette increases the odds by 11%
log_IL-1β (A7)	-0.4564	0.63	Each unit increase in IL-1β decreases the odds by 37%
No. teeth	-0.9929	0.37	Each additional tooth decreases the odds by 63%
log_IL-6 (B3)	0.2323	1.26	Each unit increase in IL-6 increases the odds by 26%
Brushing_frequency	0.5141	1.67	Each additional brushing increases the odds by 67%
log_MCP-1 (A10)	0.1876	1.21	Each unit increase in MCP-1 increases the odds by 21%

Table 3. Description of variables of selected model - Logistic Regression with Top Specific Features.

Variable	Description	Unit
BOP_total_percent.	Total percentage of bleeding on probing, indicating active gingival inflammation.	%
PPD_total	Total probing pocket depth, reflecting the severity of periodontal tissue destruction.	mm
PI_total_percentage	Total plaque index percentage, associated with oral hygiene control.	%
log_IL-8 (B7)	Log-transformed interleukin-8 level, an inflammatory biomarker related to the local immune response.	pg/mL
Cigarettes_per_day	Number of cigarettes smoked per day, a recognized risk factor for periodontitis.	cigarettes/day
log_IL-1β (A7)	Log-transformed interleukin-1β level, a pro-inflammatory marker strongly associated with tissue destruction.	pg/mL
No._teeth	Total number of teeth present, reflecting the patient's oral health history.	count
log_IL-6 (B3)	Log-transformed interleukin-6 level, an inflammatory biomarker associated with chronic processes.	pg/mL
Brushing_frequency	Daily toothbrushing frequency, a behavioral variable related to oral hygiene.	times/day
log_MCP-1 (A10)	Log-transformed monocyte chemoattractant protein-1 (MCP-1) level, involved in inflammatory cell recruitment.	pg/mL

Smoking status and brushing frequency strongly predicted disease. IL-1β, IL-6, and IL-8 also played key roles.

By combining traditional clinical factors and specific laboratory measurements, our model draws on diverse sources of information, thereby enhancing its predictive ability and providing unique insights into risk mechanisms not available through standard approaches.

5 Discussion

As mentioned before, the main objective of this study was to explore the potential of machine learning models to predict PD using clinical, sociodemographic, behavioural and molecular data. Throughout the entire workflow developed—from data collection and pre-processing to algorithm training and model evaluation—it was possible to identify the key variables associated with the risk of developing PD and to assess the discriminatory power of different

predictive approaches. The discussion that follows seeks to relate the results obtained to those in the existing literature.

Our results showed that total Bleeding on Probing (BOP) was among the strongest predictors of PD. This finding is consistent with previous works that identifies BOP as a fundamental clinical indicator of gingival inflammation and disease progression [6,13,18]. Within the group of diseases that includes gingivitis and periodontitis—both of which fall under the general concept of periodontal disease—bleeding is known to be an unequivocal clinical sign: in gingivitis, this may manifest as bleeding whilst brushing the teeth, whereas in periodontitis it may occur spontaneously for no apparent reason. In this sense, bleeding on probing (BOP) reflects active inflammation and constitutes an objectively reproducible measure that is highly informative for risk assessment.

Statistical analyses conducted in this work show the value of inflammatory mediators in PD diagnosis. All salivary inflammatory markers used to train the model showed statistically significant differences between the 'Periodontal Health' and 'Periodontal Disease' groups. IL-1β, IL-6, and IL-8 have consistently emerged as relevant biomarkers in the three main models, as indicated in previous studies which associate these mediators with periodontal inflammation and tissue degradation [23]. Consequently, clinicians can regularly use saliva-based diagnostics for the early detection, monitoring, prognosis and management of PD, through the detection of host-derived biomarkers. It should always be kept in mind that the diagnostic validity of these tests depends on their sensitivity and specificity [7,23]. Our findings reinforce the importance of certain inflammatory mediators as clinical indicators and potential therapeutic targets in PD.

Furthermore, the analysis of the significance of the variables revealed a wide range of influential factors, with particular emphasis on classic periodontal parameters described in the literature, such as the bacterial plaque index, probing depth and clinical attachment loss, as well as systemic and behavioural factors, such as diabetes and a history of smoking. This variety of significant predictors reinforces the multifactorial nature of PD and the need to consider multiple risk factors that interact with each other to obtain a robust assessment of the risk and severity of the disease [21,26]. Local, systemic and behavioural factors do not operate in an isolated manner; instead, they interact to shape the host's response to the oral biofilm and, ultimately, the clinical manifestation of the disease.

All the machine learning models tested in this study performed well, with an accuracy of over 84% in all cases, and several algorithms—including Random Forest, Decision Tree, Support Vector Machine, and Logistic Regression—achieved an accuracy of over 94%. This high performance reinforces the use of computational approaches for the early identification and prediction of PD and suggests that these models can assist clinicians with risk assessment and clinical decision-making. Although various performance measures (precision, accuracy, recall and F1-score) were calculated, the main metric used to compare the classifiers was the area under the ROC curve. The AUC was preferred because it captures overall discriminatory power independent of the choice of decision threshold and is

less vulnerable to potential class imbalance than precision. In clinical applications, the AUC is particularly suitable because it reflects the trade-off between sensitivity and specificity, which is important for minimising both false negatives and false positives in the diagnosis of PD.

With regard to the chosen model, based on the results presented, the final model selected for clinical application was a logistic regression classifier trained using a subset of the best-performing features. This choice was based not only on its near-perfect performance (AUC = 1.0, with high accuracy, precision and F-score), but also on its interpretability and relevance to clinical practice. Compared to more complex methods, such as Random Forest or Support Vector Machines, Logistic Regression provides coefficients that can be easily interpreted as odds ratios (ORs), allowing clinicians to understand the relative contribution of each variable to the risk of disease. This form of clarity is a major advantage in medical contexts, where model explanations are needed to support clinical judgement and build confidence. The analysis of the coefficients of the Logistic Regression model, together with its principal characteristics, clarified which variables contributed most strongly to the risk of PD. Specifically, the total percentage of BOP, total probing depth and total plaque index had very high ORs (approximately 1670, 4.22 and 2.31, respectively), indicating that even small increases in these clinical parameters substantially increase the likelihood of PD. On the LR model, probability estimates close to 1 indicate a high risk of periodontal disease, whilst those close to 0 indicate a low risk. It should also be pointed out that the coefficients represent adjusted effects, showing the combined influence of all variables in the model. Thus, variables with low multicollinearity, such as smoking or toothbrushing frequency, can be interpreted more directly. In contrast, inflammatory markers or correlated clinical variables should be interpreted within the context of the full multivariate model.

The LR model shows a positive coefficient for daily brushing frequency. This does not mean that more frequent brushing increases the risk of periodontal disease; rather, it reflects the multivariate adjustment. One possible explanation is that patients with poorer periodontal health (higher values for bleeding on probing, probing depth or plaque index) may brush their teeth more frequently in response to symptoms or medical advice, creating a positive association in the data. Clinically, tooth brushing remains protective—the positive coefficient indicates a statistical association after adjustment for other variables. This model should be used as a tool for jointly predicting PD risk, not for evaluating the effects of the coefficients separately.

The final proposed model remains robust across various decision thresholds. Performance analysis revealed high sensitivity and specificity, enabling the reliable detection of cases of periodontal disease without a significant increase in the false positive rate. Its high discriminatory power shows that the predicted probability distributions clearly separate the ‘Periodontal Health’ and ‘Periodontal Disease’ groups, reducing ambiguity in borderline cases.

6 Limitations and Future Work

Several limitations should be addressed. The sample size was relatively small. Thus, the models were not tested on an independent dataset. Subsequent research should evaluate these models using larger, more varied populations. Additionally, employing longitudinal data that tracks the same subjects over time may improve the prediction of disease progression.

7 Conclusion

In conclusion, we can say that this study demonstrates that interpretable machine learning models are capable of accurately predicting periodontal disease using clinical, behavioural and inflammatory data. The logistic regression model achieved high predictive accuracy whilst maintaining clinical practicality, demonstrating its suitability as a decision-support tool in periodontal care. This concise model, which includes ten variables (BOP_total_percentage, PPD_total, PI_total_percentage, log_IL-8 (B7), Cigarettes_per_day, log_IL-1β (A7), No. teeth, log_IL-6 (B3), Brushing_frequency, log_MCP-1 (A10)) confirmed established risk factors and highlighted the importance of salivary cytokines as non-invasive and quantifiable biomarkers of disease activity. These results support the practical utility of logistic regression for personalised risk assessment and management in periodontal clinical practice.

Acknowledgments. This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UID/00319/2023.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alqahtani, H.M., Koroukian, S.M., Stange, K., Schiltz, N.K., Bissada, N.F.: Identifying factors associated with periodontal disease using machine learning. Journal of International Society of Preventive and Community Dentistry **12**, 612–620 (11 2022). https://doi.org/10.4103/jispcd.JISPCD_188_22
2. Ashwood, P., Krakowiak, P., Hertz-Picciotto, I., Hansen, R., Pessah, I., de Water, J.V.: Elevated plasma cytokines in autism spectrum disorders provide evidence of immune dysfunction and are associated with impaired behavioral outcome. Brain Behav. Immun. **25**(1), 40–45 (2011). https://doi.org/10.1016/j.bbi.2010.08.003
3. BinShabaib, M., AlHarthi, S.S., Akram, Z., Khan, J., Rahman, I., Romanos, G.E., Javed, F.: Clinical periodontal status and gingival crevicular fluid cytokine profile among cigarette-smokers, electronic-cigarette users and never-smokers. Arch. Oral Biol. **102**, 212–217 (2019). https://doi.org/10.1016/j.archoralbio.2019.05.001
4. Binson, V.A., Thomas, S., Subramoniam, M., Arun, J., Naveen, S., Madhu, S.: A review of machine learning algorithms for biomedical applications. Ann. Biomed. Eng. **52**(5), 1159–1183 (2024). https://doi.org/10.1007/s10439-024-03459-3

5. BioLegend: Legendplex^{TM} mull-analyte flow assay kit for accurate quantification of multiple human th (t helper cell) cytokines from cell culture supernatant, serum, plasma and other biological samples legendplex^{TM} mull-analyte flow assay kit. Tech. rep. (2025), https://sl1nk.com/ohloU
6. Bishop, C.: Time to take gum disease seriously: The societal and economic impact of periodontitis. Economist Intelligence Unit report, London (2021), commissioned by the European Federation of Periodontology
7. Cafiero, C., Spagnuolo, G., Marenzi, G., Martuscelli, R., Colamaio, M., Leuci, S.: Predictive periodontitis: The most promising salivary biomarkers for early diagnosis of periodontitis (4 2021). https://doi.org/10.3390/jcm10071488
8. Chang, M.C., Lin, S.I., Pan, Y.H., Lin, L.D., Wang, Y.L., Yeung, S.Y., Chang, H.H., Jeng, J.H.: Il-1-induced icam-1 and il-8 expression/secretion of dental pulp cells is differentially regulated by irak and p38. J. Formos. Med. Assoc. **118**(8), 1247–1254 (2019). https://doi.org/10.1016/j.jfma.2018.11.015
9. Chapple, I.L.C., Avila-Ortiz, G., Nibali, L., Graziani, F., Sanz, M., Tonetti, M., Giannobile, W.V., et al.: Periodontal health and gingival diseases and conditions on an intact and a reduced periodontium: Consensus report of workgroup 1 of the 2017 world workshop on the classification of periodontal and peri-implant diseases and conditions. In: Journal of Clinical Periodontology. vol. 45, pp. S68–S77 (06 2018https://doi.org/10.1111/jcpe.12940
10. Ciobanu-Caraus, O., Aicher, A., Kernbach, J.M., Regli, L., Serra, C., Staartjes, V.E.: A critical moment in machine learning in medicine: On reproducible and interpretable learning. Acta Neurochir. **166**(1), 65–78 (2024). https://doi.org/10.1007/s00701-024-05892-8
11. Deng, Y., et al.: Artificial intelligence and machine learning for periodontal diagnosis and prognosis. (Complete with the exact journal details you are using) (2023)
12. of Dental, N.I., Research, C.: Probing periodontal disease. https://www.nidcr.nih.gov/news-events/nidcr-news/2024/probing-periodontal-disease (03 2024), accessed 21 Oct. 2024
13. European Federation of Periodontology: Gum diseases explained. https://www.efp.org/for-patients/gum-diseases/gum-diseases-explained/ (2021), acedido em: 19 out. 2024
14. Hasan, A., Palmer, R.M.: A clinical guide to periodontology: Pathology of periodontal disease. British Dental Journal **216**(8), 457–461 (04 2014https://doi.org/10.1038/sj.bdj.2014.299
15. Li, W., Xia, Y., Yang, J., Sanyal, A.J., Shah, V.H., Chalasani, N.P., Yu, Q.: Disrupted balance between pro-inflammatory lipid mediators and anti-inflammatory specialized pro-resolving mediators is linked to hyperinflammation in patients with alcoholic hepatitis. Frontiers in Immunology **15**, 1377236 (2024). https://doi.org/10.3389/fimmu.2024.1377236
16. Majeed, A.M., et al.: Salivary cytokines as biomarkers for periodontal disease severity: A systematic review. (Fill in exact journal details as per your original reference) (2022)
17. Manzardo, A.M., Poje, A.B., Penick, E.C., Butler, M.G.: Multiplex immunoassay of plasma cytokine levels in men with alcoholism and the relationship to psychiatric assessments. Int. J. Mol. Sci. **17**(4), 472 (2016). https://doi.org/10.3390/ijms17040472
18. Marotz, C., et al.: Machine learning for oral microbiome and periodontal disease prediction. (Complete with journal, volume, pages, doi) (2022)
19. Organization, W.H.: Oral health. https://www.who.int/news-room/fact-sheets/detail/oral-health (03 2023), accessed 10 Nov. 2024

20. Özden, F., et al.: Application of machine learning methods in periodontal disease classification. (Complete with journal details according to your original source) (2015)
21. Papapanou, P.N., Sanz, M., Buduneli, N., Dietrich, T., Feres, M., Fine, D.H., Flemmig, T.F., Garcia, R., Giannobile, W.V., Graziani, F., Greenwell, H., Herrera, D., Kao, R.T., Kebschull, M., Kinane, D.F., Kirkwood, K.L., Kocher, T., Kornman, K.S., Kumar, P.S., Loos, B.G., Machtei, E., Meng, H., Mombelli, A., Needleman, I., Offenbacher, S., Seymour, G.J., Teles, R., Tonetti, M.S.: Periodontitis: Consensus report of workgroup 2 of the 2017 world workshop on the classification of periodontal and peri-implant diseases and conditions. In: Journal of Clinical Periodontology. vol. 45, pp. S162–S170 (06 2018).https://doi.org/10.1111/jcpe.12946
22. Polinski, K.J., Bemis, E.A., Yang, F., Crume, T., Demoruelle, M.K., Feser, M., Seifert, J., O'Dell, J.R., Mikuls, T.R., Weisman, M.H., Gregersen, P.K., Keating, R.M., Buckner, J., Reisdorph, N., Deane, K.D., Clare-Salzler, M., Holers, V.M., Norris, J.M.: Association of lipid mediators with development of future incident inflammatory arthritis in an anti–citrullinated protein antibody–positive population. Arthritis and Rheumatology **73**(6), 955–962 (2021). https://doi.org/10.1002/art.41631
23. Relvas, M., Mendes-Frias, A., Gonçalves, M., Salazar, F., López-Jarana, P., Silvestre, R., da Costa, A.V.: Salivary il-1, il-6, and il-10 are key biomarkers of periodontitis severity. International Journal of Molecular Sciences **25** (8 2024https://doi.org/10.3390/ijms25158401
24. Sarker, I.H.: Machine learning algorithms, real-world applications and research directions. SN Computer Science **2**(3), 160 (2021). https://doi.org/10.1007/s42979-021-00592-x
25. Schober, P., Vetter, T.R.: Logistic regression in medical research. Anesthesia & Analgesia **132**(2), 365–366 (2021). https://doi.org/10.1213/ANE.0000000000005247
26. Steffens, J.P., Marcantonio, R.A.C.: Classificação das doenças e condições periodontais e peri-implantares 2018: guia prático e pontos-chave. Revista de Odontologia da UNESP **47**, 189–197 (8 2018).https://doi.org/10.1590/1807-2577.04704
27. Uddin, S., Khan, A., Hossain, M.E., Moni, M.A.: Comparing different supervised machine learning algorithms for disease prediction. BMC Med. Inform. Decis. Mak. **19**(1), 281 (2019). https://doi.org/10.1186/s12911-019-1004-8
28. Zhang, M., Zhang, X., Wang, K., Tang, X., Sun, J., Zhang, Y.: An update on periodontal inflammation and bone loss. Frontiers in Immunology **15**, 1406798 (2024https://doi.org/10.3389/fimmu.2024.1406798