Predicting periodontal severity via machine learning

Yuwon Oh, Juheon Oh

Overview: We develop the Machine Learning Model to classify severe cases in periodontal patients and validates our work with 11 participants.

Introduction

Current diagnostic approaches for periodontal disease rely primarily on in-clinic examinations, which limit accessibility and scalability. Breath analysis presents a promising alternative, as specific volatile sulfur compounds (VSCs) correlate strongly with periodontal pathology. Prior studies have demonstrated that elevated VSC levels contribute not only to halitosis but also to periodontal tissue destruction [1]. More recent findings indicate that baseline VSC levels can significantly predict disease progression [2]. Building on this foundation, our pilot study integrates a broad-spectrum VOC sensor with machine learning models to provide a low-cost, real-time screening tool for early detection and improved access to oral healthcare.

Results

Among 26 engineered features, the plaque measurements showed the strongest correlation with disease severity. As for the time series analysis, each clinical measurement suggested a trend; the plaque index in severe patients had high initial value and clear separation between classes in Month1 measurement, while the non-severe patients did not. Other measurements, Pocket depth and BOP exhibited similar trend, but in contrast, they did not reveal clear separation between classes (Fig 4).

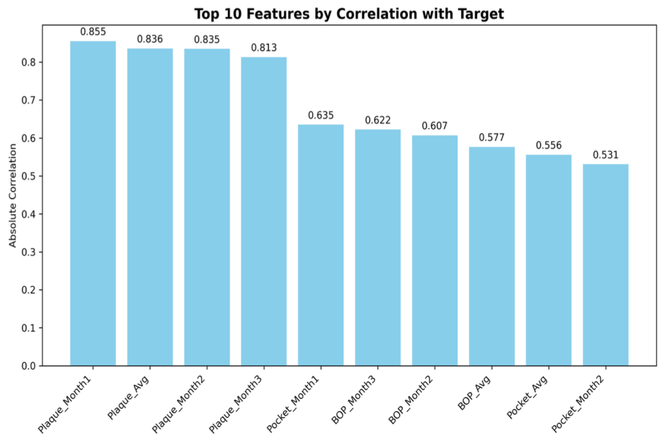

Figure 4: Top 10 features by mutual information

With plaque measurement being excluded, we got still perfect ROC AUC and 100% test accuracy, suggesting that the separation was achievable even with alternative features. Performance drops were modest compared to plaque-included model, but the cross validation scores were higher and variability higher, indicating greater generalization in uncertainty.

Model	Test Accuracy	10-fold CV Accuracy	CV Std	ROC AUC
Random Forest	1.0	0.850000	0.152753	1.0
Gradient Boosting	1.0	0.908333	0.141667	1.0
AdaBoost	1.0	0.908333	0.141667	1.0
Logistic Regression	1.0	0.825000	0.184278	1.0
SVM	1.0	0.850000	0.152753	1.0
Decision Tree	1.0	0.875000	0.154785	1.0
K-Nearest Neighbors	1.0	0.816667	0.152753	1.0
Naive Bayes	1.0	0.825000	0.184278	1.0

Table 1: External Validation performance

In an external sensor-only cohort (N = 11), the model reached an accuracy of 0.91, AUC of 1.00, and Cohen's kappa of 0.74. For severe disease, precision was 1.00 and recall was 0.67. These results indicate that, even absent clinical indices, breath derived VOC signals carry sufficient discriminative structure for meaningful severity classification.

Methods

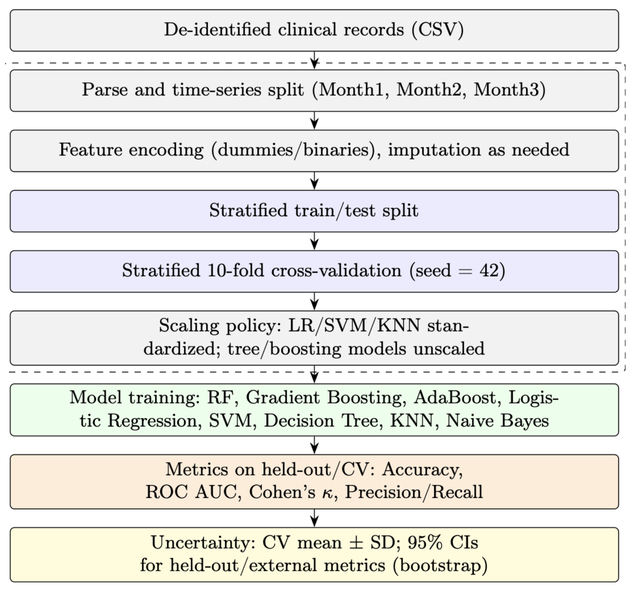

Figure 1. Leakage-safe pipeline (column-friendly): preprocessing and scaling occur inside CV folds; tree/boosting models use unscaled inputs; LR/SVM/KNN are standardized

De-identified clinical records were parsed into a Month-1–3 time series. Features were encoded (binary/dummy) and missing values imputed. We applied a stratified train/test split, then stratified 10-fold cross-validation (seed = 42) with all preprocessing performed inside each CV fold. LR/SVM/KNN were standardized; tree-based and boosting models used unscaled inputs. We evaluated eight algorithms: Random Forest, Gradient Boosting, AdaBoost, Logistic Regression, SVM, Decision Tree, KNN, and Naïve Bayes (Fig. 1). Metrics included accuracy, ROC AUC, Cohen's κ, precision, and recall.

After the machine learning process, we validated our model's framework with 11 participants. The 11 participants answer the brief survey and exhale towards our VOC sensor integrated with microcontroller (Figure 2). Each participant follows the breath sampling protocol to minimize the contamination in recordings (Figure 3).

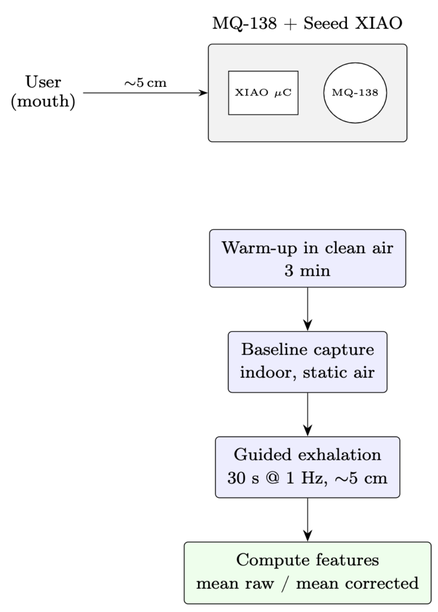

Figure 2: Breath sampling protocol with a compact MOS VOC sensor

Per participant, we computed the mean raw and mean corrected voltage across the 30 s window. A pre-specified corrected-voltage threshold of >300 arbitrary units was used to map readings to binary classes. This cutoff was measured by Youden's J Maximum and kept it fixed for external testing [3]. Evaluation metrics and reporting matched those used in the retrospective analysis.

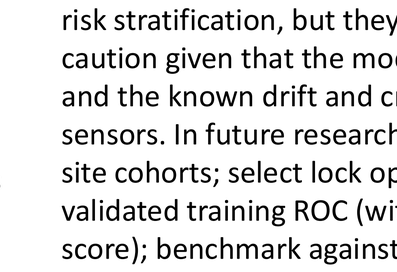

Figure 3: participant exhaling toward the VOC sensor through the mouthpiece

De-identified clinical records from the Seoul Oh's clinic in Gyeonggi-do, Republic of Korea were accessed with approval from Dr. Juheon Oh, a co-author of this study and director of the clinic. This study was an independent secondary analysis; because the study primarily involved technology feasibility testing, it met the exemption criteria from formal IRB review. External breath measurements recorded only non-identifiable information.

Conclusion

These findings support the feasibility of affordable, at-home risk stratification, but they should be interpreted with caution given that the model has small sample size, and the known drift and cross-sensitivity of metal-oxide sensors. In future research, we will expand to larger, multi-site cohorts; select lock operating points from cross-validated training ROC (with calibration metrics such as Brier score); benchmark against simple clinical rules and standard devices; and add environmental compensation and robustness checks across repeated session. By tightening statistical validation and protocol harmonization, we aim to translate this approach into a practical, scalable tool for early periodontal screening outside the clinic.

References & Acknowledgements

Ratcliff, P.A., Johnson, P.W.: The relationship between oral malodor, gingivitis, and periodontitis. A review. Journal of Periodontology. 70(5), 485–489 (1999).
Makino, Y. et al.: Association between volatile sulfur compounds and periodontal disease progression in elderly Non-Smokers. Journal of Periodontology.
Fluss R, Faraggi D, Reiser B. *Estimation of the Youden Index and its associated cutoff point.* Biometrical Journal. 47(4), 458–472 (2005)
Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman AE. *Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection.* BMC Med Res Methodol. 8(32), (2008)

The author thanks Professor Nabil Alshurafa, from Northwestern university,
graduate student Wooyoung Kang, from New York University,
and Josh(Ryugwang) Jang from University of California, Berkeley for their guidance.