Integrating Machine Learning and Epigenetic Biomarkers to Evaluate Dental Pulp Stem Cell-Based Therapies for Periodontal Tissue Regeneration

Joline Lof
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
contactjoline@uab.edu

Iain Whittaker
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
iain.whittaker930@missouri.edu

Mirthe van Emmelen
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
mvanemmelen@oregonstate.edu

Abstract

Periodontal tissue regeneration remains a difficult translational problem because the periodontium is a composite, mechanically active, microbially exposed, and immunologically dynamic system. Dental pulp stem cells have emerged as a promising cell source for periodontal repair, yet their therapeutic behavior is shaped by donor heterogeneity, epigenetic state, senescence, immune modulation, and local niche conditions. This paper develops a system-level framework for integrating machine learning with epigenetic biomarkers to evaluate dental pulp stem cell-based therapies for periodontal tissue regeneration. The framework treats epigenetic markers as latent state variables that help explain why apparently similar cell preparations produce divergent regenerative outcomes. Machine learning is positioned not merely as a predictive instrument but as an infrastructural layer that connects multimodal data, including methylation and histone-related signatures, transcriptomic profiles, imaging, clinical periodontal measures, and manufacturing metadata. The discussion emphasizes architectural trade-offs among centralized learning, federated learning, interpretable modeling, and multimodal representation learning. It also examines deployment concerns such as data standardization, consent, model drift, validation, equity, environmental sustainability, and regulatory oversight. Particular attention is given to robustness and fairness because epigenetic and clinical datasets are often small, imbalanced, and shaped by ancestry, age, inflammation, and socioeconomic context. The paper argues that successful translation will depend less on isolated algorithmic performance and more on governance, interoperability, prospective validation, and adaptive policy. A conclusion outlines a translational agenda in which machine learning and epigenetic biomarkers jointly support safer, more reproducible, and more equitable dental pulp stem cell therapies for periodontal regeneration.

Keywords

periodontal regeneration, dental pulp stem cells, epigenetics, machine learning, biomarker discovery, federated learning, governance, translational medicine, fairness, regenerative dentistry.

1. Introduction

Periodontal disease affects a substantial share of the global adult population and remains a leading cause of tooth loss, masticatory dysfunction, and systemic inflammatory burden [8]. Conventional treatment can control infection and stabilize attachment, but true regeneration of cementum, periodontal ligament, and alveolar bone is unpredictable [9]. Cell-based therapies using dental pulp stem cells have therefore attracted sustained interest because these cells are accessible, possess multilineage potential, and can exert paracrine and immunomodulatory effects relevant to periodontal repair [6]. Nevertheless, the field has struggled to move from promising laboratory findings to reliable clinical protocols. One central reason is that therapeutic outcomes depend on complex interactions among cell source, donor biology, expansion conditions, delivery matrix, host immune state, and postoperative maintenance.

Machine learning has entered this space as a response to such complexity. Deep learning and related methods have demonstrated the capacity to identify nonlinear patterns in high-dimensional biomedical data [1], and machine learning is now widely discussed as a means to improve diagnosis, prognosis, and treatment selection in medicine [2]. In regenerative dentistry, the opportunity is not simply to predict attachment gain or bone fill; it is to construct an evaluative infrastructure that can connect molecular state, manufacturing quality, clinical context, and long-term outcome [3]. Such an infrastructure must confront algorithmic bias and inequity [4], as well as the ethical challenges that arise when opaque models influence high-stakes clinical decisions [5]. The integration of machine learning and epigenetic biomarkers therefore becomes a socio-technical problem rather than a purely computational one.

This paper argues that epigenetic biomarkers can serve as latent state variables within a machine learning-enabled evaluation system for dental pulp stem cell therapies. Epigenetic marks reflect developmental history, environmental exposures, inflammatory signaling, and senescence, and they may help explain variability that conventional surface markers cannot capture. The central question is how to design, govern, and sustain a system that uses these markers to evaluate therapies without reducing regenerative medicine to a black-box prediction task. The discussion proceeds through biological foundations, biomarker interpretation, model architecture, data infrastructure, robustness, fairness, sustainability, and policy. The aim is to provide an interdisciplinary framework for researchers, clinicians, regulators, and technology developers who must coordinate across laboratories, clinics, manufacturers, and public institutions.

2. Biological and Clinical Foundations of Dental Pulp Stem Cell-Based Periodontal Regeneration

Dental pulp stem cells were first described as postnatal mesenchymal-like cells capable of forming dentin-like structures and supporting repair in vivo [6]. They reside within a perivascular niche that regulates quiescence, activation, and differentiation, a feature shared with other postnatal mesenchymal stem cell populations [7]. In periodontal regeneration, their therapeutic value is not limited to direct differentiation into periodontal ligament fibroblasts, osteoblasts, or cementoblasts. Dental pulp stem cells also release trophic factors, modulate immune responses, and influence angiogenesis and extracellular matrix remodeling [17].

These paracrine activities mean that successful therapy depends on the cell preparation behaving as a dynamic biological agent rather than as a uniform pharmaceutical ingredient.

The periodontal wound environment is unusually demanding. It is chronically exposed to oral microbial communities, subject to mechanical loading, and characterized by spatially distinct tissues that must regenerate in coordinated fashion [8]. Periodontitis involves a dysregulated host inflammatory response that can persist after microbial control, creating a niche that may suppress regeneration or promote fibrous repair [9]. Stem cell delivery strategies must therefore address cell survival, retention, localization, and integration with host tissues [10]. Clinical feasibility studies have explored dental pulp stem cells and other mesenchymal sources for bone regeneration, but outcome variability remains a major barrier [11]. This variability is not merely technical; it reflects biological heterogeneity across donors, passage numbers, culture conditions, and recipient sites.

From a systems perspective, dental pulp stem cell therapy is best understood as a multi-stage pipeline. It begins with donor selection and tissue harvest, continues through isolation, expansion, characterization, cryopreservation, and quality control, and ends with delivery, host response, and long-term maintenance. Each stage introduces variance, and each variance source may interact with others. A cell population that appears robust in vitro may senesce rapidly, lose immunomodulatory capacity, or acquire an inflammatory phenotype after transplantation. Conversely, a preparation with modest differentiation potential may produce favorable outcomes through paracrine signaling or immune modulation [17]. Evaluation frameworks must therefore capture both intrinsic cell state and contextual performance. This requirement creates a natural bridge to epigenetic biomarkers, which can encode stable or semi-stable features of cell identity, stress history, and regenerative capacity.

3. Epigenetic Biomarkers as Latent State Variables in Regenerative Systems

Epigenetic regulation encompasses DNA methylation, histone modification, chromatin accessibility, and non-coding RNA activity. These mechanisms allow cells to respond to environmental signals without changing the underlying genome, and they are increasingly recognized as central to human disease prevention and mitigation [12]. In cancer and regenerative biology, epigenetic states can determine whether cells proliferate, differentiate, senesce, or adopt inflammatory phenotypes [13]. For dental pulp stem cells, epigenetic marks may provide a more informative representation of therapeutic potential than conventional marker panels because they integrate developmental origin, donor age, inflammatory exposure, and culture-induced stress. They are not static identifiers; they are dynamic state variables that change with expansion, cryopreservation, and transplantation.

Periodontal ligament stem cells and dental pulp stem cells occupy related but distinct developmental and functional niches [14]. Their epigenetic landscapes may influence lineage bias, immunomodulatory activity, and susceptibility to senescence. Studies of stem cells in the periodontal ligament have highlighted the importance of niche-derived signals and tissue-specific programming [16]. A machine learning system that treats epigenetic features as latent variables could therefore learn associations between molecular state and regenerative outcome while remaining aware that these associations are context-dependent. For example, a methylation signature associated with osteogenic bias may be beneficial in one delivery matrix but detrimental in an inflammatory microenvironment. The value of the biomarker lies not in fixed prediction but in its ability to mediate between biological mechanism and clinical observation.

Recent work on NUP62-mediated alleviation of senescence and promotion of stemness through NSD2-dependent epigenetic reprogramming illustrates how a single regulatory axis can reshape dental pulp stem cell behavior [15]. This example is important because it connects nuclear architecture, chromatin regulation, and senescence, three domains that are difficult to capture with standard cell surface phenotyping. In an integrated evaluation system, such a mechanism could inform feature selection, guide mechanistic interpretation, and suggest stratification strategies for patients or cell batches. However, the presence of a compelling mechanism does not automatically produce a clinically useful biomarker. The marker must be measurable at scale, robust to batch effects, reproducible across laboratories, and linked prospectively to meaningful periodontal outcomes. These requirements transform epigenetics from a laboratory curiosity into an infrastructural challenge involving assay standardization, data sharing, and longitudinal governance.

4. Machine Learning Architectures for Multimodal Biomarker and Outcome Modeling

A machine learning system for evaluating dental pulp stem cell therapies must integrate heterogeneous data types. Epigenetic assays produce high-dimensional, sparse, and often noisy measurements. Transcriptomic data provide complementary information about gene expression. Clinical data include probing depth, clinical attachment level, bleeding on probing, radiographic bone levels, and patient-reported outcomes. Manufacturing data include donor characteristics, passage number, media composition, and cryopreservation history. Imaging data may include cone-beam computed tomography, intraoral scans, and histology. No single model class can absorb all these modalities without explicit architectural choices. Supervised models are useful for predicting known outcomes, while self-supervised and representation learning approaches can exploit unlabeled data to learn latent biological structure [1]. Multimodal fusion can occur early, intermediate, or late, each with trade-offs in interpretability, missing-data tolerance, and computational cost.

Interpretability is especially important because regenerative medicine decisions affect patients, clinicians, manufacturers, and regulators. Interpretable models should be preferred when they achieve comparable performance, particularly in high-stakes settings [18]. However, explainable artificial intelligence methods applied after the fact can create false confidence, especially when explanations are unstable or do not reflect causal mechanisms [19]. A more defensible architecture may combine inherently interpretable components for clinical decision support with mechanistic modules that test epigenetic hypotheses. For example, a model could use sparse linear predictors for known senescence and inflammation pathways while using a graph neural network to represent interactions among cell state, matrix properties, and host response. The system should also quantify uncertainty, because small regenerative medicine datasets rarely support confident predictions across diverse populations and protocols.

Federated and distributed learning architectures offer one way to train models across institutions without centralizing sensitive patient data. In such systems, each clinic or laboratory trains local updates and shares only model parameters or aggregates. This approach can improve robustness and respect data sovereignty, but it introduces challenges in heterogeneous data schemas, uneven label quality, and adversarial risks. Centralized repositories may simplify harmonization and reproducibility but concentrate privacy and security risks. Hybrid architectures may be most practical, with a governed central metadata layer, local computation for sensitive records, and shared benchmark tasks for validation. The design must also accommodate model versioning, audit trails, and post-deployment

monitoring. Without these elements, a machine learning model may perform well in retrospective analysis but fail when cell manufacturing protocols, patient populations, or periodontal maintenance practices change.

5. System Architecture, Data Infrastructure, and Deployment Considerations

The technical architecture of an integrated evaluation platform is inseparable from its data infrastructure. Epigenetic biomarkers require standardized assays, reference materials, and quality control metrics. Clinical outcomes require consistent definitions and longitudinal follow-up. Manufacturing data require interoperable records that link cell batches to donors, protocols, and release criteria. In practice, these data are often trapped in proprietary laboratory information systems, electronic health records, and spreadsheets. A translational platform must therefore support data extraction, semantic mapping, provenance tracking, and privacy-preserving linkage. Privacy is not a secondary concern; medical big data systems must protect confidentiality while enabling socially valuable research [20]. Consent models must anticipate secondary use, model training, and cross-border collaboration, and they must be understandable to patients who may not distinguish between clinical care and research infrastructure.

Deployment also raises questions of workflow integration. A model that predicts regenerative potential is useless if it cannot be accessed at the point of care or if it adds unacceptable burden to clinicians and laboratory staff. User interfaces should present actionable information, such as risk of senescence, probability of favorable attachment gain, or uncertainty around a recommended cell batch. They should not ask clinicians to interpret raw epigenetic values or complex feature attributions. At the same time, the system must preserve enough transparency for audit and scientific learning. This balance suggests a layered architecture: a data layer for ingestion and harmonization, a model layer for prediction and uncertainty estimation, an application layer for clinical and manufacturing decisions, and a governance layer for access control, monitoring, and accountability. Each layer must be designed for failure, because regenerative medicine pipelines are vulnerable to sample mislabeling, batch effects, and protocol drift.

Infrastructure sustainability is often overlooked in computational biomedicine. Large multimodal models require substantial energy, storage, and specialized hardware. If every laboratory trains its own foundation model, the environmental and financial costs may be prohibitive. Shared pretrained models, efficient fine-tuning, and federated learning can reduce redundancy, but they require trust, standardization, and long-term funding. The system must also remain useful in resource-constrained settings where periodontal disease burden is high and access to advanced therapies is limited. A tiered deployment model, with lightweight local inference and optional cloud-based analysis, may be more equitable than a centralized platform that assumes reliable broadband and expensive sequencing. Sustainability therefore has technical, economic, and ethical dimensions.

6. Robustness, Fairness, and Governance

Robustness in this context means that the system performs reliably across donors, laboratories, clinical sites, and time. Epigenetic assays are sensitive to tissue handling, culture conditions, and sequencing depth. Clinical outcomes are sensitive to smoking, diabetes, oral hygiene, and socioeconomic factors. A model trained on one population may learn spurious associations that do not generalize. Robustness testing should include external validation, stress testing under missing data, sensitivity analysis to assay variation, and prospective monitoring after

deployment. It should also distinguish between biological signal and technical artifact. Without such discipline, machine learning may amplify noise rather than reduce uncertainty.

Fairness is equally central. Periodontal disease and access to regenerative care are patterned by race, ethnicity, income, insurance, and geography. Epigenetic markers can reflect social and environmental exposures, including stress, pollution, and nutrition. If a model uses these markers to allocate expensive therapies, it may reproduce or deepen existing inequities [4]. Governance must therefore include representation in training data, fairness auditing across subgroups, and mechanisms for contesting model-driven decisions. Ethical principles alone cannot guarantee responsible artificial intelligence; they must be embedded in institutional practices, accountability structures, and enforceable rules [21]. Machine learning in medicine also raises questions about consent, transparency, liability, and the responsibilities of developers, clinicians, and health systems [22]. A governance framework for dental pulp stem cell therapy should specify who owns data, who can access models, how adverse events are reported, and how patients can seek redress.

Regulatory oversight must adapt without becoming paralyzed. Regenerative medicine products are often regulated as biologics or advanced therapies, while machine learning tools may be regulated as medical devices or clinical decision support software. The intersection creates ambiguity. A model that predicts the suitability of a cell batch for a patient is not merely a laboratory test; it influences manufacturing and clinical decisions. Post-market surveillance should track model performance, drift, and unintended consequences. Transparency reports, model cards, and audit logs can support oversight, but they must be meaningful rather than performative. The most credible governance model will combine regulatory standards, professional norms, patient engagement, and independent evaluation. It will also recognize that fairness and robustness are not one-time certification criteria but ongoing operational commitments.

7. Sustainability, Policy, and Translational Pathways

Translation from research to routine practice requires more than a validated algorithm. It requires reimbursement pathways, manufacturing standards, clinical training, and public trust. Payers may hesitate to cover epigenetic testing or machine learning-guided cell selection unless evidence demonstrates improved outcomes or reduced costs. Manufacturers may resist data sharing if it exposes proprietary processes. Clinicians may adopt tools only if they fit existing workflows and do not increase liability. Policymakers therefore have a role in creating incentives for interoperability, prospective trials, and equitable access. Public funding can support independent benchmarking and long-term follow-up, while regulatory sandboxes can allow controlled innovation. International coordination is desirable because epigenetic signatures, cell manufacturing practices, and periodontal care standards vary across regions.

A sustainable translational agenda should prioritize prospective validation over retrospective discovery. The field already has many candidate biomarkers and model architectures; what it lacks is a coherent evaluation ecosystem. That ecosystem should include shared benchmark datasets with clear consent, standardized outcome definitions, and blinded external validation. It should support negative results and protocol deviations as valuable data. It should also integrate patient-reported outcomes, because periodontal regeneration is meaningful only if it improves comfort, function, and quality of life. Machine learning can help by identifying subgroups that benefit most, predicting adverse responses, and optimizing cell manufacturing.

Epigenetic biomarkers can help by providing mechanistic depth and enabling earlier quality control. Neither is sufficient alone.

Policy should further address environmental and social sustainability. Advanced therapies can be resource-intensive, and their benefits may accrue to wealthy patients in urban centers. A just translational pathway would explore decentralized manufacturing, point-of-care assays, and affordable model deployment. It would also invest in prevention and basic periodontal care, recognizing that regeneration is not a substitute for public health. The governance of data and models should be transparent to patients and communities, especially those historically harmed by medical research. Finally, sustainability requires maintenance. Models degrade as populations, protocols, and microbial environments change. Biomarker assays evolve. Governance structures must be renewed. A successful system is therefore not a finished product but an adaptive institution.

8. Conclusion

Integrating machine learning with epigenetic biomarkers offers a promising path for evaluating dental pulp stem cell-based therapies in periodontal tissue regeneration. The biological problem is inherently multifactorial, and epigenetic states can provide a deeper representation of cell quality, senescence, and regenerative potential than static marker panels. Machine learning can connect these molecular features to clinical outcomes, manufacturing parameters, and patient context, but only if it is embedded in a robust data and governance infrastructure. The most important challenges are not purely algorithmic. They include standardization, privacy, interpretability, robustness, fairness, reimbursement, environmental cost, and long-term monitoring. These challenges require collaboration among biologists, clinicians, engineers, regulators, patients, and policymakers.

A responsible framework should treat epigenetic biomarkers as latent state variables rather than deterministic predictors. It should use interpretable models where possible, quantify uncertainty, validate externally, and monitor performance after deployment. It should support federated or hybrid architectures that respect data sovereignty while enabling shared learning. It should embed fairness auditing and patient engagement into the lifecycle of the system. It should also plan for sustainability, not only in computational and environmental terms but in institutional and economic terms. If these conditions are met, machine learning and epigenetics can together improve the evaluation of dental pulp stem cell therapies and help move periodontal regeneration toward more reproducible, equitable, and clinically meaningful outcomes.

References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

2. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

3. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

4. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

5. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—Addressing ethical challenges. New England Journal of Medicine, 378(11), 981–983. https://doi.org/10.1056/NEJMp1714229

6. Gronthos, S., Mankani, M., Brahim, J., Robey, P. G., & Shi, S. (2000). Postnatal human dental pulp stem cells (DPSCs) in vitro and in vivo. Proceedings of the National Academy of Sciences, 97(25), 13625–13630. https://doi.org/10.1073/pnas.240309797

7. Shi, S., & Gronthos, S. (2003). Perivascular niche of postnatal mesenchymal stem cells in human bone marrow and dental pulp. Journal of Bone and Mineral Research, 18(4), 696–704. https://doi.org/10.1359/jbmr.2003.18.4.696

8. Pihlstrom, B. L., Michalowicz, B. S., & Johnson, N. W. (2005). Periodontal diseases. The Lancet, 366(9499), 1809–1820. https://doi.org/10.1016/S0140-6736(05)67728-8

9. Kinane, D. F., Stathopoulou, P. G., & Papapanou, P. N. (2017). Periodontal diseases. Nature Reviews Disease Primers, 3, 17038. https://doi.org/10.1038/nrdp.2017.38

10. Chen, F.-M., Sun, H.-H., Lu, H., & Yu, Q. (2012). Stem cell-delivery therapeutics for periodontal tissue regeneration. Biomaterials, 33(27), 6320–6344. https://doi.org/10.1016/j.biomaterials.2012.05.048

11. Yamada, Y., Nakamura, S., Ito, K., Sugito, T., Yoshimi, R., Nagasaka, T., & Baba, S. (2010). A feasibility of useful cell-based therapy by bone regeneration with deciduous tooth stem cells, dental pulp stem cells, or bone marrow-derived mesenchymal stem cells for clinical study using tissue engineering. Tissue Engineering Part A, 16(4), 1251–1260. https://doi.org/10.1089/ten.tea.2009.0732

12. Feinberg, A. P. (2018). The key role of epigenetics in human disease prevention and mitigation. New England Journal of Medicine, 378(14), 1323–1334. https://doi.org/10.1056/NEJMra1402513

13. Jones, P. A., Issa, J.-P. J., & Baylin, S. (2016). Targeting the cancer epigenome for therapy. Nature Reviews Genetics, 17(10), 630–641. https://doi.org/10.1038/nrg.2016.93

14. Seo, B. M., Miura, M., Gronthos, S., Bartold, P. M., Batouli, S., Brahim, J., Young, M., Robey, P. G., Wang, C. Y., & Shi, S. (2004). Investigation of multipotent postnatal stem cells from human periodontal ligament. The Lancet, 364(9429), 149–155. https://doi.org/10.1016/S0140-6736(04)16627-0

15. Wang, Xiping, et al. "NUP62 alleviates senescence and promotes the stemness of human dental pulp stem cells via NSD2-dependent epigenetic reprogramming." International Journal of Oral Science 17.1 (2025): 34.

16. Ivanovski, S., Gronthos, S., Shi, S., & Bartold, P. M. (2006). Stem cells in the periodontal ligament. Oral Diseases, 12(4), 358–363. https://doi.org/10.1111/j.1601-0825.2006.01276.x

17. Caplan, A. I., & Correa, D. (2011). The MSC: An injury drugstore. Cell Stem Cell, 9(1), 11–15. https://doi.org/10.1016/j.stem.2011.06.008

18. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

19. Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

20. Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. Nature Medicine, 25(1), 37–43. https://doi.org/10.1038/s41591-018-0272-7

21. Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

22. Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. PLoS Medicine, 15(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689