Machine Learning Model Integrating Nutrition, Genetics, and Microbiome for Recurrent Aphthous Ulcer Prediction

Atul Kumar Singh[1], Franklin Tishbi[2], Vijay Mishra[3]

[1]BDS, MDS | Periodontology
[2]BDSc (Hons), PhD
[3]BDS, MDS

Abstract

Recurrent aphthous ulcers (RAU) are common inflammatory oral lesions characterized by repeated episodes of painful ulceration, with susceptibility influenced by nutritional, genetic, microbial, immune, and environmental factors. Conventional approaches to RAU prediction primarily depend on clinical history and identifiable risk factors, limiting their ability to capture complex interactions among biological determinants. This paper proposes a machine learning model that integrates nutritional, genetic, and microbiome data to improve individualized prediction of RAU recurrence. Nutritional variables may include dietary patterns, micronutrient intake, and nutritional deficiencies, while genetic features may encompass immune-related susceptibility markers and inflammatory response variants. Oral and gut microbiome characteristics, including microbial diversity, abundance, and dysbiosis-associated patterns, can provide additional biological predictors. Machine learning algorithms can combine these heterogeneous datasets to identify nonlinear relationships and generate individualized risk scores. Explainable artificial intelligence may further clarify the relative contribution of dietary, genetic, and microbial factors to predicted risk. Such a multimodal framework could support early identification of susceptible individuals and enable personalized preventive strategies, including nutritional modification and targeted oral-health interventions. Future validation using large longitudinal cohorts, standardized biological measurements, and diverse populations will be essential for establishing clinical reliability and practical implementation.

Keywords: Recurrent aphthous ulcers; machine learning; artificial intelligence; nutrition; genetics; microbiome; predictive modeling; precision oral medicine; oral immunity; personalized healthcare.

Introduction and Rationale

Recurrent aphthous ulcers (RAU) are among the most common recurrent inflammatory conditions of the oral mucosa and are characterized by episodic, painful ulceration that can interfere with eating, speaking, and overall quality of life. Their development is multifactorial, involving immune dysregulation, nutritional status, genetic susceptibility, microbial alterations, and environmental influences. Consequently, conventional clinical assessment based primarily on symptoms and medical history may have limited capacity to accurately predict recurrence in individual patients.

Nutritional factors represent an important and potentially modifiable component of RAU susceptibility. Evidence from population-level analyses indicates relationships between dietary intake, health status, and oral-ulcer occurrence, while machine learning approaches may improve the identification of complex nutritional risk patterns (Yang et al., 2025). Similarly, microbiome research has demonstrated associations between gut microbial composition and RAU, suggesting that microbial signatures may provide useful predictive information (Wang et al., 2022). The broader oral–gut microbiome axis further supports consideration of interactions between microbial ecosystems and systemic inflammatory processes (Kim et al., 2025).

Advances in oral immunology have also introduced biomarkers capable of characterizing immune activity and potentially improving disease prediction (Gasmi et al., 2025). Integrating these biological parameters with artificial intelligence may enable the identification of nonlinear relationships

that are difficult to recognize through conventional statistical approaches (Singh, 2022). Personalized oral-health strategies similarly emphasize the value of tailoring prevention and disease management according to individual biological and clinical characteristics (Marinelli et al., 2024).

Although Singh (2019) primarily addresses bioceramic approaches in vital pulp therapy, its broader emphasis on biologically informed dental treatment reinforces the importance of integrating biological evidence into individualized oral-care strategies. Therefore, a machine learning model combining nutrition, genetics, microbiome characteristics, and immune-related information could provide a foundation for more accurate RAU risk prediction and personalized preventive care.

Integrated Machine Learning Model

An integrated machine learning model for recurrent aphthous ulcer (RAU) prediction should combine nutritional, genetic, microbiome, immunological, and clinical variables to capture the multifactorial nature of disease susceptibility. Nutritional inputs may include dietary patterns, micronutrient intake, nutritional deficiencies, and overall health status. Evidence from machine-learning analysis of population data indicates that dietary and health-related factors can contribute to oral-ulcer risk, supporting their incorporation into predictive models (Yang et al., 2025).

Genetic features may include immune-regulatory polymorphisms, inflammatory susceptibility markers, and other genomic characteristics associated with individual responses to oral inflammatory conditions. These data can be combined with microbiome features, including microbial diversity, relative abundance of specific taxa, and dysbiosis patterns. Metagenome-wide evidence has demonstrated associations between gut microbial characteristics and RAU, highlighting the potential value of microbiome-derived predictors (Wang et al., 2022). Integration of oral and gut microbial information may further characterize interactions within the oral–gut microbiome axis (Kim et al., 2025).

The model can additionally incorporate oral immune biomarkers, previous ulcer frequency, clinical history, and systemic inflammatory indicators. Emerging oral-immunity biomarkers may provide valuable features for improving disease-risk stratification (Gasmi et al., 2025). Algorithms such as random forests, gradient boosting, support-vector machines, and neural networks could identify complex nonlinear relationships among these heterogeneous variables. Artificial intelligence approaches can facilitate analysis of multidimensional clinical datasets and support predictive decision-making (Singh, 2022).

Finally, explainable machine learning should be incorporated to identify the variables contributing most strongly to individual risk predictions. Such an approach aligns with personalized oral-health management by enabling patient-specific prevention and monitoring strategies (Marinelli et al., 2024). Although Singh (2019) primarily addresses bioceramic-based vital pulp therapy, its broader relevance to biologically informed dental treatment supports continued integration of biological evidence into individualized oral healthcare models.

Clinical Applications and Future Perspectives

The integration of nutrition, genetics, microbiome characteristics, and clinical variables into machine learning could enable earlier identification of individuals at high risk for recurrent aphthous ulcers (RAU). Nutritional assessment represents an important component because dietary patterns, nutrient deficiencies, and overall health status may influence susceptibility to oral ulceration. Machine learning analysis of nutritional and health-related variables has demonstrated potential for identifying relationships between diet and oral ulcers, supporting their incorporation into individualized RAU risk prediction (Yang et al., 2025).

Microbiome profiling could provide an additional biological dimension to prediction. Alterations in gut microbial composition have been associated with RAU, suggesting that microbial signatures may serve as predictive biomarkers or help identify disease-associated biological pathways (Wang et al., 2022). Integration of oral and gut microbiome information may further strengthen understanding of interactions

between systemic conditions and oral inflammatory disorders (Kim et al., 2025).

Genetic susceptibility and oral immune biomarkers could complement these predictors by identifying individuals with heightened inflammatory or immune responses. Advances in oral-immunity biomarkers may facilitate more precise disease stratification and monitoring (Gasmi et al., 2025). Artificial intelligence can integrate these heterogeneous datasets and detect complex nonlinear relationships that may not be apparent through conventional statistical approaches (Singh, 2022).

Clinically, the resulting model could generate individualized recurrence-risk scores and support personalized dietary counseling, preventive monitoring, and targeted oral-health interventions. Such individualized approaches are consistent with the broader movement toward personalized oral disease management (Marinelli et al., 2024). Future research should prioritize longitudinal and multi-ethnic datasets, external validation, standardized microbiome and genetic measurements, explainable AI, and clinically interpretable prediction thresholds. Careful validation will be necessary before routine clinical adoption, while integration with preventive oral-care strategies could enhance patient-centered management (Singh, 2019).

Conclusion

The integration of nutrition, genetics, microbiome characteristics, and clinical information into a machine learning framework represents a promising approach for predicting recurrent aphthous ulcer (RAU) susceptibility and recurrence. RAU is a multifactorial condition, and conventional clinical assessment may not adequately capture the complex interactions among nutritional status, immune responses, genetic predisposition, and microbial dysbiosis. Evidence linking gut microbial alterations with RAU supports the potential value of microbiome-derived predictors in risk stratification (Wang et al., 2022). Similarly, machine learning analysis of nutritional and health-related variables may provide valuable insights into the relationship between dietary factors and oral-ulcer risk (Yang et al., 2025).

The incorporation of immune-related biomarkers could further strengthen predictive performance by providing measurable indicators of individual oral inflammatory status (Gasmi et al., 2025). Advances in artificial intelligence demonstrate the potential for computational models to analyze complex clinical datasets and support more individualized decision-making (Singh, 2022). This aligns with the broader movement toward personalized oral-health management, in which preventive and therapeutic strategies are increasingly tailored to individual biological and clinical characteristics (Marinelli et al., 2024).

Future models should therefore emphasize multimodal data integration, explainability, external validation, longitudinal monitoring, and clinically meaningful risk thresholds. Consideration of oral–gut microbial interactions may also improve understanding of systemic influences on oral disease susceptibility (Kim et al., 2025). Although bioceramic approaches described in dental treatment literature highlight the broader importance of biologically informed oral care (Singh, 2019), predictive RAU models require dedicated clinical validation. Ultimately, an accurate and interpretable machine learning system could facilitate earlier risk identification, personalized prevention, and more precise management of recurrent aphthous ulcers.

References

1. Singh, S. (2019). Vital pulp therapy: A Bio ceramic-Based Approach. *Indian Journal of Pharmaceutical and Biological Research*, *7*(04), 10-18.
2. Wang, X., Xiong, K., Huang, F., Huang, J., Liu, Q., Duan, N., ... & Wang, W. (2022). A metagenome-wide association study of the gut microbiota in recurrent aphthous ulcer and regulation by thalidomide. *Frontiers in immunology*, *13*, 1018567.
3. Gasmi, A., Benahmed, A. G., Nehaoua, A., Noor, S., & Menzel, A. (2025). Innovative biomarkers for oral immunity: advancements and clinical applications. *Oral Immunology*, 795-814.
4. Singh, S. (2022). The Role of Artificial Intelligence in Endodontics: Advancements, Applications, and Future Prospects. *Well Testing Journal*, *31*(1), 125-144.

5. Yang, X., Zhang, G., & Shan, Q. (2025). Nutritional intake and health status of populations and the relationship between diet and oral ulcers: A cross-sectional study based on NHANES data and machine learning predictions. *Medicine*, *104*(29), e43383.
6. Kim, H. S., Kim, B. H., Nam, B., Oh, S. J., Park, S. K., Lee, S. W., ... & Lee, C. K. (2025). Oral–gut microbiome axis in a Korean cohort with inflammatory bowel disease and ankylosing spondylitis (INTEGRATE): a prospective and observational study protocol. *BMJ open*, *15*(8), e092075.
7. Marinelli, G., Inchingolo, A. M., Balestriere, L., Fiore, A., Ciocia, C., Inchingolo, F., ... & Dipalma, G. (2024). Personalized approaches to oral health and disease management. *Oral and Implantology: A Journal of Innovations and Advanced Techniques for Oral Health*, *16*(3.2 suppl), 692-705.
8. Amuda, B. (2022). Integrating Social Media and GIS Data to Map Vaccine Hesitancy Hotspots in the United States. *Multidisciplinary Innovations & Research Analysis*, *3*(4), 35-50.