Abstract for Greeks in AI 2026

Non-invasive screening for periodontitis in non-dental settings: leveraging explainable machine learning on proteomic and self-reported data

E. Stamatelou | M.X.F. Kosho | N. Nijland | A. Gallucci | N. Su | B.G. Loos

Periodontitis is a chronic inflammatory disease associated with type 2 diabetes and cardiovascular disease. Treatment of periodontitis has been shown to benefit both glycemic control and cardiovascular risk markers, underscoring the need for early identification. Periodontitis remains hard to identify outside dental settings because its assessment requires evaluation by a dental professional. Therefore, scalable, noninvasive screening tools are needed to identify high-risk individuals in medical settings, pharmacies, and at home.

To develop and internally validate two complementary machine learning (ML) approaches for screening severe periodontitis using (1) oral rinse proteomics and (2) self-reported oral health questionnaires and demographics.

Both approaches applied a shared explainable ML approach based on repeated nested cross-validation with Logistic Regression (LR) and Random Forest (RF), and SHapley Additive exPlanations (SHAP) for interpretability. For the proteomic-based model, 54 oral rinse samples from 38 patients with periodontitis and 16 controls were profiled using the Proximity Extension Assay (Olink® Inflammation and Immuno-Oncology panels, 184 proteins). XGBoost was additionally evaluated, and Boruta feature selection was applied to handle high-dimensional protein data. For the questionnaire-based model, three Dutch cohorts (n = 278) were analyzed, with 90 individuals with severe periodontitis and 188 without periodontitis.

The proteomic-based LR model reached an ROC-AUC of 0.86 (95% CI: [0.83, 0.88]), sensitivity of 0.74 (95% CI: [0.73, 0.79]), and specificity of 0.80 (95% CI: [0.75, 0.84]) for periodontitis versus controls, with Gal-1, HGF, TNFSF14, and ARG1 identified as the most discriminatory proteins, all showing higher values in periodontitis patients. The questionnaire-based LR model achieved an ROC-AUC of 0.84 (95% CI: [0.81, 0.86]), sensitivity of 0.74 (95% CI: [0.69, 0.78]), and specificity of 0.76 (95% CI: [0.74, 0.79]), with SHAP identifying self-reported gum treatment history, self-rated oral health, sex, and age as the most important predictors.

Explainable ML applied to oral rinse proteomics and questionnaires offers a promising, non-invasive approach for severe periodontitis screening in non-dental settings, supporting timely referral and potentially contributing to improved management of type 2 diabetes and cardiovascular disease. External validation in independent cohorts is warranted before clinical implementation.

Note: This abstract integrates findings from the two following manuscripts:

(1) Kosho, M.X.F., Stamatelou, E., Loos, B.G. (2026). Machine learning-based analysis of oral rinse samples to identify proteomic signatures for severe periodontitis. A pilot study. *(manuscript submitted to Scientific Reports, currently in second revision)*

(2) Stamatelou, E., Nijland, N., Gallucci, A., Su, N., | Loos, B.G. (2026). Explainable machine learning for periodontitis screening using questionnaires and demographics *(manuscript in preparation, to be submitted to Journal of Clinical Periodontology)*

Keywords: AI for Health, screening, point-of-care testing, periodontal disease, proteomics, explainable artificial intelligence