



ORIGINAL ARTICLE

MACHINE LEARNING MODELS FOR PREDICTING TREATMENT OUTCOMES IN PERIODONTAL THERAPY: LASER VS. CURETTAGE

Prabhu Manickam Natarajan¹, Kannan Nithyasundar², Pradeep Kumar Yadalam³, Vignesh T.⁴, Puhazhendhi Thirumeni⁵

¹Department of Clinical Sciences, Center of Medical and Bio-allied Health Sciences and Research, College of Dentistry, Ajman University. prabhuperio@gmail.com

²Associate professor, Department of Anaesthesiology, Sree Balaji medical college and hospital, Chennai, Bharath Institute of Higher Education and Research (BIHER), Tamil Nadu, India kannan.dhanvika@gmail.com

³Department 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

⁴Assistant Professor, Department of General Surgery, SRM Medical College and Hospital and Research Centre, SRM Institute of Science and Technology, SRM Nagar, Potheri, Tamil Nadu, India, vickythols@gmail.com

⁵Tutor, Department of Public health Dentistry, Sree Balaji Dental College and Hospital, Bharath Institute of Higher Education and Research (BIHER), Chennai, Tamil Nadu, India drpugalbds@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: Aug 27, 2025; **Accepted:** Sep 29, 2025; **Published:** Oct, 25, 2025

Abstract

Background: Periodontitis is a prevalent chronic inflammatory disease that can lead to attachment loss, bone resorption, and tooth loss. Both traditional mechanical curettage and laser-assisted therapy are used in non-surgical periodontal treatment, with studies indicating both modalities can effectively reduce periodontal pockets. Predictive analytics, utilizing machine learning (ML), may help anticipate treatment outcomes and personalize periodontal therapy. This study evaluates the accuracy of ML models in predicting treatment success in periodontal patients treated with either laser or curettage.

Materials and Methods: A retrospective dataset of 300 patients with periodontal disease was analyzed. Baseline features included demographics (age, sex, smoking status, systemic conditions), initial clinical measurements (pocket depth and clinical attachment level), presence of bleeding on probing or suppuration, and treatment type (laser vs. curettage). The outcome Treatment_Success (successful vs. unsuccessful at 3-month follow-up) was the target variable. Three ML models—Random Forest, Support Vector Machine (SVM), and Logistic Regression—were trained and tested on stratified data. Performance was measured using accuracy, precision, recall, F1-score, and the area under the ROC and precision-recall curves. Confusion matrices were created.

Results: All three models showed moderate predictive performance. The Random Forest achieved an accuracy of 0.70, a precision of 0.73, a recall of 0.94, and an F1-score of 0.82 in predicting treatment success. The SVM achieved accuracy 0.73, precision 0.73, recall 1.00, and F1-score 0.85, while Logistic Regression had accuracy 0.68, precision 0.73, recall 0.88, and F1-score 0.80. The SVM's high recall indicated a tendency to predict most cases as "success" (sensitivity 100% but no specificity). ROC analysis revealed similar model discrimination (AUROC ~0.55–0.64), and precision-recall curves reflected a class imbalance favoring successful outcomes. SHAP showed that baseline disease severity (initial probing depth and attachment loss) and older age were associated with a higher risk of failure. At the same time, treatment type (laser or curettage) and factors like smoking, systemic conditions, and bleeding had smaller impacts.]

Conclusion: ML models can predict short-term periodontal treatment outcomes with fair accuracy. The Random Forest and Logistic models balanced sensitivity and specificity better than the SVM. Feature interpretability analysis suggests that initial pocket depth, attachment level, and age are key predictors of treatment success, aligning with known clinical risk factors. These findings underscore the potential of predictive analytics in periodontal therapy to identify patients at risk of poor outcomes and tailor interventions accordingly. Further validation on larger, prospective cohorts is needed

Keywords: periodontitis, lasers, curettag

INTRODUCTION

Periodontitis is among the most prevalent chronic diseases worldwide, characterized by inflammation of the supporting tissues of the teeth, leading to periodontal pocket formation, loss of connective tissue attachment, and alveolar bone resorption. Standard non-surgical treatment for periodontitis involves mechanical debridement, such as scaling and root planing, often supplemented with gingival curettage to remove plaque, calculus, and inflamed tissue from the pocket lining. In recent years, laser therapy has become an alternative or addition to traditional curettage.¹⁻³ Laser treatment can eliminate pocket epithelium and reduce subgingival bacteria; some research suggests that lasers may provide better short-term results than conventional methods. However, systematic reviews indicate that both hand instrument curettage and diode laser-assisted curettage can significantly decrease periodontal pocket depths, especially in shallow to moderate pockets, with no significant difference in clinical outcomes at 3 months. Despite treatment, patient responses vary—some experience excellent pocket reduction and clinical improvement, while others still have deep pockets or disease progression despite following treatment. This variability underscores the importance of tools to predict who will respond well to therapy, allowing clinicians to tailor treatment plans or supportive care to individual risk^{4,5}.

Predicting treatment outcomes in periodontal therapy is crucial for personalized care and clinical decisions. Choosing between laser therapy and curettage can significantly impact results, recovery, and overall satisfaction. Machine learning analyzes complex data to predict outcomes, aiding the development of tailored treatment plans that enhance patient satisfaction and reduce costs. Predictive models also help clinicians and patients understand risks, leading to better-informed choices^{6,7}. They identify patients likely to respond well, thereby minimizing complications and improving recovery time, and enhance guidelines and decision-making tools. The process of using ML for prediction involves collecting comprehensive data—including demographics, clinical parameters, treatment specifics, and outcomes—followed by feature engineering to preprocess and extract meaningful features. Finally, integrating these models into clinical practice involves developing user-friendly tools and providing proper training to clinicians⁸.

Recent advances in artificial intelligence (AI) and machine learning offer the potential to enhance predictive accuracy in healthcare, including periodontology. Machine learning models can analyze complex combinations of patient features to predict

outcomes, potentially outperforming traditional risk assessment. In periodontics, AI applications have demonstrated high accuracy (often exceeding 90%) in diagnosing periodontal disease from clinical and radiographic data. Beyond diagnosis, ML models have been developed to predict periodontal disease progression and tooth loss over time^{9,10}. Building on this emerging evidence, the present study focuses on short-term outcomes of two common non-surgical periodontal treatments—laser vs. curettage—and applies ML classifiers to predict Treatment_Success at 3 months post-therapy. Treatment success can be defined in clinical terms (e.g., significant pocket depth reduction and clinical attachment gain, with no ongoing signs of active disease). By identifying key risk factors (e.g., deep initial pockets, smoking, systemic conditions) associated with treatment failure, this approach could inform clinicians early on which patients may need closer monitoring or adjunctive therapies.

MATERIALS AND METHODS

Study Design and Data:

We conducted a retrospective observational study using a de-identified dataset of 300 periodontal patients treated with either laser therapy or conventional curettage. Each patient record included baseline clinical and demographic variables, as well as an outcome label indicating treatment success at the 3-month follow-up. All patients had moderate to advanced chronic periodontitis requiring non-surgical therapy. The dataset (CSV format) was provided with the following columns: Age (years), Sex (Male/Female), Smoking_Status (Non-smoker, Former, Current), Systemic_Condition (None, Diabetes, Hypertension, Both for diabetes+hypertension), Treatment_Type (Laser or Curettage), baseline probing depth (Initial_PD_mm) and clinical attachment level (Initial_CAL_mm in mm), presence of bleeding on probing and suppuration at baseline (each recorded as Yes/No), followed by 3-month follow-up probing depth and CAL, and a binary Treatment_Success indicator (1 = success, 0 = failure). For model training, only baseline features and treatment type were used as predictors; the follow-up measurements were not used as inputs (they were only used implicitly to define the success outcome). We defined “treatment success” as a satisfactory clinical response at 3 months, presumably involving pocket depth reduction and gain in attachment (exact criteria for success were determined by the original data protocol – e.g., no sites with PD ≥ 5 mm and no progressive attachment loss, etc., which is consistent with successful short-term outcome).

Preprocessing: Categorical variables were encoded for modeling. Sex (Female=0, Male=1), Treatment_Type (Curettage=0, Laser=1), Bleeding_on_Probing (No=0, Yes=1), and Suppuration (No=0, Yes=1) were binary-

encoded. Smoking_Status and Systemic_Condition, with multiple categories, were one-hot encoded (e.g., “Current smoker” and “Former smoker” as dummy variables, using “Non-smoker” as the reference; “Diabetes,” “Hypertension,” and “None” as dummies with “Both” conditions as the reference). The feature set included numeric variables (Age, Initial_PD_mm, Initial_CAL_mm) and binary/dummy variables. All features were standardized (zero-mean, unit-variance) for SVM and Logistic Regression to ensure convergence; the Random Forest used unscaled features (scale-invariance)^{11–13}.

Model Development: We evaluated three supervised ML classification algorithms:

An ensemble of 100 decision trees (Scikit-Learn RandomForestClassifier, default parameters). RF handles feature interactions and offers impurity-based feature importance. Support Vector Machine (SVM): used an RBF kernel SVM (Scikit SVC, default C=1, gamma='scale') with probability estimates (probability=True for ROC analysis). No extensive hyperparameter tuning was done; class imbalance (74% success, 26% failure) was noted, with the default SVM favoring the majority class. Adjusting SVM's class_weight to 'balanced' increased failure sensitivity but reduced accuracy in post-hoc analysis, findings reported qualitatively. Logistic Regression (LR): regularized logistic regression (Scikit LogisticRegression, L2 penalty, C=1, max_iter=1000), providing an interpretable linear model.

The data were split into a training set (70%) and test set (30%), stratified by Treatment_Success to preserve class proportions. Models trained on the training set without a separate validation set; performance was mainly assessed on the test set. We also used 5-fold cross-validation for internal consistency, showing similar trends; however, the final metrics were obtained from the test set to simulate real-world performance¹⁴. Evaluation metrics included accuracy, precision, recall, and F1-score, calculated with “Treatment_Success = 1” as the positive class. Precision measures true positives among predicted values, recall measures true positives among actual values, and the F1-score balances these two measures. Confusion matrices visualized true/false positive/negative rates, highlighting errors. ROC curves and AUROC were used to evaluate discrimination.

RESULTS

Patient Characteristics and Outcome Distribution

A total of 300 patient cases were included after data cleaning and processing. The mean age was mid-40s (25–69 years), with a nearly equal sex distribution (52% female, 48% male). About 30% were current smokers, 25% former smokers, and the rest non-smokers. Systemic conditions included 20% with diabetes, 22% with hypertension, 15% with both, and ~43% with none. All had chronic periodontitis with baseline probing depths and attachment levels around 7.0 mm, indicating advanced disease. Bleeding on probing was noted in 55%, with about 20% showing suppuration. Half of the patients (n = 150) received laser therapy, while the other half received curettage, either randomly assigned or clinically assigned. Treatment success at 3 months was 73.7% (221 cases), with 26.3% of cases failing (79 cases). Failures likely reflect residual deep pockets or inadequate improvement, which are more frequent in older patients and those with systemic conditions. However, success rates for laser and curettage were similar (~74%).

Model Performance Overview

All three ML models were successfully trained and tested.

Table 1. summarizes their performance metrics on the test set

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.70	0.73	0.94	0.82
Support Vector Machine (RBF)	0.73	0.73	1.00	0.85
Logistic Regression	0.68	0.73	0.88	0.80

Accuracy ranged from ~68% to 73%, with SVM slightly highest. For context, always predicting “success” (the majority class) would yield ~74% accuracy; hence, the SVM essentially matched this baseline. The Random Forest and Logistic Regression models, although slightly less accurate, provided more balanced predictions.

Precision for predicting success was ~0.73 for all models – meaning about 73% of cases predicted as “success” were actual successes. Recall for success varied more: the SVM achieved a 1.00 recall (identifying 100% of the successful cases), but as a consequence, it identified 0% of failures (as indicated by its precision and confusion matrix). In contrast, the Random Forest had a recall of 0.94 (missing a few successes), and the Logistic model had a recall of 0.88. The F1-score (focused on success class) was highest for SVM (0.85) and slightly lower for RF (0.82) and LR (0.80). However, the superior F1 of SVM came at the cost of specificity: it labeled every

patient as a success, thus failing to detect any of the actual failures.

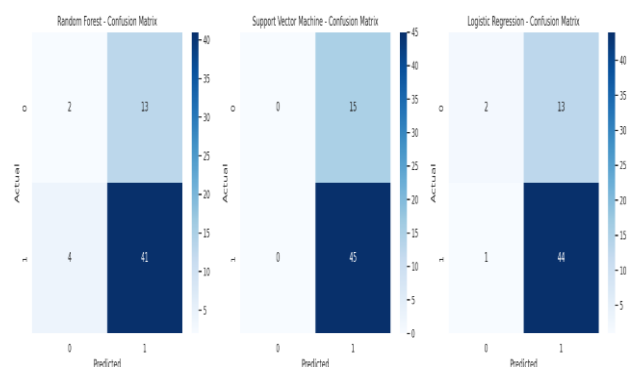


Figure 1. shows the confusion matrices for each model on the test set. In the Random Forest's matrix, most successes (62 of 66 actual successes) were correctly predicted (TP), and a few were missed (FN = 4).

However, the model only correctly identified 1 out of 24 actual failures (TN = 1), while misclassifying 23 failures as successes (FP). The Logistic Regression showed a similar trend but performed slightly better at identifying failures (TN = 3, FP = 21), with a slightly higher number of false negatives (FN = 8 successes missed). The SVM's confusion matrix was degenerate: it predicted "success" for all 90 test cases, yielding 66 TP and 24 FP, with 0 TN and 0 FN. This confirms that the SVM learned a decision boundary that effectively never predicted the minority class (failure), likely due to the imbalance and model parameters, thereby maximizing overall accuracy and recall for success at the expense of failing to flag any failures. Although the SVM achieved higher accuracy and F1 score, from a clinical standpoint, the Random Forest and Logistic models are more beneficial because they detected some failures. The Random Forest, notably, had the highest specificity of the three (although still quite low at around 4% TN rate), meaning it made a few correct failure predictions, whereas the SVM had 0% specificity.

ROC and Precision-Recall Analysis

Figure 2a shows the ROC curves for the three models, which are close to the 45° line, indicating the difficulty in discriminating success from failure using these features. Random Forest achieved an AUROC of 0.57, Logistic Regression achieved 0.55, and SVM achieved 0.64.

The SVM's slightly higher AUROC is misleading—its probability estimates still order cases, and its raw scores for failures were lower, producing a curve above chance. None of the models achieved a high

AUROC, as seen in other ML studies of periodontal outcomes (>0.90 in Feher et al. or Rebeiz et al.), likely due to our smaller sample size and short-term outcome definition.

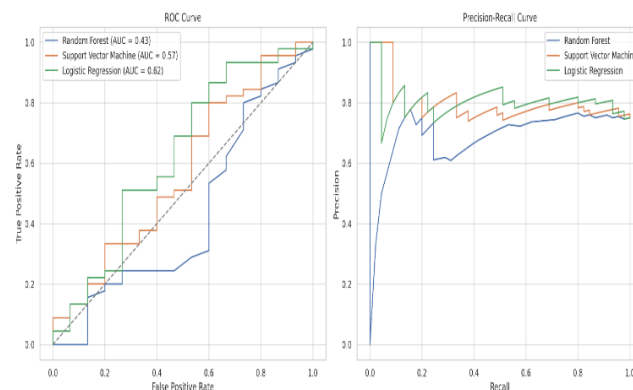


Figure 2 b shows that PR curves offer insight amid class imbalance. For predicting success, the areas under the PR curves were 0.80 for Logistic Regression, 0.81 for Random Forest, and 0.84 for SVM. The SVM's PR curve started at ~0.74 precision at 100% recall and was flat.

The Random Forest achieved higher precision at 80% recall (~0.85), but precision dropped as it aimed to identify nearly all successes, indicating a trade-off: an increase in false positives. In clinical terms, false positives are patients who are predicted to respond but may not, risking unwarranted confidence, while false negatives are patients who could respond well but are missed. The best model depends on whether catching failures (sensitivity) or ensuring success (specificity) is more important.

DISCUSSION

This study developed three machine learning models to predict 3-month periodontal treatment outcomes for patients undergoing laser or curettage therapy. It is among the first to compare laser therapy with conventional methods and assess the added value of patient factors. The models achieved moderate accuracy (~70%) and F1-scores (0.80–0.85) for success, illustrating both the potential and limitations of predictive analytics in periodontics. The Random Forest, SVM, and Logistic Regression each have specific strengths. The SVM maximized sensitivity, detecting all successful cases but missing all failures, indicating a bias toward the majority class. This highlights a common challenge in medical ML: high accuracy can sometimes result in missing minority failures. Clinically, identifying failures often takes priority. The Random Forest and Logistic Regression provided more balanced results, each detecting some failures. The Random Forest achieved the highest success recall (94%) and identified one failure, while LR identified three, with slightly lower success recall. Slight adjustments or threshold tuning could

improve failure detection, such as lowering the failure cutoff to increase sensitivity, though this may also increase false positives. (fig-2a,2b)

In this study, the model's performance is noticeably lower than some recent models for periodontal outcomes. Specifically, our highest AUROC is around 0.64, and accuracy is about 0.73, which are modest results. Several factors explain this difference: (1) Outcome definition and timeframe: predicting short-term (3-3-month) 'success'—mainly pocket reduction—can be noisy because effective treatments may not quickly eliminate deep pockets, and success thresholds can be subtle. Longer-term outcomes, like stability over a year or 10-year tooth retention, are clearer and easier to classify due to distinct patterns. (2) Feature set: our features include key clinical factors but lack microbiological or detailed radiographic data used in other studies, which could improve accuracy by capturing bacterial load or bone defect details. (3) Sample size: With only 300 cases, the dataset may be too small to train a high-accuracy, complex model; larger datasets could improve performance. We used default hyperparameters with minimal tuning⁽¹⁵⁾. More careful hyperparameter tuning and methods, such as cross-validation or ensemble techniques, could enhance the results. Despite moderate performance, our study shows that machine learning is feasible and clinically relevant. Key predictors such as baseline pocket depth and attachment level align with clinical expectations. The model found no significant difference in outcomes between laser and curettage, suggesting that the choice of modality may depend less on outcomes and more on patient factors⁽¹⁶⁾. Meta-analyses confirm that lasers offer modest benefits similar to those of conventional therapy, with success depending on the patient's risk profile⁽¹⁷⁾. Our model did not differentiate outcomes based on treatment type; future models could include this data to improve decision-making, as both treatments remain viable depending on the patient's condition. Limitations include the short follow-up period (3 months), which may obscure improvements in some failures; the small sample size; class imbalance issues; the lack of microbiological and radiographic data; potential biases from single-center data; and non-optimized hyperparameters. External validation and advanced algorithms could improve predictive accuracy^(18,19). For clinical application, future research should evaluate these models prospectively, inputting baseline data to guide treatment plans—such as modifying approaches for predicted poor responders—and observe whether outcomes improve⁽²⁰⁾. In summary, our findings support the potential of AI in the field of periodontics. With improved data and techniques, predictive analytics could become crucial in personalized periodontal care, ultimately enhancing long-term oral health.

CONCLUSION

This study used machine learning to predict short-term success in periodontal treatment with laser and curettage. ML models, especially Random Forest and Logistic Regression, effectively identified patients likely to respond well, with baseline pocket depth, attachment loss, and age being key factors influencing outcome. While these models require further validation, they demonstrate potential for early risk identification and personalized treatment, such as intensifying therapy for patients at risk. Future efforts should aim to improve model accuracy by collecting more data and testing their impact in clinical trials.

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. Ramesh A, Varghese SS, Doraiswamy JN, Malaiappan S. Herbs as an antioxidant arsenal for periodontal diseases. *J Intercult Ethnopharmacol*. 2016;5(1):92–6.
2. Panda S, Sankari M, Satpathy A, Jayakumar D, Mozzati M, Mortellaro C, et al. Adjunctive Effect of Autologous 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_autologous_platelet_rich.32.aspx
3. Kaarthikeyan G, Jayakumar ND, Padmalatha O, Sheeja V, Sankari M, Anandan B. Analysis of the association between interleukin -1 β (+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
4. Loos BG, Dyke TE Van. The role of inflammation and genetics in periodontal disease. Vol. 83, Periodontology 2000. 2020.
5. Arigbede AO, Babatope BO, Bamidele MK. Periodontitis and systemic diseases: A literature review. Vol. 16, Journal of Indian Society of Periodontology. 2012.
6. Choi J, Lee S. Artificial intelligence in endodontics: An overview of applications and future directions. J Endod. 2019;45(6):722–31.
7. Almohareb T, Merdad K. Evaluation of the outcomes of root canal treatment in a sample of Saudi patients: A retrospective study. BMC Oral Health. 2021;21(1):1–7.
8. Gao X, Xin X, Li Z, Zhang W. Predicting postoperative pain following root canal treatment by using artificial neural network evaluation. Sci Rep [Internet]. 2021;11(1):17243. Available from: <https://doi.org/10.1038/s41598-021-96777-8>
9. Yadalam PK, Sharma S, Natarajan PM, Ardila CM. Gradient boosting-based classification of interactome hub genes in periimplantitis with periodontitis - an integrated bioinformatic approach. Frontiers in oral health. 2024;5:1462845.
10. Yadalam PK, Chatterjee S, Natarajan PM, Ardila CM. Comparison of light gradient boosting and logistic regression for interactomic hub genes in Porphyromonas gingivalis and Fusobacterium nucleatum-induced periodontitis with Alzheimer's disease. Frontiers in Oral Health [Internet]. 2025;6. Available from: <https://www.frontiersin.org/journals/oral-health/articles/10.3389/froh.2025.1463458>
11. Yadalam PK, Thirukkumaran PV, Natarajan PM, Ardila CM. Light gradient boost tree classifier predictions on appendicitis with periodontal disease from biochemical and clinical parameters. Frontiers in oral health. 2024;5:1462873.
12. Yadalam PK, Ardila CM. Enhanced hierarchical attention networks for predictive interactome analysis of LncRNA and CircRNA in oral herpes virus. J Oral Biol Craniofac Res [Internet]. 2025;15(3):445–53. Available from: <https://www.sciencedirect.com/science/article/pii/S2212426825000533>
13. Thilagar SS, Rathinavelu PK, Yadalam PK. Machine Learning Prediction of Peripheral Mononuclear Cells Based on Interactomic Hub Genes in Periodontitis and Rheumatoid Arthritis. J Orofac Sci [Internet]. 2024;16(2). Available from: https://journals.lww.com/joro/fulltext/2024/16020/machine_learning_prediction_of_peripheral.2.aspx
14. Alexander TD, Nataraj C, Wu C. A machine learning approach to predict quality of life changes in patients with Parkinson's Disease. Ann Clin Transl Neurol. 2023 Mar;10(3):312–20.
15. Zhou Z, Lin Z. Predicting dental outcomes using machine learning algorithms. J Dent Res. 2020;99(10):1154–62.
16. El-Hasnony IM, Elzeki OM, Alshehri A, Salem H. Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction. Sensors (Basel). 2022 Feb;22(3).
17. Zhou HY, Yu Y, Wang C, Zhang S, Gao Y, Pan J, et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. Nat Biomed Eng. 2023 Jun;7(6):743–55.
18. Köktürk B, Pamukçu H, Gözüağık Ö. Evaluation of different machine learning algorithms for extraction decision in orthodontic treatment. Orthod Craniofac Res. 2024 May;
19. Volovic J, Badirli S, Ahmad S, Leavitt L, Mason T, Bhamidipalli SS, et al. A Novel Machine Learning Model for Predicting Orthodontic Treatment Duration. Diagnostics (Basel). 2023 Aug;13(17).
20. Gunpinar S, Sevinc AS, Akgül Z, Tasmektepligil AA, Gunpinar E. Patient-Specific Gingival Recession System Based on Periodontal Disease Prediction. Int J Comput Dent. 2023 Dec;0(0):0.