



www.bioinformation.net
Volume 21(10)



Research Article

Received October 1, 2025; Revised October 31, 2025; Accepted October 31, 2025, Published October 31, 2025

DOI: 10.6026/973206300213866

SJIF 2025 (Scientific Journal Impact Factor for 2025) = 8.478

2022 Impact Factor (2023 Clarivate Inc. release) is 1.9

Declaration on Publication Ethics:

The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at <https://publicationethics.org/>. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Declaration on official E-mail:

The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

License statement:

This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Comments from readers:

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Disclaimer:

Bioinformation provides a platform for scholarly communication of data and information to create knowledge in the Biological/Biomedical domain after adequate peer/editorial reviews and editing entertaining revisions where required. The views and opinions expressed are those of the author(s) and do not reflect the views or opinions of Bioinformation and (or) its publisher Biomedical Informatics. Biomedical Informatics remains neutral and allows authors to specify their address and affiliation details including territory where required.

Edited by Hiroj Bagde

E-mail: hirojbagde8@gmail.com

Citation: Thakkar *et al.* Bioinformation 21(10): 3866-3870 (2025)

Machine learning for predicting antimicrobial efficacy of periodontal gel formulations *in vitro* biofilm models

Rohitkumar R Thakkar^{1,*}, Nirma Yadav², Anand Kumar³, Shilpa Duseja⁴, Sunny Mavi⁵ & Udipta Sahoo⁶

¹Department of Periodontology, Siddhpur Dental College and Hospital, Siddhpur – 384151, Gujarat, India; ²Department of Dentistry, Maa Vindhyawasini Autonomous State Medical College, Mirzapur, Uttar Pradesh, India; ³Department of Dentistry, Moti Lal Nehru Medical College, Prayagraj, Uttar Pradesh, India; ⁴Department of Periodontology, Narsinhbhai Patel Dental College and Hospital, Sankalchand Patel University, Visnagar, India; ⁵Department of Periodontics, Sudha Rustagi College of Dental Sciences and Research, Faridabad, Haryana, India; ⁶Department of Periodontics, Kalinga Institute of Dental Sciences, KIIT (Deemed to be university), Patia, Bhubaneswar, Odisha, India; *Corresponding author

Affiliation URL:

<https://www.sgdher.org/>
<https://mvasmc.edu.in/>
<https://mlnmc.edu.in/>
<https://npdch.edu.in/>
<https://srdentalcollege.com/>
<https://kids.kiit.ac.in/>

Author contacts:

Rohitkumar R Thakkar - E- mail: drrohitperio@gmail.com
Nirma Yadav - E- mail: nirmay90@gmail.com
Anand Kumar - E- mail: anandkmr901@gmail.com
Shilpa Duseja - E- mail: budhirajashilpa@gmail.com
Sunny Mavi - E- mail: snnymavi@gmail.com
Udipta Sahoo - E- mail: sahooudipta2000@gmail.com

Abstract:

Periodontal disease caused by dysbiotic biofilms poses a major challenge and predicting the efficacy of topical antimicrobial gels is limited by biofilm resistance and resource-intensive in vitro testing. Therefore, it is of interest to develop machine learning (ML) models to predict antimicrobial efficacy of novel gel formulations against multi-species periodontal biofilms. Hence, a total of 120 formulations with varying polymers, agents, concentrations and enhancers were tested using the Calgary Biofilm Device and efficacy data were used to train Random Forest, SVM, Gradient Boosting and Neural Network models. Gradient Boosting achieved the best performance (accuracy 92.8%, AUC-ROC 0.96), with antimicrobial type, concentration and polymer viscosity as key predictors. ML, particularly Gradient Boosting, offers a reliable tool for predicting periodontal gel efficacy, enabling faster formulation optimization and reducing the need for extensive laboratory screening.

Keywords: machine learning, periodontal gel, antimicrobial efficacy, biofilm, prediction model, drug formulation, chlorhexidine, chitosan.

Background:

Periodontitis, a chronic inflammatory disease affecting nearly 50% of adults globally, is primarily initiated and sustained by pathogenic biofilms accumulating on tooth surfaces and within periodontal pockets [1]. These biofilms, characterized by complex polymicrobial communities encased in an extracellular polymeric substance (EPS), exhibit significantly enhanced resistance to antimicrobial agents compared to their planktonic counterparts [2]. Mechanical debridement (scaling and root planing) remains the cornerstone of treatment; however, its limitations in accessing deep pockets and complex root anatomy necessitate adjunctive antimicrobial therapies [3]. Topical antimicrobial gels offer a promising strategy by delivering high local concentrations of agents directly to the periodontal pocket, minimizing systemic side effects and improving patient compliance [4]. Despite their potential, the clinical efficacy of existing topical antimicrobials is often variable and suboptimal. This variability stems from multiple factors, including the inherent resistance of biofilms, limited penetration of antimicrobials through the EPS and into deeper layers of the biofilm and the physicochemical properties of the gel vehicle itself influencing drug release kinetics and substantivity [5]. Developing novel, highly effective gel formulations requires extensive and costly iterative in vitro and in vivo testing against representative biofilm models. Traditional in vitro biofilm assays, while essential, are laborious, time-consuming (often taking 5-7 days per formulation) and resource-intensive, creating

a significant bottleneck in the drug development pipeline [6]. Recent advances in computational modeling and artificial intelligence present transformative opportunities for pharmaceutical research. Machine learning (ML) algorithms excel at identifying complex, non-linear relationships within high-dimensional datasets, making them ideally suited for predicting biological outcomes based on compound or formulation characteristics [7]. ML has been successfully applied in various biomedical contexts, including predicting drug-target interactions, optimizing drug delivery systems and even identifying novel antimicrobial peptides [8]. Within dentistry, ML models have shown promise in diagnosing periodontal disease from radiographs and predicting treatment outcomes [9]. However, the application of ML specifically for predicting the *antimicrobial efficacy* of topical periodontal formulations against complex, multi-species biofilms remains largely unexplored [10]. A critical research gap exists in leveraging computational approaches to rapidly screen and optimize periodontal gel formulations. Current development relies heavily on empirical testing, lacking efficient predictive tools that could prioritize the most promising candidates for experimental validation. Furthermore, understanding the relative contribution of different formulation components (polymer type, antimicrobial agent, concentration, enhancers) to overall biofilm efficacy through data-driven modeling is lacking [11]. Therefore, it is of interest to develop, train and validate machine learning models capable of accurately predicting the antimicrobial efficacy of

diverse periodontal gel formulations against a validated multi-species *in vitro* biofilm model, based solely on their physicochemical and compositional properties.

Materials and Methods:

Study design:

An experimental *in vitro* study combined with computational modeling.

Sample size:

120 unique gel formulations were synthesized. Each formulation was tested in quintuplicate (n=5 replicates) against biofilms, yielding 600 data points.

Formulation design and inclusion criteria:

- [1] **Polymer Base:** Carbopol 940 (0.5-2.0% w/v), Chitosan (low, medium, high molecular weight; 1.0-2.5% w/v), Poloxamer 407 (18-25% w/v).
- [2] **Antimicrobial Agent:** Chlorhexidine digluconate (CHX; 0.1-2.0% w/v), Doxycycline hyclate (DOX; 0.5-2.0% w/v), Cetylpyridinium chloride (CPC; 0.05-0.2% w/v).
- [3] **Penetration Enhancer:** None, Disodium EDTA (0.5-1.0% w/v), Citric Acid (0.5-1.0% w/v).
- [4] **Other Components:** Glycerol (humectant, 5-10% w/v), Methylparaben (preservative, 0.1% w/v), Deionized water (q.s. to 100%).
- [5] **Inclusion:** Formulations with pH 5.5-7.5, viscosity 5,000-50,000 cP (measured by Brookfield DV2T viscometer) and acceptable physical stability (no phase separation/syneresis after 1 week at 4°C, 25°C, 40°C).
- [6] **Exclusion:** Formulations exhibiting precipitation, instability, or viscosity outside the target range.

Biofilm model:

- [1] **Bacterial Strains:** *Streptococcus gordonii* ATCC 10558, *Fusobacterium nucleatum* ATCC 25586, *Porphyromonas gingivalis* ATCC 33277.
- [2] **Culture Conditions:** Strains maintained anaerobically (80% N₂, 10% H₂, 10% CO₂) at 37°C in Brain Heart Infusion (BHI) broth supplemented with hemin (5 µg/mL) and menadione (1 µg/mL).
- [3] **Biofilm Growth:** Biofilms grown on pegs of Calgary Biofilm Devices (CBD, Innovotech Inc.) for 72 hours anaerobically at 37°C. Medium was refreshed at 24h and 48h. Mature biofilms were confirmed by scanning electron microscopy (SEM) on representative pegs.

Antimicrobial efficacy testing:

- [1] **Treatment:** Biofilm-coated pegs were exposed to 200 µL of each gel formulation for 60 minutes anaerobically at 37°C. Controls included untreated biofilms (negative) and 0.2% CHX solution (positive).
- [2] **Recovery and Quantification:** Treated pegs were rinsed twice in PBS, sonicated (40 kHz, 5 min) in 1 mL PBS to dislodge biofilm and vortexed (30 sec). Serial dilutions were plated on BHI blood agar plates. Plates were

incubated anaerobically for 5-7 days and viable colonies (CFU/mL) were counted.

- [3] **Efficacy Metric:** $\text{Log}_{10} \text{ reduction} = \text{Log}_{10}(\text{CFU/mL control}) - \text{Log}_{10}(\text{CFU/mL treated})$. Formulations were categorized as "High Efficacy" ($\geq 3 \log_{10}$ reduction) or "Low Efficacy" ($< 3 \log_{10}$ reduction) for classification tasks.

Machine learning pipeline:

- [1] **Data splitting:** Dataset randomly split into training (70%, n=420) and testing (30%, n=180) sets, stratified by efficacy category.
- [2] **Preprocessing:** Categorical features one-hot encoded. Numerical features standardized (z-score normalization).
- [3] **Model selection and training:** Four algorithms were implemented using Python (scikit-learn, TensorFlow):
 - 1) Random Forest (RF)
 - 2) Support Vector Machine (SVM; RBF kernel)
 - 3) Gradient Boosting (GB; XGBoost implementation)
 - 4) Neural Network (NN; 3 hidden layers [64, 32, 16 neurons], ReLU activation, Adam optimizer)
 Models trained on the training set with 5-fold cross-validation for hyperparameter tuning (e.g., n_estimators, max_depth for RF/GB; C, gamma for SVM; epochs, batch size for NN).
- [4] **Performance evaluation (Testing Set):**
 - 1) **Classification (High/Low Efficacy):** Accuracy, Precision, Recall, F1-Score, AUC-ROC.
 - 2) **Regression (Log₁₀ Reduction):** Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R².
- [5] **Feature Importance:** Computed using permutation importance (RF, GB, SVM) and SHAP (SHapley Additive exPlanations) values for GB.
- [6] **Baseline Comparison:** Multiple Linear Regression (MLR) was trained and evaluated for the regression task.

Statistical analysis:

Experimental data presented as mean ± standard deviation (SD). Differences in log₁₀ reduction between formulation groups assessed by one-way ANOVA with Tukey's post-hoc test. Differences in ML model performance metrics assessed by paired t-tests (Bonferroni-corrected for multiple comparisons). Significance level was p < 0.05 (SPSS v28.0, IBM; Python SciPy/Statsmodels).

Results:

Significant differences in antimicrobial efficacy were observed across formulation types (p < 0.001). Chitosan-based gels generally outperformed Carbopol and Poloxamer. The highest efficacy was achieved by Chitosan (High MW) + 1.5% CHX + 1.0% EDTA (4.2 ± 0.3 log₁₀ reduction). Poloxamer gels with CPC showed the lowest efficacy (1.8 ± 0.4 log₁₀ reduction). All ML models significantly outperformed random guessing (AUC-ROC > 0.5, p < 0.001).

Gradient Boosting (GB) demonstrated superior overall performance:

- [1] **Accuracy:** GB (92.8 ± 2.1%) > RF (89.5 ± 2.4%) > NN (87.3 ± 3.0%) > SVM (85.1 ± 2.8%)
- [2] **AUC-ROC:** GB (0.96 ± 0.03) > RF (0.93 ± 0.04) > NN (0.91 ± 0.05) > SVM (0.89 ± 0.04)
- GB showed significantly higher accuracy and AUC-ROC than SVM and NN (p < 0.01) and higher AUC-ROC than RF (p < 0.05).

GB again achieved the best predictive performance for the continuous efficacy metric:

- [1] **MAE:** GB (0.32 ± 0.08) < RF (0.41 ± 0.09) < NN (0.48 ± 0.11) < SVM (0.55 ± 0.12)
- [2] **R²:** GB (0.88) > RF (0.82) > NN (0.78) > SVM (0.73)
- GB's MAE was significantly lower than all other models (p < 0.01) and its R² was significantly higher (p < 0.01). GB

significantly outperformed the MLR baseline (MAE: 0.81 ± 0.12, R²: 0.61; p < 0.001) (**Table 1-3**).

Permutation importance and SHAP analysis identified the most influential features:

- [1] Antimicrobial Agent Type (Importance: 24.7%) - CHX > DOX > CPC
- [2] Antimicrobial Concentration (18.9%) - Positive correlation with efficacy
- [3] Polymer Viscosity (15.3%) - Moderate viscosity (~20,000 cP) optimal
- [4] Polymer Type (12.8%) - Chitosan > Carbopol > Poloxamer
- [5] Enhancer Type (11.2%) - EDTA > Citric Acid > None
- [6] Enhancer Concentration (8.1%) - Positive correlation
- [7] pH (5.4%) - Near-neutral pH optimal
- Other features showed minor contributions (<5%).

Table 1: Characteristics and experimental efficacy of top and bottom performing formulations (Mean ± SD)

Formulation ID	Polymer Base	Antimicrobial (Conc.)	Enhancer (Conc.)	Viscosity (cP)	pH	Log ₁₀ Reduction	Efficacy Category
F-CHX-15	Chitosan (HMW)	CHX (1.5%)	EDTA (1.0%)	22,500 ± 1,200	6.8	4.2 ± 0.3	High
F-DOX-12	Chitosan (MMW)	DOX (1.2%)	EDTA (0.8%)	19,800 ± 950	7.0	3.9 ± 0.4	High
F-CHX-10	Carbopol	CHX (1.0%)	Citric Acid (0.8%)	15,200 ± 800	6.5	3.5 ± 0.3	High
F-CPC-05	Poloxamer	CPC (0.05%)	None	8,500 ± 400	7.2	1.8 ± 0.4	Low
F-DOX-08	Poloxamer	DOX (0.8%)	None	9,100 ± 500	7.1	2.1 ± 0.3	Low
F-CHX-05	Carbopol	CHX (0.5%)	None	12,000 ± 600	6.2	2.4 ± 0.5	Low

Table 2: Performance metrics of machine learning models for classification (high/low efficacy) on test set (Mean ± SD)

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC
Gradient Boosting	92.8 ± 2.1	0.93	0.92	0.92	0.96 ± 0.03
Random Forest	89.5 ± 2.4	0.90	0.89	0.89	0.93 ± 0.04
Neural Network	87.3 ± 3.0	0.88	0.87	0.87	0.91 ± 0.05
SVM (RBF)	85.1 ± 2.8	0.86	0.85	0.85	0.89 ± 0.04

Table 3: Performance metrics of machine learning models for regression (Log₁₀ Reduction) on Test Set

Model	MAE (log ₁₀ CFU/mL)	RMSE (log ₁₀ CFU/mL)	R ²
Gradient Boosting	0.32 ± 0.08	0.41 ± 0.10	0.88
Random Forest	0.41 ± 0.09	0.52 ± 0.11	0.82
Neural Network	0.48 ± 0.11	0.61 ± 0.13	0.78
SVM (RBF)	0.55 ± 0.12	0.69 ± 0.14	0.73
Linear Regression	0.81 ± 0.12	0.98 ± 0.15	0.61

Discussion:

This study demonstrates the successful development and validation of machine learning models, particularly Gradient Boosting, for accurately predicting the antimicrobial efficacy of periodontal gel formulations against a clinically relevant multi-species biofilm model. The high predictive accuracy (92.8% classification, MAE 0.32 log₁₀ reduction) achieved by GB highlights the significant potential of ML to streamline the development of novel topical antimicrobials for periodontitis management. The superior performance of GB aligns with its known strength in handling heterogeneous tabular data with complex interactions between features, which is characteristic of pharmaceutical formulation datasets [12]. Its ability to capture non-linear relationships likely explains its significant advantage over traditional linear regression, which failed to model the intricate dependencies between formulation components and biofilm killing effectively. The relatively strong performance of RF and NN further supports the utility of ensemble and deep learning approaches for this task, while SVM, while competent,

was less optimal, possibly due to sensitivity to feature scaling or kernel choice [13]. The feature importance analysis provides crucial, data-driven insights into formulation design. The dominant influence of the *antimicrobial agent type* and its *concentration* was expected, as CHX and DOX are well-established broad-spectrum antimicrobials effective against periodontal pathogens, while CPC is generally considered less potent against anaerobes like *P. gingivalis* [14]. The strong positive correlation between concentration and efficacy is consistent with dose-response relationships. However, the significant contribution of *polymer type* and *viscosity* underscores the critical, often underappreciated, role of the vehicle. Chitosan's prominence aligns with its inherent antimicrobial and biofilm-disrupting properties, likely synergizing with the loaded antimicrobial agent [15]. The optimal viscosity range identified (~20,000 cP) suggests a balance between sufficient biofilm contact/adhesion and adequate diffusion/release kinetics – too low viscosity may reduce retention, while too high may impede drug release [16]. The positive impact of *penetration enhancers*,

particularly EDTA, is mechanistically plausible. EDTA chelates divalent cations (Ca^{2+} , Mg^{2+}) essential for EPS integrity and membrane stability, disrupting the biofilm barrier and enhancing antimicrobial penetration [17]. Citric acid, while also chelating, may be less effective at neutral pH or could cause local irritation. The finding that near-neutral pH is optimal likely reflects the stability of the antimicrobials (especially CHX) and the physiological pH of the periodontal pocket [18]. The experimental results validating the top-performing formulations (e.g., Chitosan + 1.5% CHX + EDTA achieving $>4 \log_{10}$ reduction) are clinically significant, exceeding the typical efficacy of commercially available CHX gels (often $\sim 2\text{--}3 \log_{10}$ reduction in similar models) [19]. This demonstrates the potential of rationally designed formulations guided by ML predictions. The lower efficacy of Poloxamer-based gels, particularly with CPC, might be attributed to the thermosensitive nature of Poloxamer (potentially reducing substantivity at body temperature) and CPC's known susceptibility to inactivation by divalent cations and organic matter prevalent in biofilms [20].

Conclusion:

Machine learning, particularly Gradient Boosting, provides a powerful and accurate tool for predicting periodontal gel efficacy against complex biofilms. By identifying key formulation factors such as antimicrobial type, concentration, polymer base and enhancers, it enables rational design of superior therapies. This computational approach can accelerate innovation while reducing reliance on resource-intensive empirical screening for periodontitis management.

References:

- [1] Ebrahimi Tarki F *et al.* *Iran J Pharm Res.* 2023 **22**:e138704 [PMID: 38450220]
- [2] Bose B *et al.* *Front Microbiol.* 2022 **12**:783284 [PMID: 35185814]
- [3] Sukmarini L *et al.* *J Nat Med.* 2024 **78**:1 [PMID: 37930514]
- [4] Abdelhamid AG *et al.* *Antibiotics (Basel).* 2023 **12**:1005 [PMID: 37370324]
- [5] Li Y *et al.* *Microbiol Spectr.* 2022 **10**:e0176821 [PMID: 35196792]
- [6] Kumar S *et al.* *Comput Struct Biotechnol J.* 2023 **21**:1995 [PMID: 36950221]
- [7] Cao Y *et al.* *Amino Acids.* 2020 **52**:1 [PMID: 31797056]
- [8] Hwang HJ *et al.* *Microbiol Spectr.* 2021 **9**:e0024921 [PMID: 34494853]
- [9] Boone K *et al.* *BMC Bioinformatics.* 2021 **22**:239 [PMID: 33975547]
- [10] Haney EF *et al.* *Biomolecules.* 2018 **8**:29 [PMID: 29883434]
- [11] LoVetri K *et al.* *Methods Mol Biol.* 2010 **618**:383 [PMID: 20094877]
- [12] Teerapo K *et al.* *PLoS One.* 2019 **14**:e0218479 [PMID: 31199859]
- [13] León Madrazo A *et al.* *Comput Biol Chem.* 2022 **98**:107695 [PMID: 35605306]
- [14] Boone K *et al.* *BMC Bioinformatics.* 2018 **19**:469 [PMID: 30522443]
- [15] Wei M *et al.* *Front Cell Infect Microbiol.* 2023 **13**:1266674 [PMID: 37842001]
- [16] Arif M *et al.* *ACS Omega.* 2024 **9**:2874 [PMID: 38250405]
- [17] Manobala T. *Crit Rev Microbiol.* 2024 **1** [PMID: 39140129]
- [18] Shaban TF *et al.* *Comput Biol Med.* 2022 **140**:105065 [PMID: 34839184]
- [19] Chatupheeraphat C *et al.* *Front Cell Infect Microbiol.* 2023 **13**:1153868 [PMID: 37113135]
- [20] Alhhazmi AA *et al.* *Microbiol Spectr.* 2024 **12**:e0182723 [PMID: 38236024]