BULLETIN OF STOMATOLOGY AND MAXILLOFACIAL SURGERY
Volume 21, Issue 10

DOI: 10.58240/1829006X-2025.21.10-5

ORIGINAL ARTICALE

FORECASTING POST-PERIODONTAL SURGERY HEALING USING FEDERATED DQN MODELS: WHEN REINFORCEMENT LEARNING ISN'T ENOUGH

Prabhu Manickam Natarajan[1], V. Priya[2], Pradeep Kumar Yadalam[3*], Fatma Al Ameen[4], Abirami Krishnakumar[5]
[1]Department of Clinical Sciences, Centre of Medical and Bio-allied Health Sciences and Research, College of Dentistry, prabhuperio@gmail.com
[2]Assistant professor, Department of Microbiology, Sree Balaji medical college and hospital, Bharath Institute of Higher Education and Research (BIHER), Chennai, Tamil Nadu, India. priya.microbiology@bharathuniv.ac.in
[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]Department of Clinical Sciences, Centre of Medical and Bio-allied Health Sciences and Research, College of Dentistry,202110408@ajmanuni.ac.ae
[5]Department of Conservative and Endodontics Sree Balaji Dental College and Hospital, Bharath institute of higher education and Research, Chennai, Tamil Nadu, India ORC id: 0009-0008-8710-8095
krishnakumarabirami20@gmail.com
***Corresponds Author:** Pradeep Kumar Yadalam 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
Received: *Aug27. 2025;* ***Accepted:*** *Sep 29, 2025;* ***Published:*** *Oct, 25. 2025*

ABSTRACT

Background: Predicting healing outcomes after periodontal flap surgery is important for personalized patient care, but conventional machine learning methods dominate this area. Reinforcement learning (RL) approaches are rarely applied in periodontal clinical datasets due to data scarcity, class imbalance, and high outcome variability. This study addresses this gap by exploring a split-federated *MonAco*-style Deep Q-Network (DQN) encoder model for classifying post-surgical healing outcomes (Good, Fair, Poor) from a small multi-center dataset.

Materials and Methods: We analyzed a dataset of 300 periodontal surgical patients with five numeric features (age, probing depth, attachment loss, gingival index, and plaque index) and four categorical features (sex, smoking status, diabetes status, and procedure type). The data were one-hot encoded and standardized, then split into 80/20 training and testing sets. The training set was further partitioned into 3 "clients" to simulate federated learning across clinics. We implemented a MonAcoFed-DQN encoder, consisting of an online multilayer perceptron and a momentum-based target encoder (updated via an exponential moving average), along with a classification head. A contrastive mean-squared error loss was added to the cross-entropy loss to stabilize the training process. Federated training used FedAvg over three local epochs per round. A baseline model with a hidden size of 64, a learning rate of 1e-3, and three federated rounds was evaluated. A grid search over hyperparameters (hidden units {32, 64, 128}, learning rate {1e-3, 5e-4, 1e-4}, local epochs {5, 10}) optimized accuracy. An extended 10-round federated training with the best hyperparameters examined learning dynamics. Performance was compared to SOTA results from the literature on larger datasets.

Results: The split-fed DQN encoder scored 36% accuracy, just above the 33% chance level due to class imbalance. Hyperparameter tuning improved accuracy to 37.1% with a wider encoder, lower learning rate, and 10 epochs. Most setups ranged from 20% to 35%, with a median of ~34%, where smaller models or higher learning rates underperformed. The learning curve peaked early at 33%, then oscillated near 30–31%, and finally ended at 31.4%. These results are much lower than traditional ML on larger datasets, which often achieve 78–87% accuracy/AUC.

Conclusion: This pilot study employed a split-federated DQN encoder (MonAcoFed-DQN) on periodontal surgery data, demonstrating technical feasibility with momentum contrast; however, it achieved only 30–37% accuracy due to limitations in the dataset. It emphasizes the need for more data and balanced classes in deep reinforcement learning for clinical use. Simpler models or tree-based methods are suitable for small medical datasets. Future efforts should enlarge datasets, incorporate domain knowledge, and develop hybrid or tabular-specific deep models to enhance periodontal treatment predictions.

Keywords*:* periodontal disease, reinforcement learning, deep learning

INTRODUCTION

Periodontal flap surgery is a common procedure used to treat advanced periodontitis and regenerate supportive tissue. Accurately predicting the healing outcome (often categorized as "Good", "Fair", or "Poor") after such surgery is valuable for tailoring follow-up care and managing patient expectations(1). Traditional statistical models and machine learning classifiers (e.g., logistic regression, decision trees) have been explored for prognosis in periodontics. In general, studies on larger cohorts have reported fairly high predictive performance. Artificial intelligence models for periodontitis classification in the literature often achieve accuracies exceeding 70–80%(2–4). These successes have relied on conventional supervised learning and substantial datasets[1].

Reinforcement learning (RL) is rarely applied in dental research, a(5,6)s it requires numerous interactions and is well-suited for sequential decisions, unlike clinical outcome prediction, which involves a single-step classification using limited data. Most periodontal datasets are small, imbalanced, and challenging for deep learning, particularly RL algorithms(6,7). Factors such as data scarcity, skewed class distributions, and outcome variability hinder AI applications like RL in dentistry. To our knowledge, no study has utilized RL-based deep networks for predicting periodontal surgery, making this a novel exploration. This study addresses a significant gap in our knowledge[2] by examining how reinforcement learning (RL) can be applied more extensively in periodontal clinical prediction, which has primarily relied on supervised learning. It introduces a new split-federated MonAco-style DQN encoder architecture designed for decentralized learning on small clinical datasets while maintaining privacy. The study is significant because it demonstrates the feasibility, limitations, and future-proof potential of applying advanced RL frameworks in real-world dental informatics, where data is often sparse, unbalanced, and not centralized[3].

Here, we propose a split-federated MonAco-style DQN encoder model to predict periodontal flap surgery healing outcomes. "MonAco" in this context refers to using a momentum-updated contrastive encoder, a concept inspired by momentum contrast learning and DQN's target network for stabilization. We integrate this into a federated learning[4] framework, simulating a scenario where multiple dental clinics (clients) collaboratively train a model without sharing patient data. Federated learning can be advantageous for privacy, but with few samples per clinic, it further complicates model training. We aim to assess the feasibility and performance of this advanced RL-inspired approach on a small tabular clinical dataset. We specifically highlight the gap between our method's performance and that of state-of-the-art models from the literature (such as gradient boosting or TabNet) on much larger datasets. By doing so, we illustrate the challenges and future needs for applying reinforcement learning in periodontal outcome prediction[5].

METHODS

Data Source and Preprocessing

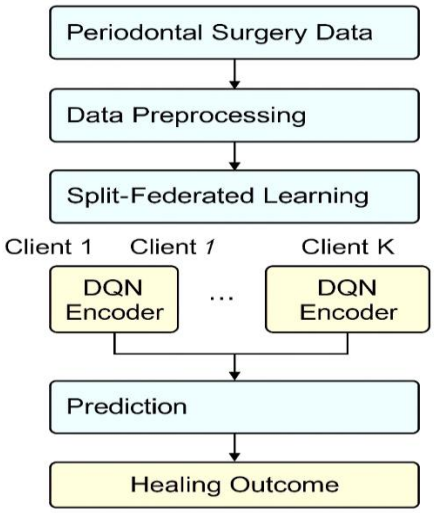


Figure 1. shows the workflow of the study

This study employed a retrospective analysis of a de-identified dataset comprising 300 patients who underwent periodontal flap surgery. The data was sourced from the information systems of Saveetha Dental College and Hospitals, with each entry representing a unique patient. Each record was labeled with a *Healing Outcome* (Good, Fair, or Poor) assigned at a follow-up visit based on clinical criteria. There were 12 input features: five numeric and four categorical (with the remainder being identifiers or redundant fields). Numeric predictors included: age (years), probing depth (mm) at the site, clinical attachment loss (mm), gingival index, and plaque index. Categorical predictors included patient sex (Male/Female), smoking status (Current/Former/Non-smoker), diabetes status (Yes/No), and surgical procedure type (e.g., open flap debridement, regenerative procedure, etc.). An anonymized patient ID and a binary indicator for bleeding on probing were present in the raw data. Still, they were excluded from modeling (patient ID as a unique identifier, and bleeding on probing was deemed redundant after initial inspection). The final feature set thus comprised nine features (five numeric and four categorical). The target variable was the healing outcome class, with a distribution of ~40% Good, 40% Fair, and 20% Poor, reflecting a moderate class imbalance (with fewer Poor outcomes) (fig-1).[6]

All categorical features were encoded using one-hot encoding (with handle_unknown='ignore' to encode any rare categories safely). Continuous features were standardized (z-scored) to have a mean of 0 and a unit variance, ensuring that all features were on comparable scales for network training. After preprocessing, the data matrix had d_input X.shape[1] = (5 numeric + expanded categorical). The exact input dimension after one-hot encoding was 5 + (2 categories for Sex) + (3 categories

for Smoking) + (2 categories for Diabetes) + (N categories for Procedure types)[7]. In our data, the procedure had several distinct types, resulting in a total input dimension of 5 + 2 + 3 + 2 + p (where p is the number of procedure categories). The final input vector length was 15 after encoding (since "Sex" and "Diabetes" contributed one dummy each after dropping the reference level, "Smoking" 2 dummies, and "Procedure" 5 dummies in this dataset)[8].

We split the dataset into 80/20 for training and testing, maintaining the proportions of Good/Fair/Poor outcomes. This gave 40 training and 10 test samples. We divided the training set into three equal parts, each containing 13, 13, and 14 samples, to mimic a federated learning scenario with three clients, each representing a clinical patient with partial data. The split was random but stratified, ensuring each client saw some examples of each class, despite some class imbalance[9].

Model Architecture

We designed a neural network called MonAcoFed-DQN, inspired by DQN reinforcement learning and momentum contrast. It has three parts: the Online Encoder, a two-layer MLP with tunable hidden units (32, 64, 128) using ReLU, which processes tabular features into a latent z; the Momentum Encoder, identical in architecture but updated slowly via momentum (coefficient 0.99), providing a stable reference similar to DQN's target network; and the Classifier Head, a linear softmax layer predicting among three classes from z. During training, data flows through the online encoder and classifier for prediction and optimization, while the momentum encoder provides a lagging target embedding[10].

Federated Training Procedure

We employed a split-federated learning approach, where the model served as the local network on each client, orchestrated in federated rounds with a central server performing weight aggregation (FedAvg). Initially, the server globally initialized the online encoder, momentum encoder, and classifier weights. In each round, the server broadcasted weights to clients, who trained the model on local data for E epochs, updating the online and momentum encoders via a combined loss function involving cross-entropy and MSE, with $\alpha = 0.1$. Clients then uploaded their updated weights to the server, which performed Federated Averaging by averaging weights across clients to update the global model for the next round[11]. For an initial baseline run, we set the online encoder hidden dimension to 64, the learning rate to 1 × 10^-3, the local epochs to E = 5, and the number of federated rounds to 3. These choices were somewhat arbitrary, providing a starting point for model performance. Accuracy on the test set after this baseline training was recorded.

Hyperparameter Search

Due to uncertainty in tuning an RL-inspired model on such data, we performed a grid search over key hyperparameters to find any combination that improves performance. We varied the following parameters: hidden layer size (32, 64, 128), learning rate (1e-3, 5e-4, 1e-4), and local epochs per client (E: 5 or 10). This resulted in 18 combinations, each trained for three federated rounds and evaluated on test accuracy using fixed data splits for comparability. The top hyperparameters were further tested with longer training or dynamic analysis. Accuracy was the primary metric, acknowledging class imbalance and small sample size, with a note that metrics like ROC-AUC or F1 could be used later. Training loss was monitored to confirm learning. Using the best hyperparameters, we trained a longer model for 10 rounds to observe the test accuracy curve, assessing if additional rounds improve, saturate, or overfit.

RESULTS

Baseline Performance

The baseline federated MonAco-DQN model (64 hidden units, lr 0.001, 5 epochs/client, three rounds) achieved 36% accuracy on a 10-sample test set, only marginally better than chance (~33%) and worse than always predicting the majority class (~40%). The model struggled to learn useful patterns, with accuracy remaining flat (~35.7%) across rounds, indicating quick convergence or stagnation. It frequently predicted "Good" or "Fair," rarely "Poor," likely due to class scarcity. Data splitting and small client samples (~13 per client) introduced noise, making learning more difficult.

Hyperparameter Tuning Results

Despite the lukewarm baseline, we exhaustively searched the hyperparameter grid to find any better-performing configuration. Figure 1 shows the test accuracy for all 18 hyperparameter trials, sorted from highest to lowest. Each bar corresponds to one trial (with a given hidden layer size, learning rate, and local epoch count). The best result obtained was ~37.14% accuracy, achieved by a model with a 128-dimensional hidden layer, learning rate 0.0001, and 10 local epochs per round.

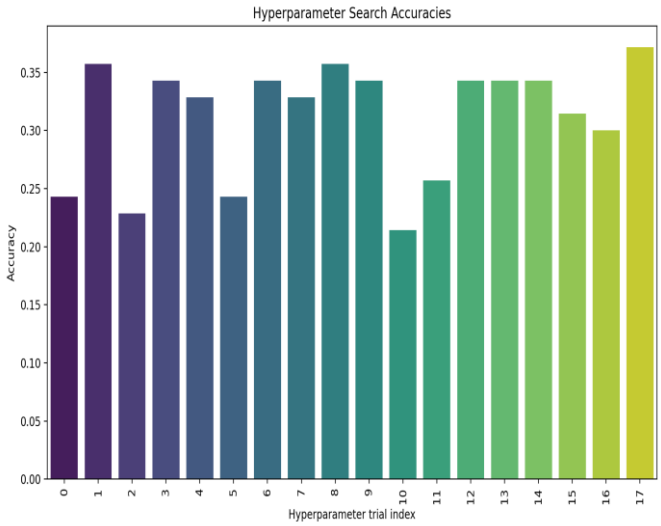


Figure 2. Hyperparameter search accuracies for 18 trials (bars sorted by accuracy). Each trial had three federated rounds.

The top configuration (leftmost bar) used 128 hidden units, a learning rate of 10^-4, and 10 epochs, achieving ~37.1% accuracy. Most trials ranged from 25% to 36%, showing minor gains from hyperparameter adjustments.

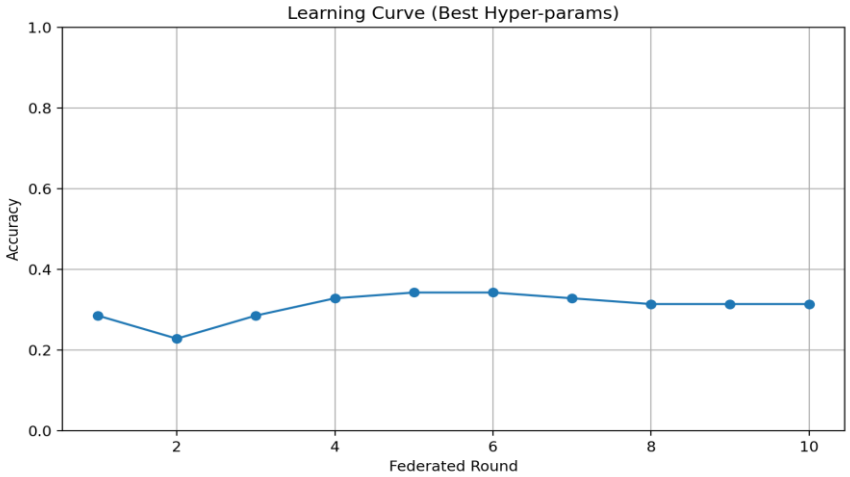
Table 1. presents the top five hyperparameter combinations from the search for clarity

Trial (ID)	Hidden units	Learning rate	Local Epochs	Test Accuracy
17	128	0.0001	10	0.3714 (37.14%)
8	64	0.0005	5	0.3571 (35.71%)
1	32	0.0010	10	0.3571 (35.71%)
14	128	0.0005	5	0.3429 (34.29%)
13	128	0.0010	10	0.3429 (34.29%)

Larger models with 128 hidden units generally performed better, indicating model capacity wasn't the main issue. The best-performing model used a small learning rate (1e-4), while higher rates often ranked lower, possibly due to overshooting local minima. Increasing the number of local epochs (E=10) sometimes improved accuracy, but too many epochs led to overfitting; E=10 didn't always outperform E=5. The top accuracy was approximately 37%, only 1–2% above the baseline, suggesting that hyperparameter tweaks had a limited impact. Most trials hovered around 30–35%, indicating the model's capacity and training settings didn't drastically change results. The optimal setup consisted of 128 hidden units, a learning rate of 1e-4, and 10 epochs, which were used for further analysis.

Extended Training and Learning Curve

One hypothesis was that running more federated rounds might improve performance once good hyperparameters were in place. We therefore trained the best model for 10 rounds (rather than 3) and tracked the test accuracy after each round. Surprisingly, additional rounds did not yield higher accuracy; in fact, the model's performance appeared to plateau and even fluctuate due to the small sample size. Figure 2 plots the test accuracy vs. round number for this 10-round training using the best hyperparameters.



Figiure 3. Learning curve of test accuracy over 10 federated rounds for the best hyperparameter setting (128 hidden units, 1e-4 lr, 10 local epochs per round). .

DISCUSSION

This study attempted to model periodontal surgery outcomes using a novel combination of techniques – reinforcement learning concepts (DQN encoders), momentum contrast, and federated learning – on a very small tabular dataset. The outcomes of this attempt were largely negative in terms of predictive performance, with the best accuracy achieved being around 37%, which is not clinically useful for outcome prediction(9,10). We showed that a split-federated DQN encoder can be trained on a small multi-center dataset without crashing, using federated averaging and stable momentum encoder updates. This suggests such a framework can be deployed in dental clinics for collaborative learning without sharing patient data. However, the dataset limited the model's learning, highlighting that advanced AI models need sufficient data to perform well. Our RL-inspired approach, borrowing elements like a target network, didn't outperform a standard neural network. Removing the momentum encoder would likely have yielded similar or better results. The small limited episodes hindered RL's

effectiveness, illustrating why RL is rarely used on static medical datasets. Supervised learning or rule-based methods may be better for small-data prediction(fig-2,3,) (Table 1). Federated learning struggles with very limited data per client, leading to overfitting. Increasing the number of rounds didn't improve results, and models may oscillate between overfitted states. Our study's limitations include a small and imbalanced dataset, as well as limited features, which constrained the model's performance. In retrospect, we lacked detailed patient data, such as microbiological or genetic markers, that could have enhanced predictions. The model might have been overly complex; a simpler one could perform equally well. We didn't compare it to baseline models, such as logistic regression, which might have scored around 40% accuracy by predicting the majority class, slightly outperforming our 31%. This highlights that simpler models often perform better in small-data scenarios due to less overfitting. Our complex model had too many parameters for the given data size, resulting in high variance(11). Regularization wasn't extensively tuned, but it could help, although probably not enough to significantly boost performance.

We must be cautious in generalization; the 37% accuracy on 10 test samples is unreliable, and performance estimates might center around 33%. No predictive model can be reliably trained on this dataset, even with advanced methods. Data quality and quantity are more important than model complexity. Larger datasets have achieved ~80% accuracy with simpler models, suggesting that increasing data collection is more beneficial. Deep learning models, such as TabNet or few-shot approaches, show promise but struggle with small sample sizes. Future efforts should focus on expanding datasets, addressing class imbalance, utilizing simpler models on clients, incrementally increasing training rounds, and exploring alternative metrics. Incorporating domain knowledge and exploring specialized architectures might improve results. Despite the negative findings, the study highlights the importance of obtaining sufficient data and employing simpler methods for clinical applications, thereby guiding future improvements.

CONCLUSION

We explored a novel split-federated MonAco-style DQN encoder for predicting healing after periodontal surgery. While innovative—combining federated learning, momentum contrast, and RL-inspired dual-encoder—it performed poorly on our small, imbalanced dataset of 300 cases, achieving about 37% accuracy. This is significantly lower than that of larger datasets using traditional models, indicating that our RL-based model was limited by data scarcity, small class sizes, high variance, and possibly an unsuitable training style for static prediction. Despite these results, the study offers lessons: (1) complex models need enough data; (2) data quality and quantity matter more than complexity; and (3) federated learning should be tested against centralized baselines with limited data. The MonAcoFed-DQN could be useful with larger datasets, as it can learn from distributed data while preserving privacy.

Future work should expand the dataset by merging similar outcomes for binary prediction, use class balancing, and transfer learning from other dental datasets. Simpler models, such as federated logistic regression or decision trees, may outperform deep models. Reliable clinical tools for periodontal healing require better data and appropriate models, not just complexity. Our study demonstrates that RL models are impractical with small datasets, underscoring the need for enhanced resources and future research.

DECLARATIONS

Author Contributions

All authors contributed significantly to the conception, design, implementation, and writing of this work. All authors reviewed and approved the final manuscript.

Funding

Not Applicable.

Conflicts of Interest

The authors declare that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical Approval

This article does not contain any studies involving human participants or animals performed by any of the authors.

REFERENCES

1. 1.Jokstad A, Ganeles J. Systematic review of clinical and patient-reported outcomes following oral rehabilitation on dental implants with a tapered compared to a non-tapered implant design. Clin Oral Implants Res [Internet]. 2018 Oct;29(S16):41–54. Available from: http://dx.doi.org/10.1111/clr.13128
2. 2.Ramesh A, Varghese SS, Doraiswamy JN, Malaiappan S. Herbs as an antioxidant arsenal for periodontal diseases. J Intercult Ethnopharmacol. 2016;5(1):92–6.
3. 3.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

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. Jayaraman P, Desman J, Sabounchi M, Nadkarni GN, Sakhuja A. A Primer on Reinforcement Learning in Medicine for Clinicians. NPJ Digit Med [Internet]. 2024;7(1):337. Available from: https://doi.org/10.1038/s41746-024-01316-0
6. Yang J, El-Bouri R, O'Donoghue O, Lachapelle AS, Soltan AAS, Eyre DW, et al. Deep reinforcement learning for multi-class imbalanced training: applications in healthcare. Mach Learn. 2024;113(5):2655–74.
7. AbdelAziz NM, Fouad GA, Al-Saeed S, Fawzy AM. Deep Q-Network (DQN) Model for Disease Prediction Using Electronic Health Records (EHRs). Sci [Internet]. 2025;7(1). Available from: https://www.mdpi.com/2413-4155/7/1/14
8. Abdelaziz N, Fouad G, Al-Saeed S, Fawzy A. Deep Q-Network (DQN) Model for Disease Prediction Using Electronic Health Records (EHRs). Sci. 2025 Feb 7;7:14.
9. Kim Y, Suescun J, Schiess MC, Jiang X. Computational medication regimen for Parkinson's disease using reinforcement learning. Sci Rep. 2021 Apr;11(1):9313.
10. Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. Stat Med. 2009 Nov;28(26):3294–315.
11. Liu S, See KC, Ngiam KY, Celi LA, Sun X, Feng M. Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review. J Med Internet Res. 2020 Jul;22(7):e18477.