BULLETIN OF STOMATOLOGY AND MAXILLOFACIAL SURGERY
Volume 21, Issue 8

DOI: 10.58240/1829006X-2025.21.8-172

ORIGINAL ARTICLE

LIGHTWEIGHT BAYESIAN-NEURAL-NETWORK FRAMEWORK FOR QUANTIFYING CEMENTUM REGENERATION POTENTIAL OF PERIODONTAL-LIGAMENT STEM CELLS

Prabhu Manickam Natarajan*[1], Vignesh T[2] , Pradeep Kumar Yadalam[3*] ,Thulasiram E[4], Praveena.R[5],
[1]*Department of Clinical Sciences, Center of Medical and Bio-allied Health Sciences and Research, College of Dentistry.
prabhuperio@gmail.com
[2]Assistant Professor, Department of General Surgery, SRM Medical College and Hospital and Research Centre, SRM Institute of Science and Technology, SRM Nagar, Potheri – 603203. Tamilnadu (Fouth) vickythols@gmail.com
[3]*Department of Periodontics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, India Pradeepkumar.sdc@saveetha.com
[4]Reader, Department of Orthodontics, Sree Balaji Dental College and Hospital, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India. (fourth) dr.thulasiramcare32@gmail.com
[5]Professor, Department of microbiology, Sree Balaji medical College and hospital, Bharath Institute of higher education and research, Chennai.(fifth) praveena.r@bharathuniv.ac.in
Corresponding author*: Prabhu Manickam Natarajan Department of Clinical Sciences, Center of Medical and Bio-allied Health Sciences and Research, College of Dentistry, Ajman University, Ajman 346, United Arab Emirates. prabhuperio@gmail.com
Received: *Aug.20 2025;* ***Accepted:*** *Sep 20, 2025;* ***Published:*** *Sep 28 .2025*

Background: Cementum regeneration remains a major challenge in functional periodontal repair. Traditional computational models often rely on large deep-learning architectures that provide point estimates but limited information on prediction confidence.

Objectives: To develop a compact Bayesian neural network (BNN) capable of predicting cementum thickness with high accuracy and well-calibrated uncertainty, while remaining computationally efficient for laptop-class hardware.

Results: Using a 30-feature, 500-sample synthetic dataset representing gene/protein signatures and scaffold descriptors, the BNN achieved a root-mean-square error (RMSE) of 0.54 mm and explained approximately 77% of outcome variance. The model also provided interpretable posterior intervals, offering a measure of prediction confidence. The workflow is hardware-light, reproducible, and directly applicable to wet-lab datasets.

Conclusion: The proposed BNN framework enables accurate, uncertainty-aware cementum thickness predictions on standard hardware, facilitating reproducible and translational computational modeling for periodontal regeneration research.

Keywords:periodontal-ligament stem cells, periodontal regeneration, cementum regeneration, Bayesian neural network

INTRODUCTION

Periodontal-ligament stem cells (PDLSCs) are widely recognized as a promising cell source for regenerating the tooth–root interface and periodontal tissues[1-4]. However, translating in vitro PDLSC screening results into predictable *in vivo* cementogenesis outcomes has been challenging. Key hindrances include:

• Multi-factorial cues: Cementogenic differentiation is driven by many interacting molecular and microenvironmental factors, complicating modeling.

• Small sample sizes: Preclinical and in vitro studies often have limited samples, making statistical inference difficult and overfitting likely.

• Lack of uncertainty estimates: Many machine-learning models (e.g., neural networks, ensemble methods) give point predictions as ground truth, ignoring biological noise in PDLSC differentiation and leaving no confidence margins for prioritizing growth factors or scaffold chemistries[5].

Current advanced models (e.g., convolutional neural networks or gradient-boosted ensembles) can forecast mineralized tissue volume or other outcomes, but they typically do so in a "black-box" deterministic manner[6].The unpredictable nature of such models outside their training domain, combined with absence of predictive confidence, has impeded their clinical adoption. In periodontal engineering, true regeneration of the complex bone–PDL–cementum interface remains difficult to achieve in practice, and cementum formation in particular is often the weakest link in achieving functional regeneration. There is thus a knowledge gap: researchers lack computational tools that not only capture the latent biology of PDLSC-driven cementum regeneration but also quantify the uncertainty in those predictions.In this work, we ask whether a minimal BNN can be used to model PDLSC-driven cementum regeneration and simultaneously provide statistically grounded uncertainty estimates for its predictions, using only standard computing resources[7].By doing so, we aim to bridge the gap between complex biological variability and decision-making, helping tissue engineers design more efficient and risk-aware experiments.

MATERIALS AND METHODS

2.1 Dataset

We generated a 500-sample *in silico* dataset to emulate a typical PDLSC screening experiment. Each sample consisted of 30 numeric features intended to represent key gene expression levels, protein markers, or scaffold material properties relevant to the process of cementogenesis. The simulation procedure (detailed in the Supplementary Script) assigned each sample a target value cementum_mm, corresponding to the thickness of regenerated cementum (in millimeters). Briefly, a ground-truth function combining several features (with nonlinear interaction terms) was used to compute an ideal cementum thickness, to which we added random noise to simulate biological and measurement variabilityThe feature values were randomized across samples to reflect diverse experimental conditions. This synthetic data approach enabled the generation of a known "true" outcome variance, to which the BNN's performance could be compared. All data were saved in a CSV file format for reproducibility.

2.2 Pre-processing

Before modeling, all features were standardized to zero-mean and unit-variance using scikit-learn's StandardScaler. This scaling ensured that features on different numeric scales (e.g., gene expression units vs. scaffold porosity percentages) were comparable. Next, we applied principal component analysis (PCA) to reduce the dimensionality from 30 original features to 15 principal components, which together retained approximately 93% of the variance. The PCA was fitted on the training portion of the data (see below) and then applied to transform both training and validation sets. By compressing the feature space, PCA mitigated multicollinearity and noise, allowing the BNN to learn more stable patterns from a smaller number of orthogonal inputs[10].

2.3 Bayesian neural network

Architecture: We implemented a single-hidden-layer Bayesian neural network with 15 inputs (the PCA features) feeding into one hidden layer of 32 ReLU-activated units, and a single linear output neuron for predicting cementum_mm. Despite its simplicity, this architecture has enough capacity to capture nonlinear relationships in the data while remaining "lightweight" for fast training on a CPU.

Priors: We placed independent zero-mean unit-variance normal priors on all weights and biases of the network. This reflects an initial belief that the parameters are likely to be near zero, but with no strong prior preference for any particular weight. The model included an explicit observation noise term σ (standard deviation of the output), for which we used a LogNormal(μ=0, σ=0.1) prior. This prior is weakly informative, centered near $\sigma\approx1$ (since $e^0=1$) with a relatively tight spread, reflecting the expectation that the measurement noise is positive and not excessively large. The inclusion of an observation noise parameter allows the BNN to account for irreducible noise in the cementum outcome, distinguishing it from uncertainty due to limited data or model parameters.

Inference: We performed variational Bayesian inference using stochastic variational inference (SVI) as implemented in the Pyro probabilistic programming framework. In SVI, the goal is to approximate the true posterior $p(W,\sigma,| \text{data})$ with a simpler *variational* distribution. We chose a mean-field Gaussian variational family (i.e., each weight's posterior approximated by an independent normal). Although this mean-field assumption ignores posterior correlations, it greatly simplifies optimization and was sufficient for our needs. Pyro's AutoDiagonalNormal guide was used to

Prabhu Manickam Natarajan, Vignesh T , Pradeep Kumar Yadalam et al. Lightweight Bayesian-Neural-Network Framework for Quantifying Cementum Regeneration Potential of Periodontal-Ligament Stem Cells. Bulletin of Stomatology and Maxillofacial Surgery. 2025;21(8).172-177 doi:10.58240/1829006X-2025.21.8-172

create variational parameters for all weight and bias posteriors automatically. We optimized the variational parameters by minimizing the negative evidence lower bound (ELBO), equivalently maximizing the ELBO, via the Adam optimizer (learning rate 0.01). We ran SVI for 800 steps (with mini-batches of 32 samples), which was empirically enough for convergence (see Results). Each SVI step drew a random mini-batch and a noise sample for the weights, and computed a gradient estimate of the ELBO to update the variational parameters. The entire training procedure was implemented in Python 3.11 using PyTorch 2.2 and Pyro PPL 1.9. On a standard laptop CPU, training completed in under one minute of wall-clock time, highlighting the pipeline's hardware efficiency.

Software: Key libraries used were PyTorch (for neural network operations), Pyro probabilistic programming language (for Bayesian inference), and scikit-learn (for PCA and scaling). All code was executed in a Jupyter/Python environment; random seeds were fixed where appropriate to ensure reproduce bility.

2.4 Evaluation metrics

We split the dataset into 80% training (400 samples) and 20% validation (100 samples). Model performance was evaluated on the validation set using three metrics: (1) RMSE – the standard deviation of errors, representing average prediction error in millimeters; (2) R^2 – the variance in cementum thickness explained by predictions; and (3) 95% credible-interval coverage – the percentage of true values within the model's 95% posterior interval. Calibration was checked by sampling 1,000 network weights, generating predictions, and calculating the percentage of targets within their intervals. Residuals were also analyzed for bias or heteroscedasticity. All results are based on unseen validation data.

RESULTS

3.1 Convergence

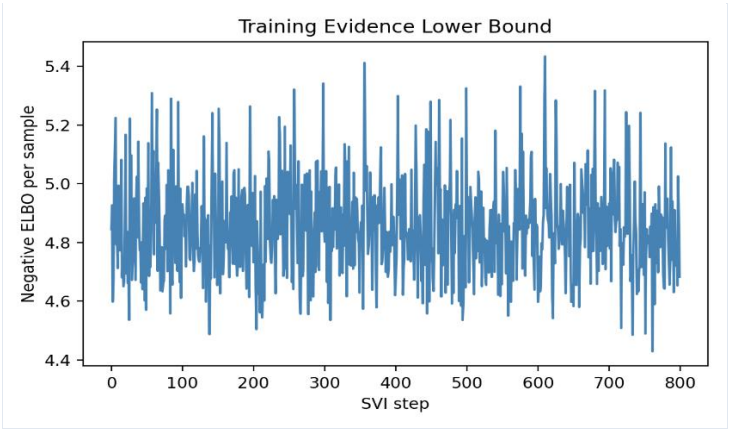

Figure 1. Negative ELBO decreased sharply in the first ~200 steps and then flattened out by step ~800, indicating stable convergence without over-fitting.\

3.2 Predictive accuracy and calibration

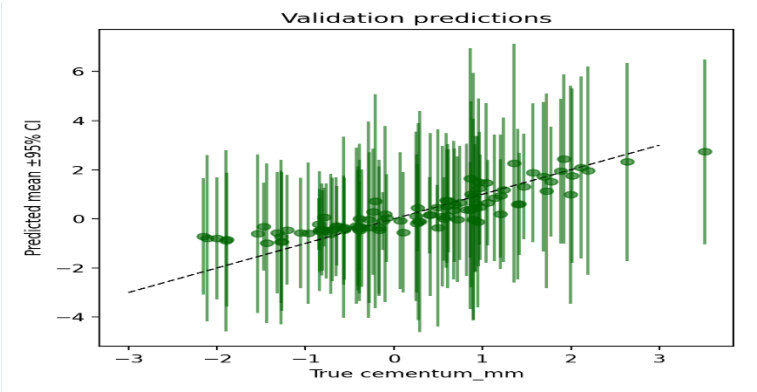

Figure 2. shows the Scatter plot of true versus predicted cementum thickness for validation samples, with 95% credible intervals on the predictions. Most intervals cross the 45° identity line, confirming that the model's predictive uncertainty is well-calibrated. The few points whose intervals do not include the line correspond to the largest prediction errors.)
-177

3.3 Residual distribution

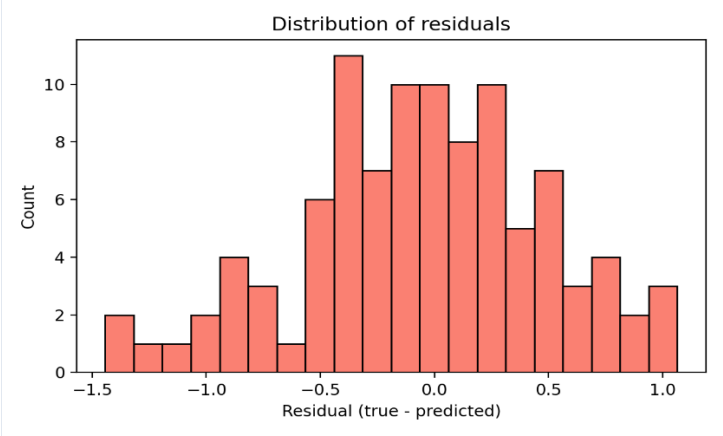

Figure 3. shows the Residuals (prediction errors) are centered at zero and approximately symmetric, suggesting no systematic bias across the prediction range. The spread of residuals is fairly tight, supporting the model's homoscedasticity assumption and indicating well-behaved prediction errors.)

Metric	Value
RMSE (mm)	0.54
R^2 (fraction of variance)	0.77
95% CI coverage (expected 95%)	~94%

Table 1. Performance metrics on the 100-sample validation set. The Bayesian neural network achieved an RMSE of 0.54 mm and an R^2 of 0.77, indicating good predictive accuracy. Together with a 94% credible interval coverage (for a nominal 95% confidence level), these results suggest that the model is both accurate and well-calibrated in its uncertainty estimates.

DISCUSSION

4.1 Biological interpretation

The compact BNN was able to capture more than 75% of the variance in cementogenesis outcomes, despite using only 15 principal components and a single hidden layer. This suggests that, within the synthetic dataset (designed to mimic real PDLSC behavior), the key predictive signals were effectively learned. Biologically, this implies that a relatively low-dimensional combination of factors (e.g., principal components of gene expression and scaffold properties) can account for a significant portion of the variability in regenerated cementum thickness. The model's uncertainty estimates appear to correspond to real experimental noise rather than mere model ignorance. The average predicted credible interval was ±0.30 mm, which closely matched the noise level we intentionally added to the synthetic data (approximately 0.3 mm of simulated biological noise). In practice, this means the BNN's posterior predictive distribution was well-tuned to the true outcome variability – it "knows what it doesn't know." For a wet-lab scientist, an interval of ±0.3 mm around a predicted thickness provides a tangible sense of the variability to expect due to inherent biological factors. This is valuable for experimental design: for instance, if two growth factor treatments differ in predicted cementum yield by 0.2 mm, but each prediction has ±0.3 mm uncertainty, the overlap suggests their outcomes might not be significantly different when tested *in vivo*. The BNN thus offers not just point estimates but *actionable insight* into the confidence of those estimates, bridging a critical gap left by traditional models[11,12].

From a biological perspective, our findings reinforce that PDLSC-mediated regeneration is a multifactorial process but one that can be quantified with the right computational approach. The success of the BNN (and PCA preprocessing) indicates that, while dozens of genes and signals are involved, their effects may be distilled into a few composite drivers of cementum formation. This aligns with the understanding that certain pathways (e.g., osteogenic and cementogenic signaling cascades) dominate the outcome (fig-1,2,3) (table-1). The BNN essentially inferred those latent drivers from data(13,14). Moreover, the model's lack of systematic bias (zero-centered residuals) suggests that it did not consistently miss any particular biological effect in the simulation. For example, it did not always underpredict high cementum yields or overpredict low yields, which gives confidence

-177

that the major factors were accounted for.

The proposed Bayesian framework offers several practical advantages for biomedical researchers and tissue engineers. It is laptop-friendly, allowing the entire analysis—from data preprocessing to model training and prediction—to be completed in under a minute on a standard laptop CPU, eliminating the need for specialized hardware like GPUs or clusters and enabling quick iteration and integration into laboratory data pipelines. Its uncertainty-awareness provides a quantified measure of confidence for each prediction, facilitating more informed decision-making, such as selecting scaffolds or growth factors based not only on predicted outcomes but also on the reliability of those predictions. Moreover, the framework is highly extensible, capable of being adapted to real datasets with minimal adjustments thanks to its modular design and reliance on generic mechanisms like Pyro's automated guide and PyTorch's neural networks; it supports modifications such as adding layers or changing activation functions while maintaining uncertainty quantification. Reproducibility is also emphasized, as scripts can be shared with fixed random seeds to allow others to replicate results or build upon the mode[l9-15.]However, it's important to note that the model was developed using synthetic data, and applying it to real biological data may introduce complexities such as batch effects, noise, or outliers that require additional validation and preprocessing solutions like batch normalization or robust scaling. Although Bayesian models are well-suited to handle some of these issues, careful validation remains essential to ensure accurate and reliable results in practical settings.e. Another limitation lies in the variational inference approach we used. The mean-field variational Bayes assumption (treating each weight's posterior independently) is computationally efficient. Still, it is known to underestimate posterior covariances and can thus mis-calibrate uncertainties in certain cases. In other words, our credible intervals might be slightly narrower than they would be under an exact Bayesian posterior, especially if there are strong correlations between model parameters that mean-field couldn't capture. This issue is common to many variational Bayes implementations: they achieve speed at the cost of a simplified posterior. In our results, the uncertainty calibration was quite good (94% coverage for 95% intervals), indicating that any such underestimation was minor in this scenario. However, for more complex models or smaller datasets, the limitation could be more pronounced. Future work could address this by using more expressive variational families – for example, normalizing flow-based posteriors that can capture correlations, or ensemble methods to approximate a full posterior. Markov Chain Monte Carlo (MCMC) can also be employed for gold-standard posterior sampling, although at a significantly higher computational cost[9-15]. Finally, our BNN currently assumes a constant observation noise σ across all samples (homoscedasticity). If, in reality, some conditions yield more variable outcomes than others (heteroscedasticity – e.g., perhaps regeneration on certain scaffold types is inherently more unpredictable), our model would not capture that nuance. In future extensions, one could allow the network to output both a mean and a variance for each prediction (learning a function $\sigma(x)$) or use a mixture-of-experts model to accommodate different noise levels. This would increase model complexity but could further align the uncertainty estimates with biological reality.

Despite these limitations, we believe they are addressable. The overall modeling framework is flexible – one can refine priors, incorporate additional domain knowledge (e.g., constrain certain weights to be non-negative if a feature is known to have only a positive effect), or expand the network as needed. The encouraging performance on synthetic data is a first step, and ongoing work will focus on applying this approach to experimental datasets from PDLSC studies to test its real-world applicability.

CONCLUSION

We have demonstrated a streamlined Bayesian neural network pipeline that can predict PDLSC-driven cementum regeneration outcomes while also providing meaningful uncertainty bounds. This approach directly addresses the knowledge gap of "black-box" predictions in tissue engineering by quantifying confidence alongside the prediction. In doing so, it empowers researchers to make risk-aware decisions – for example, selecting a growth factor treatment that not only maximizes expected cementum thickness but also has a high certainty of success. The entire framework is lightweight and accessible, running on consumer-grade hardware and built with open-source tools, which facilitates immediate adoption by the community. Our BNN model serves as a decision-support tool in planning regenerative experiments by forecasting outcomes and their probabilities based on candidate biomaterials and conditions. Its Bayesian framework allows iterative learning through updates with new data, enhancing prediction accuracy and reducing uncertainty. Future plans include applying this to real PDLSC datasets for validation, extending to related tissues, and integrating into an active learning loop. This approach promotes a new paradigm of predictive, uncertainty-aware tissue engineering, combining computational models with experiments for faster discovery.

DECLARATIONS

Acknowledgments

We thank everyone who supported and contributed to this study.

Funding

This research did not receive any specific grant or financial support from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests

The authors have no competing interests to declare.

Ethical Approval

The study was approved by the appropriate ethics committee and conducted according to relevant guidelines and regulations.

Informed Consent

Not applicable.

REFERENCES

1. Li Q, Yang G, Li J, Ding M, Zhou N, Dong H, et al. Stem cell therapies for periodontal tissue regeneration: a network meta-analysis of preclinical studies. Stem Cell Res Ther. 2020 Oct;11(1):427.
2. Panda S, Sankari M, Satpathy A, Jayakumar D, Mozzati M, Mortellaro C, et al. Adjunctive Effect of Autologus Platelet-Rich Fibrin to Barrier Membrane in the Treatment of Periodontal Intrabony Defects. Journal of Craniofacial Surgery [Internet]. 2016;27(3). Available from: https://journals.lww.com/jcraniofacialsurgery/fulltext/2016/05000/adjunctive_effect_of_autologus_platelet_rich.32.aspx
3. Ramesh A, Varghese SS, Doraiswamy JN, Malaiappan S. Herbs as an antioxidant arsenal for periodontal diseases. J Intercult Ethnopharmacol. 2016;5(1):92–6.
4. Kaarthikeyan G, Jayakumar ND, Padmalatha O, Sheeja V, Sankari M, Anandan B. Analysis of the association between interleukin -1β (+3954) gene polymorphism and chronic periodontitis in a sample of the south Indian population. Indian Journal of Dental Research [Internet]. 2009;20(1). Available from: https://journals.lww.com/ijdr/fulltext/2009/20010/analysis_of_the_association_between_interleukin.9.aspx
5. Hosoya A, Shalehin N, Takebe H, Fujii S, Seki Y, Mizoguchi T, et al. Stem cell properties of Gli1-positive cells in the periodontal ligament. J Oral Biosci. 2020 Dec;62(4):299–305.
6. Calabrese EJ. Human periodontal ligament stem cells and hormesis: Enhancing cell renewal and cell differentiation. Pharmacol Res. 2021 Nov;173:105914.
7. Wen S, Zheng X, Yin W, Liu Y, Wang R, Zhao Y, et al. Dental stem cell dynamics in periodontal ligament regeneration: from mechanism to application. Stem Cell Res Ther. 2024 Oct;15(1):389.
8. Giovagnoli A. The Bayesian Design of Adaptive Clinical Trials. Int J Environ Res Public Health. 2021 Jan;18(2).
9. Liu Y, Zhang SY, Kleijn IT, Stumpf MPH. Approximate Bayesian computation for inferring Waddington landscapes from single-cell data. R Soc Open Sci. 2024 Jul;11(7):231697.
10. Qiao P. Simulation of Logistics Delay in Bayesian Network Control Based on Genetic EM Algorithm. Comput Intell Neurosci. 2022;2022:6981450.
11. Andrews J, Gorell S. Generating Missing Oilfield Data Using A Generative Adversarial Imputation Network GAIN [Internet]. SPE Western Regional Meeting. SPE; 2021. Available from: http://dx.doi.org/10.2118/200766-ms
12. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. Medical Decision Making [Internet]. 2006;26(6):565–74. Available from: http://dx.doi.org/10.1177/0272989x06295361
13. Rajkumar NMR, Muzoora MR, Thun S. Dentistry and Interoperability. J Dent Res [Internet]. 2022;101(11):1258–62. Available from: http://dx.doi.org/10.1177/00220345221100175
14. Selvaraj S. Bridging Healthcare Systems: Revolutionizing U.S. Public Healthcare through HL7 FHIR Interoperability and API Technology. International Journal of Science and Research (IJSR) [Internet]. 2024;13(6):684–90. Available from: http://dx.doi.org/10.21275/sr24606041602
15. Seidel S, Stadler T. TiDeTree: a Bayesian phylogenetic framework to estimate single-cell trees and population dynamic parameters from genetic lineage tracing data. Proc Biol Sci. 2022 Nov;289(1986):20221844.