AI-Driven Epigenomic Profiling of Stem Cell Senescence and Stemness for Regenerative Dental Medicine

Geeta J. Gupta
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
ggupta59@ucf.edu

Mikkel Gutierrez
Department of Computer Science, George Mason University, Fairfax, VA, USA.
mgutierrez@gmu.edu

Ole A. Yang
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
y.ole@uab.edu

Abstract

Regenerative dental medicine is increasingly dependent on understanding the molecular states of dental pulp stem cells, particularly the balance between senescence and stemness that determines reparative capacity. Epigenomic profiling offers high-resolution insight into the regulatory programs governing these states, but the resulting datasets are large, heterogeneous, and context dependent. Artificial intelligence provides powerful representational capacities to model such complexity, yet successful translation into dental practice requires system-level design rather than isolated model development. This paper examines AI-driven epigenomic profiling for stem cell senescence and stemness across infrastructure, model architecture, governance, deployment, robustness, fairness, sustainability, and policy. A successful system must integrate multimodal data harmonization, interpretable deep learning, federated and privacy-preserving learning, rigorous biological validation, and staged clinical evaluation. Dental settings present distinctive constraints including limited computational resources, small cohort sizes, and the need to align with broader oral health equity goals. The discussion emphasizes structural trade-offs between centralized accuracy and decentralized confidentiality, between model complexity and clinical explainability, and between rapid translation and long-term safety monitoring. By addressing these dimensions, AI-driven epigenomic profiling can evolve from a promising computational concept into a trustworthy component of regenerative dental medicine.

Keywords

artificial intelligence; epigenomics; cellular senescence; dental pulp stem cells; regenerative medicine; data governance.

1. Introduction

Regenerative dental medicine has moved beyond the mechanical restoration of damaged tissues toward biological strategies that seek to activate endogenous repair or engineer replacement structures. At the center of this transition is the behavior of dental pulp stem cells,

which retain the capacity to differentiate into multiple cell types and to support dentin repair. However, their therapeutic performance is constrained by the gradual accumulation of cellular senescence, a state characterized by irreversible growth arrest and the secretion of inflammatory factors. At the same time, stemness, understood as the preservation of self-renewal and multilineage differentiation potential, is required for sustained regenerative function. Artificial intelligence has been examined broadly as a methodological infrastructure for health care [1], and the convergence of human and machine intelligence is increasingly viewed as a necessary condition for high-performance medicine [2]. In the dental context, this convergence can be directed toward resolving the molecular ambiguity between senescence and stemness through epigenomic profiling.

Deep learning provides a general representational framework for detecting hierarchical patterns in high-dimensional data [3], and its adaptation to genomics has expanded the analytical repertoire available to computational biology [4]. Epigenomic measurements are particularly well suited to these methods because they contain distributed regulatory signals that are not easily captured by traditional linear models. Early work in deep learning for computational biology demonstrated that neural networks could learn biological structure from raw molecular profiles without relying exclusively on hand-engineered features [5]. In the present application, such structure includes DNA methylation changes, histone modification patterns, chromatin accessibility, and noncoding RNA activity. These layers jointly define the epigenetic landscape that shapes whether a dental pulp stem cell remains quiescent and stem-like or enters a senescent state.

Epigenomic aging has been formalized through methylation clocks that estimate biological age from a limited number of CpG sites [6], establishing the principle that epigenetic state is both measurable and predictive. The hallmarks of aging framework further situates cellular senescence and stem cell exhaustion as interconnected processes that drive age-related decline [7]. For regenerative dental medicine, this means that a computational system must not treat senescence as a single marker but as a dynamic property of a larger regulatory network. AI-driven profiling therefore requires an architectural perspective that spans data acquisition, normalization, model training, validation, deployment, and governance. Without such a perspective, even accurate models may fail under clinical conditions because of data shift, privacy constraints, or lack of interpretability.

2. Epigenomic Data Infrastructure and Systemic Complexity

The construction of an AI-driven epigenomic profiling system begins with the recognition that genome-wide methylation analysis is analytically demanding and subject to platform-specific artifacts [8]. Different sequencing and array technologies produce different coverage distributions, error profiles, and normalization requirements. A regenerative dental medicine program must therefore design data infrastructure capable of ingesting heterogeneous methylome data while preserving biological variation. The dynamic nature of DNA methylation across human tissues further complicates this task because regulatory regions may be methylated differently depending on developmental stage, anatomical site, and disease context [9]. For dental pulp stem cells, these differences are important because methylation states at enhancer and promoter regions can gate the expression of genes that control senescence and stemness.

System-level analysis of DNA methylation data requires more than preprocessing pipelines. It demands statistical integration across molecular layers and careful distinction between technical variability and true biological signal [10]. In practice, a dental research consortium

may collect whole-genome bisulfite sequencing, reduced representation bisulfite sequencing, and targeted array data from different institutions. Each data type has distinct sparsity, dynamic range, and missingness patterns. An infrastructure that harmonizes these modalities must include metadata standards, versioned reference epigenomes, and reproducible processing workflows. It must also be designed with the understanding that epigenomic state is not a fixed property but a trajectory influenced by donor age, inflammation, mechanical stress, and therapeutic manipulation. This systemic complexity is one of the primary reasons why traditional bioinformatics approaches struggle to produce robust clinical markers for dental pulp stem cell quality.

Beyond technical harmonization, the infrastructure must support longitudinal sampling and multimodal linkage. A single methylome snapshot may reveal correlative patterns, but it cannot by itself distinguish causal regulators from downstream consequences. Therefore, the system should be organized around integrated data objects that combine methylation, histone modification, chromatin accessibility, transcriptomic, and proteomic information when available. Such integration creates a richer representation of regulatory state and enables AI models to learn shared latent factors across omics layers. In dental medicine, where sample sizes are often smaller than in oncology or cardiovascular genomics, careful infrastructure design can partially compensate by reducing noise and increasing signal density before modeling even begins.

3. AI Architectures for Senescence and Stemness Profiling

The choice of AI architecture for epigenomic profiling must reflect the structure of biological sequence data and the need for context-dependent prediction. Sequence-based models have demonstrated that deep networks can predict molecular phenotypes from DNA sequence by integrating long-range interactions [11]. Attention mechanisms, originally developed for natural language processing, enable models to weigh the importance of distal regulatory elements when predicting gene expression or chromatin state [12]. In the stem cell senescence domain, such models can be used to identify regulatory regions that become differentially methylated with aging or that switch their chromatin state during the transition from stemness to senescence. The architecture must therefore be capable of learning representations that span multiple genomic scales, from individual CpG sites to broad chromatin domains.

However, the application of deep learning in biology and medicine is not without obstacles [13]. Clinical translation requires careful handling of label noise, class imbalance, and the limited availability of well-annotated samples. In dental pulp stem cell research, senescence may be experimentally induced by oxidative stress, replicative exhaustion, or pharmacological agents, but these protocols do not always produce identical epigenetic states. An AI architecture that learns a shared senescence latent space across different induction models can improve generalization, but it also risks collapsing biologically meaningful differences. This trade-off mirrors broader challenges in machine learning for medicine, where model complexity must be balanced against the need for actionable and clinically interpretable outputs [14]. For regenerative dentistry, the ideal architecture is not necessarily the deepest or most parameter-rich, but one that combines predictive performance with sensitivity to the biological pathways relevant to pulp repair.

Multimodal architectures offer a promising direction. A model can be organized as a set of modality-specific encoders feeding a shared representation layer, followed by task-specific heads for classifying senescence state, predicting stemness potential, or estimating epigenetic age. Attention-based fusion can learn to weight methylation and chromatin features according

to their relevance in a given biological context. Transfer learning from larger epigenomic datasets may provide useful initial representations, but dental-specific fine-tuning is essential because oral tissues have distinct developmental origins and regenerative constraints. The system should also generate uncertainty estimates and attribution maps so that clinicians and researchers can assess whether a prediction depends on biologically plausible regulatory elements rather than technical artifacts.

4. Data Governance, Privacy, and Biological Validation

Governance is a central rather than peripheral concern for AI-driven epigenomic profiling. Epigenomic data are inherently identifiable and sensitive because they can reveal information about disease risk, aging, and potential regenerative capacity. Federated learning has been proposed as a mechanism to facilitate multi-institutional collaborations without sharing raw patient data [15]. In this model, local institutions train models on their own data and share only model updates with a central coordinator. Swarm learning extends this concept by enabling decentralized learning without a central aggregator, thereby reducing the risk of a single point of failure or control [16]. These approaches are appealing for dental consortia that must reconcile privacy regulations with the statistical need for larger training populations. Privacy-preserving generative models can also support data sharing by producing synthetic epigenomic profiles that retain population-level structure while reducing re-identification risk [17]. The system-level decision is not simply whether to use one of these methods, but how to combine them within a governance framework that covers access control, auditability, consent, and breach response.

Biological validation must be embedded in the same governance structure. A recent experimental investigation demonstrated that a nucleoporin-associated mechanism alleviates senescence and promotes stemness in human dental pulp stem cells through NSD2-dependent epigenetic reprogramming [18]. Such findings illustrate the type of mechanistic knowledge that AI models should be designed to surface rather than obscure. When a model identifies a candidate regulatory region or pathway, the result must be testable in reductionist experiments using dental pulp stem cell cultures. Validation requires iterative feedback between computational prediction and laboratory confirmation, with explicit criteria for accepting or rejecting a candidate biomarker. This integration also protects against overfitting to cohort-specific noise and helps ensure that AI-derived signatures correspond to reproducible biological processes.

The governance architecture must manage the special status of epigenomic data throughout the model lifecycle. Consent forms should distinguish between clinical use, research use, and commercial algorithm development. Data provenance must be preserved so that a model trained on heterogeneous populations can be audited for representation gaps. Furthermore, the system should maintain versioned linkage between raw data, processed features, model outputs, and experimental validation records. Without this traceability, it becomes difficult to determine whether a deployment failure arises from data drift, model degradation, or a flawed biological assumption. In dental regenerative medicine, where clinical translation may proceed through small pilot studies, such traceability is especially important because errors can propagate quickly without large external datasets to expose them.

5. Deployment and Clinical Integration in Regenerative Dental Medicine

Deployment of AI-driven epigenomic profiling must be integrated with the broader trajectory of stem cell biology and regenerative therapeutics. The demonstration that differentiated cells

can be reprogrammed to a pluripotent state established a powerful conceptual foundation for manipulating cell fate [19]. In the dental field, the goal is rarely full pluripotency but rather controlled enhancement of dental pulp stem cell stemness for dentin repair and pulp regeneration. At the same time, the aging of stem cells involves cell-intrinsic and niche-dependent changes that reduce regenerative capacity [20]. A clinically useful AI system must therefore provide information not only about whether a cell population is senescent, but also about how it can be therapeutically modulated. This implies that model outputs should be linked to actionable variables such as culture conditions, epigenetic modifiers, or donor selection criteria.

Clinical integration presents structural challenges distinct from model development. A dental hospital or research clinic often lacks the computational infrastructure of a large genomic center. Deployment strategies may therefore prefer lightweight inference models that can operate on edge devices or institutional servers rather than requiring continuous cloud connectivity. Model compression, quantization, and knowledge distillation can reduce the computational footprint, but these techniques must be validated to ensure that they do not distort epigenetic attribution. A further issue is the temporal mismatch between model training and clinical use. Epigenomic assays may evolve, sequencing platforms may change, and patient populations may shift over time. The deployment system must include continuous monitoring of input distributions, calibration of predicted probabilities, and scheduled revalidation against newly labeled samples.

The clinical decision support layer should translate model outputs into formats that align with dental workflows. For example, a predicted senescence score may be paired with a recommended assay panel or a suggested culture intervention. However, such recommendations must be treated as decision aids rather than autonomous directives. Dental clinicians, tissue engineers, and bioinformaticians should jointly review model outputs for high-stakes decisions such as the use of autologous stem cells in regenerative procedures. Staged deployment is advisable, beginning with retrospective evaluation on biobanked samples, followed by prospective observational studies, and only then by interventional trials. Each stage generates evidence not only about predictive accuracy but also about usability, workflow disruption, and unintended consequences.

6. Robustness, Fairness, and Risk Management

Robustness and fairness are often treated as technical afterthoughts in genomic machine learning, but they are structural requirements for responsible deployment. Algorithmic bias in health care can arise from seemingly neutral design choices when training data fail to represent the populations that will ultimately be served [21]. In AI-driven epigenomic profiling, this problem is compounded by epigenetic variation across ancestry, age, sex, environmental exposure, and oral health history. If a model is trained predominantly on data from one demographic group, its senescence predictions may be systematically miscalibrated for others. The system must therefore include fairness audits that evaluate performance across population subgroups and that assess the potential for differential harm. Such audits should be conducted before deployment and repeated as the model is updated.

Risk management extends beyond fairness to encompass model drift, adversarial perturbation, and overreliance on single biomarker classes. A responsible machine learning framework for health care emphasizes prospective clinical evaluation, transparent reporting, and continuous monitoring [22]. In the dental context, a model that misclassifies a highly senescent cell population as stem-like could lead to ineffective regenerative therapy, while the opposite error

may result in unnecessary discard of potentially therapeutic cells. These asymmetric risks should be encoded in decision thresholds and in the economic evaluation of model deployment. The system should also distinguish between epistemic uncertainty, which can be reduced with more data, and aleatoric uncertainty, which arises from inherent biological variation. Communicating both forms of uncertainty to clinicians is essential for appropriate trust calibration.

Accountability requires a clear assignment of responsibility across model developers, data stewards, and clinical users. When a model recommendation is followed and harm occurs, the causal pathway may involve data quality, model error, clinician interpretation, or patient-specific factors. A robust risk management system should maintain audit logs that capture model version, input data characteristics, output predictions, and human decisions. These logs enable retrospective analysis and facilitate regulatory reporting. In regenerative dental medicine, where the use of AI may be combined with cell therapy, the regulatory landscape may involve both medical device frameworks and biologics oversight. The AI governance system must be designed to accommodate this overlap rather than assuming a single regulatory category.

7. Sustainability and Policy Implications

Sustainability in AI-driven epigenomic profiling has multiple dimensions: computational, institutional, economic, and environmental. Training large models on genome-wide epigenomic data consumes substantial energy and may require specialized hardware. Smaller dental research groups may be unable to bear these costs independently, which raises questions about open infrastructure and shared compute resources. Healthy aging research has emphasized the value of preventive and regenerative strategies as long-term investments [23], and the same logic applies to the computational infrastructure needed to support them. A sustainable system should favor modular designs in which models can be retrained incrementally rather than from scratch, and in which shared reference datasets reduce redundant preprocessing. Open-source tooling, public model registries, and community-defined standards can lower entry barriers while preserving scientific rigor. However, open sharing must be balanced against privacy and the risk of misuse of sensitive epigenomic data.

Policy implications follow from the recognition that AI-driven epigenomic profiling operates at the intersection of data protection, regenerative medicine, and oral health equity. Large-scale biological AI projects have demonstrated that centralized infrastructure can produce broadly useful resources, but they also illustrate the concentration of expertise and compute in a small number of institutions [24]. For dental medicine, this concentration may widen existing disparities if only well-resourced centers can deploy advanced profiling. Policy makers should consider funding distributed networks that include community dental clinics and academic institutions serving underrepresented populations. Data sharing agreements should promote interoperability while respecting sovereignty and consent, particularly for populations that have historically experienced exploitative research practices.

International coordination is also necessary because epigenetic reference data, model updates, and therapeutic protocols may originate in different jurisdictions. Regulatory harmonization can reduce duplication and enable multi-country validation studies, but it must not dilute privacy protections. In parallel, the dental profession should develop competency frameworks for interpreting AI-derived epigenomic information, ensuring that clinicians are not positioned as passive recipients of algorithmic outputs. The long-term vision is not the replacement of dental expertise by automated profiling, but the augmentation of clinical

judgment with transparent, reproducible, and context-aware molecular evidence. Achieving this vision requires sustained investment in both computational infrastructure and the human capacity to use it responsibly.

8. Conclusion

AI-driven epigenomic profiling of stem cell senescence and stemness has the potential to transform regenerative dental medicine by providing molecular insight into the quality and therapeutic potential of dental pulp stem cells. However, the path from computational model to clinical utility is a systems problem rather than a narrow machine learning task. It requires careful attention to data infrastructure, architectural design, governance, privacy, biological validation, deployment, robustness, fairness, sustainability, and policy. The trade-offs are structural: centralized accuracy must be weighed against decentralized confidentiality, model complexity against interpretability, and rapid translation against long-term safety. Dental medicine brings distinctive constraints related to small cohorts, limited computational resources, and the need for equitable access. A well-designed system integrates these constraints from the outset, combining multimodal epigenomic data with interpretable and federated learning approaches. It embeds AI predictions within experimental validation cycles and clinical workflows, and it subjects deployed models to continuous monitoring and fairness audits. By approaching AI-driven epigenomic profiling as an infrastructure challenge rather than a collection of isolated models, researchers and clinicians can develop trustworthy tools that support the regenerative promise of dental pulp stem cells while maintaining scientific rigor and social responsibility.

References

1. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.

2. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

3. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

4. Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: New computational modelling techniques for genomics. Nature Reviews Genetics, 20(7), 389–403.

5. Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. Molecular Systems Biology, 12(7), 878.

6. Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biology, 14(10), R115.

7. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. Cell, 153(6), 1194–1217.

8. Laird, P. W. (2010). Principles and challenges of genome-wide DNA methylation analysis. Nature Reviews Genetics, 11(3), 191–203.

9. Ziller, M. J., Gu, H., Müller, F., Donaghey, J., Tsai, L. H. Y., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., Gnirke, A., & Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature, 500(7463), 477–481.

10. Teschendorff, A. E., & Relton, C. L. (2018). Statistical and integrative system-level analysis of DNA methylation data. Nature Reviews Genetics, 19(3), 129–147.

11. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., Assael, Y., Jumper, J., Kohli, P., & Kelley, D. R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. Nature Methods, 18(10), 1196–1203.

12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

13. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P. M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., ... Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society Interface, 15(141), 20170387.

14. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

15. Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R. R., & Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 10(1), 12598.

16. Warnat-Herresthal, S., Schultze, H., Shastry, K. L., Manamohan, S., Mukherjee, S., Garg, V., Sarveswara, R., Händler, K., Pickkers, P., Aziz, N. A., Ktena, S., ... Schultze, J. L. (2021). Swarm learning for decentralized and confidential clinical machine learning. Nature, 594(7862), 265–270.

17. Beaulieu-Jones, B. K., Wu, Z. S., Williams, C., Lee, R., Bhavnani, S. P., Byrd, J. B., & Greene, C. S. (2019). Privacy-preserving generative deep neural networks support clinical data sharing. Circulation: Cardiovascular Quality and Outcomes, 12(7), e005122.

18. Wang, X., Wang, L., Zhou, L., Chen, L., Shi, J., Ge, J., ... & Zou, D. (2025). 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), 34.

19. Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell, 126(4), 663–676.

20. Sharpless, N. E., & DePinho, R. A. (2007). How stem cells age and why this makes us grow old. Nature Reviews Molecular Cell Biology, 8(9), 703–713.

21. 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.

22. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337–1340.

23. Kaeberlein, M., Rabinovitch, P. S., & Martin, G. M. (2015). Healthy aging: The ultimate preventative medicine. Science, 350(6265), 1191–1193.

24. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589.