International Research Journal of Medicine and Surgery (IRJMedS), 2026;3(3):11-16 IRJMedS

Review Article | ISSN (O): 3048-6955 DOI: 10.47857/irjmeds.2026.v03i03.074

Artificial Intelligence-driven Oral Microbiome Analysis: Current Applications and Future Directions

SK Roy Chowdhury[1]*, Jui S Karmarkar[2], Nazia Khan[3], Manjunath Badni[4], Sumaiya Kausar[5]

[1]Armed Forces Medical Services (AFMS) Veteran, Department of Oral and Maxillofacial Surgery, Pune, Maharashtra, India, [2]Department of Oral and Maxillofacial Surgery, Dr D Y Patil Dental College and Hospital, Pimpri, Pune, Maharashtra, India, [3]Clinical Microbiology, Basic Medical Science, College of Medicine, Majmaah University, Al-Majmaah, Riyadh, Saudi Arabia, Department, [4]Department of Oral Pathology and Microbiology, Subbaiah Institute of Dental Sciences, HH Road, Purle, Shimoga, Karnataka, India, [5]Department of Microbiology, Kasturba Medical College, Mangalore, India. *Corresponding Author's Email: skrchow@gmail.com

Abstract

The oral cavity harbours one of the most diverse microbial communities in the human body, and disruption of this ecosystem is now known to contribute to caries, periodontal disease, oral cancer and a range of systemic diseases. High-throughput sequencing now characterises these communities in unprecedented detail, but the resulting data are high-dimensional, sparse, compositional and noisy, which constrains classical statistics. Artificial intelligence (AI), specifically machine learning (ML) and deep learning (DL), provides powerful tools to model this complexity, identify microbial signatures and convert community profiles into clinically useful predictions. This review outlines the state of the art of AI-powered oral microbiome analysis. It begins with the biology of the oral microbiome and the sequencing and bioinformatic pipelines used to process samples into feature tables. It then reviews the algorithms most often used for microbiome data, including regularised regression, random forests, convolutional networks and representation-learning networks, and highlights their use in periodontal disease, dental caries, oral cancer, gastrointestinal cancer and the oral–systemic axis. Common issues such as small and heterogeneous cohorts, batch effects, poor external validation and the balance between predictive performance and interpretability are discussed. Finally, future directions including explainable AI, multi-omics integration, standardised reference resources and prospective clinical evaluation are outlined. Driven by a mechanism-aware and rigorously tested approach, the integration of AI and oral microbiome analysis is poised to underpin a future of non-invasive diagnostics and personalized oral care.

Keywords: Artificial Intelligence, Deep Learning, Machine Learning, Metagenomics, Oral Microbiome, Precision Dentistry.

Introduction

The oral cavity, with several hundred bacterial taxa, along with archaea, fungi, viruses and protozoa, is a densely colonised and ecologically complex habitat, second only to the gut in microbial diversity, and with each surface of the oral cavity having its own unique flora (1). Most of these organisms can now be catalogued at species level and form the taxonomic backbone of community studies of the mouth and the aerodigestive tract (2). This community is not just a collection of residents; rather, it creates structured biofilms that play an active role in host physiology, and a transition to dysbiosis is closely associated with the most common oral diseases as well as with conditions in the rest of the body (3, 4). New-generation sequencing, which can now produce millions of reads per sample, is transforming the characterization of these communities. The data-interpretation challenge has outpaced data generation; microbiome feature tables are high-dimensional, sparse and compositional, and are confounded with technical and biological variation, characteristics that often overwhelm traditional univariate statistics. Artificial intelligence (AI) offers an alternative approach (5). Machine learning (ML) and deep learning (DL) models are widely adopted in biology and medicine because they can learn non-linear, multivariate patterns directly from data, integrate inputs from different data types and generalise to unseen samples. This review aims to summarise the current state of the use of AI in oral

(Received 15th March 2026; Accepted 12th August 2026; Published 20th August 2026)

microbiome data, its successes, its failures and its potential future directions. The availability of large-scale oral microbiome datasets creates up challenges to go beyond descriptive analysis of the microbial composition. In conventional microbiome investigations, the focus is frequently on changes in particular taxa between healthy and ill groups, although disease-associated changes might involve complex interactions among many microbes, changes in community structure and functional shifts. AI-based methods are particularly suited to this complexity, as they can simultaneously assess various microbiological variables and detect trends that would not be obvious with conventional statistical methods. This is particularly pertinent for diseases such as periodontitis, dental caries and oral cancer, because the disease-associated dysbiosis is more likely to be a microbial community rather than a single pathogenic organism.

The oral microbiome is also crucial in maintaining local ecological balance and communicating with host tissues. Microbial communities may affect immunological responses, epithelial integrity, food metabolism and resistance to colonisation by potentially dangerous pathogens. Thus, changes in this microbial equilibrium may reflect information about illness onset and progression. Importantly, microbial changes are not isolated to the oral cavity. The oral–systemic axis notion is supported by growing data linking oral microbial communities with systemic illnesses such as gastrointestinal malignancies and other diseases. Such findings have spurred interest in the use of oral microbial profiles as potential non-invasive markers of broader health condition.

The purpose of this review is to:

(a) Current uses of artificial intelligence, machine learning and deep learning in oral microbiome research including its usage in microbial profiling, biomarker discovery, disease classification and prediction are summarised.

(b) Address the use of AI-based oral microbiome analysis in oral and systemic disorders, focusing on dental caries, periodontal disease, oral cancer and the oral-systemic axis.

(c) Key challenges, limitations, and future directions of AI-based oral microbiome research include data heterogeneity, model interpretability, external validation, multi-omics integration, and the translation of AI-derived microbial signatures into clinically useful, non-invasive, and personalised diagnostic approaches.

Methodology

This article is a narrative review of the current applications of artificial intelligence in oral microbiome analysis. Peer-reviewed, English-language literature was identified through searches of the PubMed/MEDLINE, Scopus, IEEE Xplore and Google Scholar databases. Search terms combined concepts relating to the oral microbiome (oral microbiome, oral microbiota, saliva, periodontitis, dental caries, oral cancer), high-throughput sequencing (16S ribosomal RNA, shotgun metagenomics) and computational analysis (artificial intelligence, machine learning, deep learning, neural networks, classification). Foundational methodological papers widely used bioinformatic tools and representative applied studies across periodontal disease, dental caries and oral and gastrointestinal cancers were prioritized, and the reference lists of retrieved articles were screened for additional relevant sources. Because the aim was to synthesize the state of the field across biological, computational and clinical domains rather than to answer a single focused question, a narrative rather than a systematic approach was adopted and no formal quality scoring was performed. The selected literature was organized to follow the analytical workflow, from the biology of the oral microbiome and sequencing pipelines, through machine learning and deep learning methods and their clinical applications, to explainability, current limitations and future directions.

Results

Oral Microbiome in Health and Disease

The healthy oral microbiome is well-diversified and highly redundant, and community membership reflects a systematic variation among habitats within the oral cavity. The mechanisms that keep the ecosystem in check are host defence, salivary flow and inter-microbial competition, and a breakdown of these mechanisms is the basis for disease (6). The low-pH environment of dental caries favours acidogenic and aciduric bacteria, while subgingival

dysbiosis of anaerobic and proteolytic taxa produces the chronic inflammation and alveolar bone loss of periodontitis (3). Beyond the mouth, oral organisms and oral dysbiosis have been linked to cardiovascular disease, adverse pregnancy outcomes, diabetes and multiple malignancies, making the oral microbiome both a local pathogenic factor and a systemic sentinel (7). Because disease states manifest as a change in the makeup of the whole community rather than as the presence or absence of a single organism, pattern-recognition methods such as AI are very appropriate to this problem.

From Samples to Feature Tables: Sequencing and Bioinformatic Pipelines

AI models run on structured feature tables, and the quality of those features limits the performance of the model. There are two predominant sequencing strategies. Amplicon sequencing of the 16S ribosomal RNA (rRNA) gene is inexpensive and common, and determines community composition at genus or, with well-curated references, species level. Shotgun metagenomics sequences all the deoxyribo-nucleic acid (DNA) in a sample, allowing strain-level resolution and functional profiling at higher cost, and provided the foundation for the large multi-site surveys that first mapped the healthy human microbiome (7). The raw reads are processed using well-proven pipelines to generate features, including quality control, denoising, taxonomic assignment and abundance estimation. Commonly used tools include QIIME 2, which provides end-to-end reproducible analysis (8); DADA2, which performs high-resolution inference of amplicon sequence variants (9); mothur, which is used to characterise communities (10); and Kraken 2, which provides rapid metagenomic classification (11). Species-level reference databases for the oral cavity and aerodigestive tract also enhance the biological interpretability of the resulting tables (12). The output is usually a matrix of taxa or genes by samples, and it is the substrate used by all downstream AI models.

Machine Learning and Deep Learning Approaches to Microbiome Data

The mainstay of microbiome prediction is still classical machine learning. Relative-abundance features are frequently used in regularised logistic regression, support vector machines (SVM) and, most consistently, tree-based ensembles such as random forests and gradient boosting for classification of disease status. A ground-breaking meta-study across numerous metagenomic cohorts led to the formulation of a reusable computational framework and showed that random forests are likely to perform well and that feature selection can have a beneficial effect on generalisation (13). Further testing using other benchmarks confirmed that model complexity is not always beneficial, and that proper cross-validation and honest external testing are more important than the choice of model (14). These results apply directly to oral microbiome studies, which are often small.

With a large amount of data, deep learning becomes a further tool in this toolbox. Convolutional and recurrent architectures can learn taxonomic signatures directly from sequence reads, as is done for metagenomic fragments in sequence-classification frameworks without an a priori reference tree (15). Representation-learning methods, such as autoencoders, capture informative low-dimensional profiles from sparse, high-dimensional abundance data to aid disease prediction, and phylogeny-aware convolutional networks learn the evolutionary hierarchy among taxa by embedding that structure in the network, something flat feature tables miss (16). As always, there is a trade-off: while deep models can detect subtle, non-linear patterns, they require more data and careful attention to avoid overfitting, and they are harder to interpret, which limits their use in the clinic.

Applications in Periodontal Disease and Dental Caries

Saliva is a readily accessible, non-invasive sample, and this is one reason periodontal disease is among the most well-studied oral conditions for AI-based microbiome analysis. Salivary bacterial profiles have been used to train machine-learning models that distinguish periodontal disease from health, classify disease severity with meaningful accuracy and reveal informative taxa beyond the classic periodontal pathogens (17). These models point to lower-cost screening approaches that could detect

disease before it is apparent on conventional probing and radiography. Biomarker-discovery approaches that combine statistical testing with effect-size assessment have been a recurring strength of the field, revealing the distinct taxa and clades that discriminate disease groups and thus providing a source of biological hypotheses within a predictive model (18). Dental caries has been approached in a similar way, with supervised models linking community signatures of cariogenic factors to disease activity, although caries remains less studied than periodontology.

Applications in Oral and Gastrointestinal Cancers

With the promise of non-invasive cancer detection, oncology is a key driver of AI-and-oral-microbiome research. A large-scale review of learning-based algorithms used to analyze oral microbiome data concluded that models ranging from logistic regression to random forests and neural networks can reveal patterns of microbes associated with oral squamous cell carcinoma (OSCC) that could provide a cost-effective means of early detection and risk stratification, but also noted that clinical translation still requires protocol standardisation and larger, more diverse, longitudinal cohorts (19). Consistent with this, the microbiome has been shown to predict oral cancer at the community level, an example of a candidate biomarker layer (20). Integrated sequencing-plus-ML pipelines, including convolutional networks, have been applied successfully at other body sites such as the skin microbiome, illustrating approaches transferable to the oral cavity (21). Because oral organisms spread throughout the aerodigestive tract, oral microbiome signatures have also been explored as a tool to detect distant malignancies; there is systematic evidence linking certain oral taxa, such as Fusobacterium nucleatum, to colorectal cancer and supporting the potential of oral-derived signatures as a non-invasive diagnostic pathway (22).

Explainability, Biomarker Discovery and Multi-omics Integration

For AI to be used in clinical practice, predictions must be reliable and explainable. Model-based explanation methods now attribute a prediction to individual features, so that clinicians and researchers can see which taxa drive a particular prediction and learn to distinguish genuine signal from artefact (23). Reproducibility also depends on interpretability and on a shared vocabulary and consistent reporting standards across the field. Looking further, the microbiome is only one molecular layer. Incorporating taxonomic profiles together with metatranscriptomics, metabolomics, host immune markers and clinical metadata is likely to yield more comprehensive and mechanistically sound models, and the steady accumulation of oral reference genomes, including tens of thousands of metagenome-assembled genomes that reveal new taxa, will increase the dimensionality of the feature space available for such integrative learning (24).

Discussion

Challenges and Limitations

Several obstacles temper current enthusiasm. Oral microbiome cohorts are often small and single-centre, raising the risk of overfitting and of inflated performance estimates that do not survive external validation. Microbiome data are compositional and sparse, so naive application of standard algorithms can mislead unless appropriate transformations are used. Technical variation in sampling site, collection method, DNA extraction, sequencing platform and bioinformatic pipeline introduces batch effects that models may learn instead of biology, and inconsistent case definitions across studies hamper comparison. Deep models add the further problem of opacity, which is poorly compatible with regulated clinical decision-making. Finally, most published models are diagnostic classifiers evaluated retrospectively; prospective evidence that AI-guided microbiome testing improves patient outcomes is still lacking.

Future Directions

Addressing these limitations will be essential for progress. Increasing sample sizes and the number of participating centres, and implementing longitudinal cohorts with harmonised protocols, would allow models to generalise across groups, while harmonised reference resources and reporting would increase comparability (2, 25). Where accuracy and interpretability must be balanced, explainable and phylogeny-aware architectures

should be preferred (16, 23). Beyond correlation lies mechanism-based prediction, a natural extension of multi-omics and host–microbe integration. Non-invasive microbiome screening could be introduced into routine dental practice through point-of-care implementations that couple rapid sequencing or targeted assays with lightweight, validated models. Most importantly, the field needs to shift its focus from retrospective analysis to prospective clinical trials that evaluate whether AI-based oral microbiome analysis genuinely improves diagnosis, prediction and treatment selection.

Conclusion

With the advent of AI, oral microbiome sequencing, which yields high-dimensional and complex data, now has a primary means of interpretation. Machine learning and deep learning models have already identified periodontal disease, caries and several cancers from community profiles, identified candidate biomarkers and indicated non-invasive, cost-effective diagnostics. The field nevertheless remains in its infancy: small cohorts, technical heterogeneity, limited external validation and complex, opaque models all limit clinical translation. There is an urgent need for larger and more standardized datasets, models that are explainable and informed by biology, integration of multi-omics data and prospective clinical evaluation to inform the path forward. With responsible implementation, AI-powered oral microbiome analysis could transform oral healthcare, ushering in a new era of precision and prevention.

Abbreviations

AI: Artificial Intelligence, DL: Deep Learning, DNA: Deoxyribonucleic Acid, ML: Machine Learning; OSCC: Oral Squamous Cell Carcinoma, Rrna: Ribosomal RNA, SVM: Support Vector Machine.

Acknowledgement

None.

Author Contributions

SK Roy Chowdhury: conceptualization, supervision and critical review of the manuscript. Jui S Karmakar: literature review and manuscript writing. Nazia Khan: literature review and manuscript writing. Manjunath Badni: critical review and editing of the manuscript. Sumaiya Kausar: literature review and drafting of the manuscript. All authors read and approved the final version of the manuscript.

Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability

Data sharing is not applicable to this article, as no new data were created or analyzed in this review.

Declaration of Artificial Intelligence (AI) Assistance

AI-assisted technologies were not used in the preparation of this manuscript. The final content and scientific accuracy were reviewed and approved by the authors.

Ethics Approval

Ethical approval was not required, as this is a review of previously published literature and did not involve human participants or animals.

Funding

None.

References

1. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. The human oral microbiome. Journal of bacteriology. 2010;192(19):5002-17. doi: 10.1128/JB.00542-10
2. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New insights into human nostril microbiome from the expanded Human Oral Microbiome Database (eHOMD): a resource for the microbiome of the human aerodigestive tract. mSystems. 2018;3(6):e00187-18. doi: 10.1128/mSystems.00187-18
3. Kilian M, Chapple ILC, Hannig M, *et al*. The oral microbiome – an update for oral healthcare professionals. Br Dent J. 2016;221(10):657-66. doi: 10.1038/sj.bdj.2016.865
4. Gao L, Xu T, Huang G, Jiang S, Gu Y, Chen F. Oral microbiomes: more and more importance in oral cavity and whole body. Protein Cell. 2018;9(5):488-500. https://doi.org/10.1007/s13238-018-0548-1
5. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44. https://doi.org/10.1038/nature14539
6. Willis JR, Gabaldón T. The human oral microbiome in health and disease: from sequences to ecosystems. Microorganisms. 2020;8(2):308. https://doi.org/10.3390/microorganisms8020308

7. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14. doi: 10.1038/nature11234
8. Bolyen E, Rideout JR, Dillon MR, *et al*. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:852-57. https://doi.org/10.1038/s41587-019-0209-9
9. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581-83. https://doi.org/10.1038/nmeth.3869
10. Schloss PD, Westcott SL, Ryabin T, *et al*. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537-41 https://doi.org/10.1128/AEM.01541-09
11. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(257). https://doi.org/10.1186/s13059-019-1891-0
12. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. PLoS Comput Biol. 2016;12(7):e1004977. https://doi.org/10.1371/journal.pcbi.1004977
13. Topcuoglu BD, Lesniak NA, Ruffin MT, *et al.* A framework for effective application of machine learning to microbiome-based classification problems. mBio. 2020;11:e00434-20. https://doi.org/10.1128/mbio.00434-20
14. Liang Q, Bible PW, Liu Y, Zou B, Wei L. DeepMicrobes: taxonomic classification for metagenomics with deep learning. NAR Genom Bioinform. 2020;2(1):lqaa009. https://doi.org/10.1093/nargab/lqaa009
15. Oh M, Zhang L. DeepMicro: deep representation learning for disease prediction based on microbiome data. Sci Rep. 2020;10(6026). https://doi.org/10.1038/s41598-020-63159-5
16. Reiman D, Metwally AA, Sun J, Dai Y. PopPhy-CNN: a phylogenetic tree embedded architecture for convolutional neural networks to predict host phenotype from metagenomic data. IEEE J Biomed Health Inform. 2020;24(10):2993-3001. doi: 10.1109/JBHI.2020.2993761
17. Kim EH, Kim S, Kim HJ, *et al*. Prediction of chronic periodontitis severity using machine learning models based on salivary bacterial copy number. Front Cell Infect Microbiol. 2020;10:571515. doi: 10.3389/fcimb.2020.571515
18. Segata N, Izard J, Waldron L, *et al*. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(R60). doi:10.1186/gb-2011-12-6-r60
19. Soghli N, Khormali A, Mahboubi D, Peng A, Miguez PA. Recent advancements in artificial intelligence-powered cancer prediction from oral microbiome. Periodontol 2000. 2025;98(1):181-213. https://doi.org/10.1111/prd.70000
20. Su SC, Chang LC, Huang HD, *et al*. Oral microbial dysbiosis and its performance in predicting oral cancer. Carcinogenesis. 2021;42(1):127-35. https://doi.org/10.1093/carcin/bgaa062
21. Sun T, Niu X, He Q, Chen F, Qi RQ. Artificial intelligence in microbiomes analysis: a review of applications in dermatology. Front Microbiol. 2023;14:1112010. https://doi.org/10.3389/fmicb.2023.1112010
22. Negrut RL, Cote A, Maghiar AM. Exploring the potential of oral microbiome biomarkers for colorectal cancer diagnosis and prognosis: a systematic review. Microorganisms. 2023;11(6):1586. https://doi.org/10.3390/microorganisms11061586
23. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765-74. https://doi.org/10.48550/arXiv.1705.07874
24. Zhu J, Tian L, Chen P, *et al*. Over 50,000 metagenomically assembled draft genomes for the human oral microbiome reveal new taxa. Genomics Proteomics Bioinformatics. 2022;20(2):246-59. https://doi.org/10.1016/j.gpb.2021.05.001
25. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome. 2015;3(31). https://doi.org/10.1186/s40168-015-0094-5

How to Cite: Chowdhury SKR, Karmakar JS, Khan N, Badni M, Kausar S. Artificial Intelligence-driven Oral Microbiome Analysis: Current Applications and Future Directions. Int Res J Med Surg. 2026;3(3):11-16.
DOI: 10.47857/irjmeds.2026.v03i03.074