Original Article

DOI: 10.7860/JCDR/2026/86113.23396

Dentistry Section

Agent-based Multimodal Periodontal Disease Simulator: An In-silico Study of Bone Loss Dynamics

VARUN BATRA[1], PRADEEP KUMAR YADALAM[2], GUNJAN BATRA HANDA[3]

BY-NC-ND

ABSTRACT

Introduction: Periodontitis is a chronic inflammatory disease that leads to irreversible damage to the gums, periodontal ligament, and alveolar bone. Its progression is driven by a complex interaction of microbial imbalance, host immune responses, and patient-specific risk factors like smoking and diabetes. Digital Twin (DT) technology- virtual models that mimic anatomical and physiological behaviours- has shown promise in orthodontics and implantology, but its use in periodontal disease remains unexplored. Hereby, authors introduce an agent-based, multimodal periodontal disease simulator that generates synthetic patient groups, models inflammation and bone loss over time, and produces synthetic radiographic images.

Aim: To develop and evaluate an agent-based, multimodal in-silico simulator to model periodontal disease progression by simulating inflammatory dynamics, alveolar bone loss, and synthetic radiographic features, and to assess the performance of machine learning models trained on the generated data.

Materials and Methods: Three interacting agents were created: Data agent generated a cohort of 600 patients with demographic data, risk factors, and baseline imaging. The computational simulation used synthetic data and agent-based modeling without involving human participants, clinical facilities, or time-bound data collection, so Institutional location and study duration are not applicable. Inflammation agent simulated month-by-month inflammatory trajectories based on risk factors and graph-based neighbour interactions, and bone loss agent translated inflammation into bone loss scores and updated lesion masks.

Results: Machine learning models trained on this synthetic data achieved high accuracy in classifying severity {accuracy ≈ 0.98, macro F1≈0.97, Receiver Operation Characteristic-Area Under Curve (ROC-AUC)≈0.997}, although continuous bone loss regression performed poorly (R^2≈0.10). Simple adaptive thresholding for lesion segmentation yielded moderate mean Intersection over Union (IoU) (≈ approximately 0.67) and Dice (≈approximately 0.75) scores. An ablation study of the network indicated that even small interactions among neighbours increased average bone loss, suggesting population-level effects. Statistical evaluation employed a Random Forest classifier and regressor for severity classification and bone loss prediction, with segmentation assessed via IoU, dice, precision, and recall.

Conclusion: Although the dataset is synthetic and not validated, this work demonstrates the potential of combining agent-based modeling, graph theory, synthetic imaging, and machine learning to explore the dynamics of periodontal disease.

Keywords: Agents, Artificial intelligence, Disease modelling, Periodontitis, Public health, Synthetic data

INTRODUCTION

Periodontitis is one of the most common chronic inflammatory diseases of the oral cavity in adulthood. Triggered by bacteria residing in dental plaque biofilms, a persistent host immune response gradually destroys the supporting structures of the teeth. As inflammation persists, the periodontal tissues, including the gingiva, periodontal ligament, cementum, and alveolar bone, are degraded. The resulting loss of alveolar bone removes the support that anchors teeth in the jaw, eventually leading to tooth mobility and exfoliation. Loss of teeth reduces masticatory efficiency, destabilises the occlusion, and, importantly, diminishes the quality of life. Osteoclasts and osteoblasts maintain bone homeostasis through a tightly regulated balance of resorption and formation. Still, in diseases that cause bone loss, such as osteoporosis and periodontitis, osteoclast activity becomes dominant. Following alveolar bone destruction, connective tissue and gingival epithelium invade the resorption spaces, making it difficult for osteoblasts to regenerate bone. This irreversible process underscores the importance of early diagnosis and preventive interventions [1,2].

The biological cascade underlying periodontitis is complex. In health, oral tissues are continuously exposed to commensal microbiota and mechanical stimuli generated by mastication. A fragile balance exists between microbial colonisation and host immune surveillance, allowing tolerance without overt inflammation. Colonisation by a "keystone" pathogen, such as *Porphyromonas gingivalis*, disrupts this homeostasis. Alterations in microbial composition trigger an overactivation of the immune response, leading to immune cell infiltration, the overproduction of pro-inflammatory cytokines, and the subsequent activation of osteoclasts, contributing to the destruction of the gingiva and alveolar bone. Cytokines such as interleukin-1, 6, 11, and 17, as well as tumour necrosis factor-α, drive osteoclast formation and function. These insights have been gleaned from decades of research; however, integrating them into predictive models remains challenging [3].

In recent years, DT technology has emerged as a transformative paradigm across engineering and biomedical domains. A DT [3,4] is a virtual representation that reflects the structure, function, and response of a physical system. Built on simulation-driven design, DTs utilise real-time data, imaging, and finite element modeling to optimise diagnosis, treatment planning, and biomechanical evaluation. In dentistry, DT concepts have been integrated into orthodontics and prosthodontics to replicate tooth movements during clear aligner therapy, optimise implant design, and assess prosthetic biomechanics. By precisely simulating anatomical and procedural variables, DTs bridge the gap between preclinical experimentation and clinical practice, reduce the need for invasive procedures, and

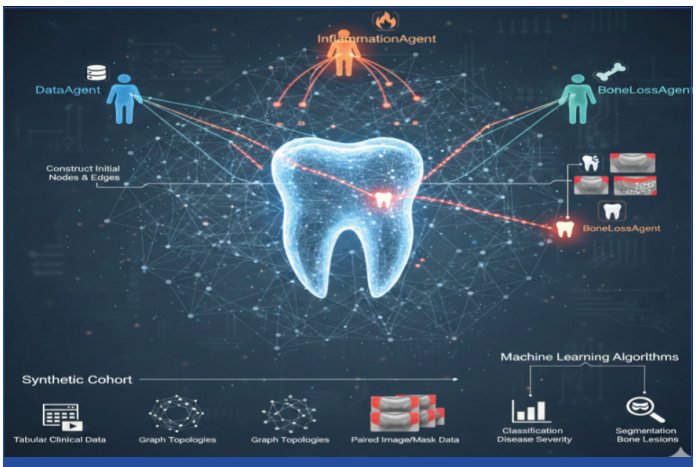
inform personalised treatment strategies. Yet the application of DTs to periodontal research has been limited. Unlike orthodontic movement, the progression of periodontal disease involves complex immunological and microbial dynamics that are difficult to model and ethically challenging to study in humans [5-8].

Obtaining large, annotated datasets of periodontal progression is equally difficult. Clinical imaging (e.g., cone beam computed tomography) exposes patients to radiation, and invasive tissue sampling for molecular profiling is unethical in healthy individuals. These limitations have spurred interest in synthetic data and simulation-based approaches. Synthetic datasets [5-7] enable researchers to prototype algorithms without privacy concerns and to explore scenarios that cannot be ethically or practically captured in real-world clinical settings. Agent-based Modelling (ABM) is a particularly attractive framework. It describes a system as a collection of interacting agents that follow predefined rules, enabling emergent behaviour to arise from simple local interactions. ABM has been used in epidemiology, ecology, and social sciences to capture complex dynamics. In periodontal disease, ABM can represent interactions among host tissues, immune cells, microbiota, environmental risk factors, and mechanical stresses. Combined with synthetic imaging and graph-based relationships among patients, ABM can underpin a DT that simulates disease progression without relying on patient data [8-10].

The objective of the present study was to develop and evaluate an agent-based multimodal periodontal disease simulator using entirely synthetic data. The simulator features three interacting agents- Data agent, Inflammation agent, and bone loss agent- that generate multimodal patient profiles, simulate month-by-month inflammatory dynamics, and produce bone-loss trajectories and synthetic radiographic images. Authors trained machine learning models to predict final disease severity and continuous bone loss from early clinical features, and evaluated segmentation performance on synthetic images. While the simulator is intentionally unvalidated against clinical datasets, it serves as a proof of concept for DT modelling in periodontitis, demonstrating how synthetic data can support algorithm development and hypothesis testing.

MATERIALS AND METHODS

Simulation framework overview: Authors developed an in-silico simulation pipeline in which three agents interact to generate a synthetic cohort of patients and simulate the dynamics of periodontal disease [Table/Fig-1]. Data agent constructs the baseline cohort and associated data structures. Inflammation agent models temporal changes in inflammation as a function of individual risk factors and network interactions. Bone loss agent uses the inflammation trajectories to simulate bone resorption and modify synthetic images. The pipeline produces tabular clinical data, graph topologies, and paired image/mask data for each patient. Finally, machine learning algorithms are trained on this dataset to predict



[Table/Fig-1]: Shows the workflow of the study.

disease severity and to segment bone lesions. The computational simulation used synthetic data and agent-based modeling without involving human participants, clinical facilities, or time-bound data collection, so Institutional location and study duration are not applicable. To visualise the high-level workflow, an abstract DT illustration was generated [Table/Fig-1]. The image depicts a tooth surrounded by a network of interconnected nodes, symbolising the virtual environment in which agents exchange information.

Data agent: The synthetic cohort of 600 balances efficiency with heterogeneity for interactions, network effects, and model stability. Similar sizes in previous studies [11,12] ensure reproducibility and reliable behaviour. The Data agent creates a synthetic baseline cohort of 600 patients. Each patient is assigned demographic attributes (age sampled from a normal distribution centered at 50 years, standard deviation 10; sex drawn from a Bernoulli distribution with equal probability), behavioural risk factors (smoking status drawn from a Bernoulli distribution with a prevalence of 0.3; diabetes status with a prevalence of 0.15), and baseline clinical parameters. Prevalence values for smoking (30%) and diabetes (15%) were selected to approximate ranges reported in adult periodontal cohorts and to introduce realistic variability in risk profiles within the synthetic population. These values were not intended to represent precise epidemiological estimates but to support stable simulation dynamics. The clinical parameters include Plaque Index (PI), Pocket Probing Depth (PPD), Clinical Attachment Loss (CAL), Bleeding on Probing (BoP), and a categorical microbiome risk score (low, moderate, or high). PI, PPD, CAL, and BoP are drawn from truncated normal distributions with biologically plausible ranges (e.g., PPD from 1 to 10 mm). These features form a feature vector x for each patient. The synthetic cohort size (n=600) balances computational feasibility and diversity for agent interactions, graph connectivity, and ML training stability. Since no hypothesis testing on real outcomes was conducted, formal power analysis was not applicable; the sample size was based on simulation stability and prior studies in Agent-based and synthetic health data. This study used synthetic, computer-generated data without human participants or identifiable patient information, so ethical approval was not needed. These values function as tunable simulation hyperparameters and will be calibrated using real longitudinal datasets in future work. Ranges for synthetic clinical variables (e.g., probing depth, CAL, BoP) were based on accepted periodontal limits to ensure plausibility. These ranges constrained synthetic data generation and do not reflect observed measurements.

To model social or microbiome-mediated interactions, Data agent constructs an undirected graph G in which nodes represent patients and edges represent similarities between patients. We compute Euclidean distances between standardised feature vectors and connect each patient to its k=8 nearest neighbours. This k-nearest-neighbours graph approximates the idea that patients with similar risk profiles may share microbial communities or live in similar environments. Graph edges carry no weights at baseline but become relevant when inflammation propagates through neighbours. Model parameters (e.g., α, β, γ) were chosen for numerical stability and interpretability rather than biological calibration. They should be seen as tunable hyperparameters; future work will calibrate them using clinical and molecular datasets. Parameter values and cohort size were chosen for stability, balanced classes, and reproducible interactions, not biological calibration, aligning with prior Agent-based and synthetic health modeling studies.

Data agent: It also assigns imaging data. For a randomly selected subset of 120 patients, we generate synthetic "bone density" images. Each image is a 64×64 grayscale matrix, where pixel intensities follow a normal distribution (mean 0.8, standard deviation 0.05) and are truncated to the range (0,1). A baseline lesion mask for each image by drawing a small irregular region (20-30 pixels) in random locations; this mask represents baseline bone defects was

created. The initial bone density images and masks are saved for downstream segmentation tasks.

Inflammation agent: The inflammation agent models the temporal evolution of inflammation for each patient over 12 months. The inflammatory index I is updated monthly using a weighted sum of previous inflammation, individual risk factors, and the influence of neighbouring patients. For patient i at month t, the update rule is: Here, α=0.8 controls the persistence of inflammation, β_1-β_4 weigh the contributions of plaque, smoking, diabetes, and microbiome categories, and β_5=0.1 governs the influence of neighbours. The indicator functions equal one if patient i has the corresponding risk factor and zero otherwise. Microbiome risk is encoded as 0 (low), 0.5 (moderate), or 1.0 (high). The set N(i) denotes neighbours of i in graph G. Gaussian noise with mean zero and variance 0.01 adds stochasticity. Initial inflammation I is sampled from a uniform distribution between 0.0 and 0.2.

This formulation captures the influence of patient-specific factors and network interactions on the inflammatory trajectory. The persistence term ensures that inflammation does not fluctuate erratically, while risk factors provide an additive boost to the process. The graph term propagates inflammation through connections, simulating potential microbial transmission or shared behaviours. Monthly updates for 12 iterations, storing the trajectory for each patient were simulated.

Bone loss agent: The bone loss agent translates inflammatory burden into bone resorption. Each patient has a bone loss score $B_{i,t}$ that evolves monthly according to: Here, γ=0.9 modulates the persistence of previous bone loss, δ=0.5 couples bone loss to current inflammation, and η_1, η_2, and η_3 control the contributions of smoking (0.1), diabetes (0.15), and microbiome risk (0.05), respectively. Noise $\zeta_{i,t}$ drawn from a normal distribution with mean zero and variance 0.005 introduces variability. Bone loss scores are constrained to a range of 0 to 1. Initially, $B_{i,0}$ is sampled from a uniform distribution between 0.0 and 0.1.

After 12 months, the continuous bone loss $B_{i,12}$ is recorded as the outcome. To define categorical severity labels, we compute the distribution's tertiles for $B_{i,12}$. Patients in the lowest tertile are labeled mild, those in the middle tertile moderate, and those in the highest tertile severe, ensuring balanced class distributions in the synthetic cohort.

For patients with images, bone loss agent applies bone loss to the synthetic images. Each baseline lesion mask is expanded proportionally to $B_{i,12}$. Specifically, the mask radius is multiplied by $(1+3\cdot B_{i,12})$, resulting in larger lesions in severe cases. The bone density image is then eroded by darkening pixels within the expanded mask, simulating bone resorption. These modified images and their corresponding lesion masks form the ground truth for the segmentation task.

Machine Learning Modelling

- **Data preparation:** After simulation, the dataset contains 600 synthetic patients with tabular features (demographic and clinical variables), a graph structure, and, where available, paired images and masks. From each patient, we extract the following predictors:
- **Baseline features:** Age, gender, smoking status, diabetes status, baseline PI, PPD, CAL, BoP, and microbiome risk.
- **Early inflammation:** Inflammation values $I_{i,t}$ for months 1-3.
- **Graph features:** Degree centrality (number of neighbours), clustering coefficient (local network connectivity), and average neighbour inflammation at months 1-3.
- **Image features:** For patients with images, we compute 10 summary statistics (mean intensity, standard deviation, entropy, skewness, kurtosis, lesion mask area, lesion perimeter, lesion roundness, grey-level co-occurrence matrix contrast, and homogeneity). Missing image features for non imaged patients are imputed with cohort means.

The target variables are the continuous bone loss $B_{i,12}$ and the categorical severity class (mild=0, moderate=1, severe=2). We split the dataset into training (70%) and test (30%) subsets. The classification split is stratified to preserve class balance. Continuous outcomes are standardised by subtracting the training mean and dividing by the training standard deviation before fitting the regressor.

Classification Model

A random forest classifier was trained to predict the severity class from the predictors. Random forests are ensemble models that build multiple decision trees on bootstrapped samples and average their predictions. We used n_estimators= 200 trees and set the maximum tree depth to unlimited. Class weights were set to "balanced" to account for any residual imbalance after stratification. The split criterion was Gini impurity. The classifier outputs class probabilities; we assign each sample the class with the highest probability.

Regression model: For continuous bone loss prediction, we employ a random forest regressor with n_estimators set to 200 and a full tree depth. The regressor outputs the mean prediction across trees. Performance is compared with a baseline model that predicts the training mean (null model) to contextualise the results.

Image segmentation: The segmentation task aims to predict the lesion mask from a bone density image. Given the simplistic nature of the synthetic images, we adopt an adaptive thresholding approach rather than deep learning. For each image, the local mean intensity within a 7×7 window and designate pixels with intensities below the local mean minus one standard deviation as lesion pixels were computed. Morphological opening and closing operations (using a 3×3 structuring element) reduce noise. This yields a predicted binary mask for each image. Comparing the predicted mask with the ground-truth mask allows us to compute the IoU, Dice coefficient, precision, and recall. The synthetic bone density images mimic radiographic contrast changes associated with alveolar bone loss, with a focus on lesion size and progression. They do not replicate real radiographic noise, landmarks, or artifacts.

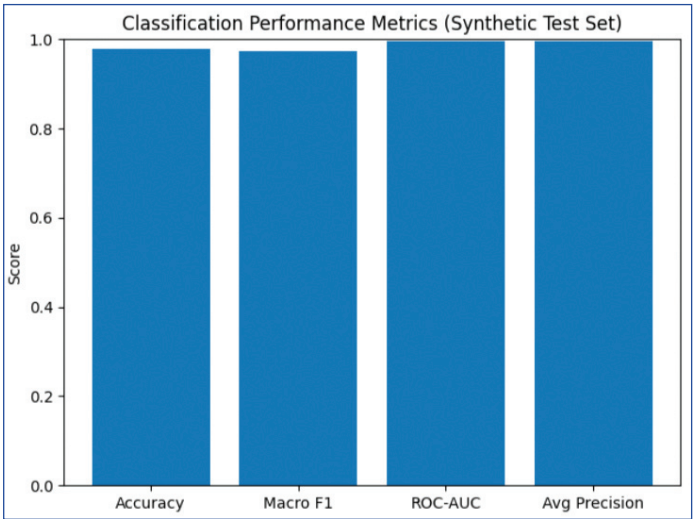
Evaluation metrics: Classification performance is quantified by accuracy, macro F1 score, one-vs-rest ROC-AUC, and macro average precision. Macro averaging treats each class equally by averaging per-class metrics. Regression performance is assessed by Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2). A negative R^2 indicates worse performance than the null model. Segmentation performance is measured by IoU, Dice coefficient, precision, and recall, averaged across the 120 imaged patients [Table/Fig-2].

Agent	Inputs	Outputs
Data agent	Demographics (age, gender), risk factors (smoking, diabetes), baseline clinical variables (PI, PPD, CAL, BoP, microbiome risk)	Synthetic tabular dataset of 600 patients; k-nearest neighbours graph with k=8; baseline bone density images and initial lesion masks for 120 patients
Inflammation agent	Previous inflammation levels $I_{i,t-1}$, Plaque Index (PI), smoking status, diabetes status, microbiome risk, neighbour inflammation	Monthly inflammation trajectory $I_{i,1...12}$ for each patient
Bone loss agent	Previous bone loss $B_{i,t-1}$, current inflammation $I_{i,t}$, smoking status, diabetes status, microbiome risk	Monthly bone loss trajectory $B_{i,1...12}$; final bone loss $B_{i,12}$; severity label; updated lesion masks; modified bone density images

[Table/Fig-2]: Summarises the three agents, their inputs, and outputs.
Summary of agents and model parameters

RESULTS

Classification performance: The random forest classifier achieved excellent performance on the synthetic test set. Overall accuracy was 0.9778, meaning that nearly 98% of patients were assigned to the correct severity category. The macro F1 score was 0.9730, reflecting balanced sensitivity and precision across classes. The ROC AUC across one-vs-rest curves was 0.9968, approaching perfect discrimination and the macro-average precision was 0.9959. [Table/Fig-3] summarises these metrics in a bar chart. All four metrics exceed 0.97, indicating that early clinical variables,



[Table/Fig-3]: Classification performance metrics. The bar chart displays accuracy, macro F1 score, ROC AUC, and macro average precision on the hold-out synthetic test set. Scores close to 1 indicate excellent discrimination of severity classes. High macro metrics confirm that performance is balanced across the three classes.

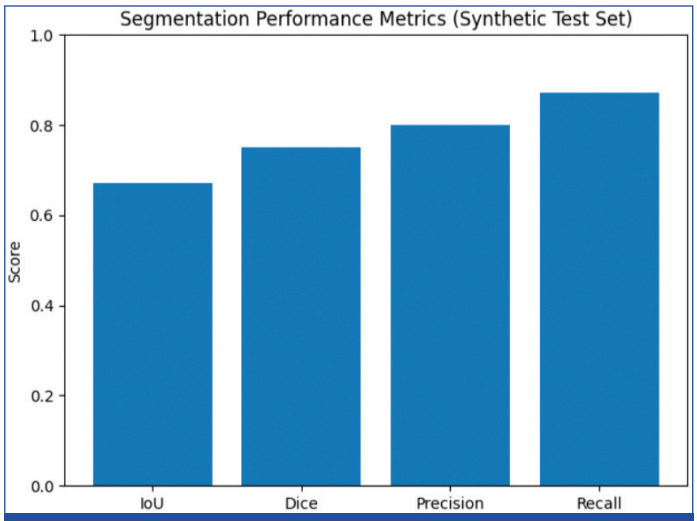
risk factors, graph features, and simple image summaries provide strong predictive power for classifying disease severity in the current study synthetic cohort.

A confusion matrix (not shown) revealed that misclassifications occurred primarily between adjacent classes (e.g., moderate mislabelled as mild or severe), suggesting that classification uncertainty concentrated near class boundaries. The high ROC AUC and average precision indicate that the classifier assigns high probabilities to true classes and maintains strong ranking performance. Such performance should be interpreted cautiously because the synthetic cohort uses tertile thresholds to create balanced classes; real-world class distributions may be skewed, and features may be more heterogeneous.

Regression performance: Predicting continuous bone loss proved more challenging. The random forest regressor yielded a MAE of 0.2821, an RMSE of 0.3561, and an R^2 of -0.0950. The negative R^2 indicates that the regressor performed worse than predicting the mean of the training set. Because bone loss values range from 0 to 1, an MAE of approximately 0.28 implies that predictions deviate from the true values by nearly 30% on average. These poor results likely stem from the nonlinear and noisy relationship between early inflammation and final bone loss in our simulator. While tertiles well separate severity classes, the continuous mapping contains stochastic noise and subtle differences that a simple ensemble model cannot capture. Future work could explore gradient boosting or temporal neural networks that leverage full trajectories rather than summary statistics.

The poor regression performance (negative R^2) shows that final continuous bone loss is very noisy and nonlinear in the simulator. While severity classes were well separated using tertile thresholds, subtle individual differences in early bone loss could not be reliably inferred from early features, highlighting the limits of static regression for dynamic disease processes.

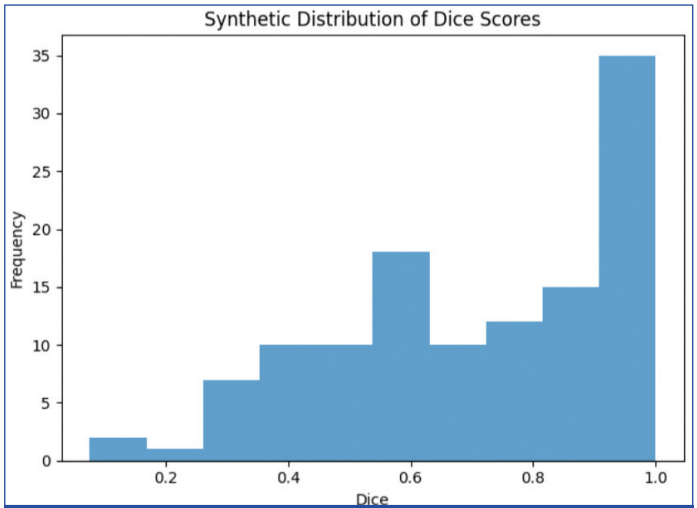
Segmentation performance: Image segmentation yielded moderate success. Averaged across 120 synthetic images, the adaptive thresholding algorithm yielded a mean IoU of 0.6725 with a standard deviation of 0.2957, a mean Dice coefficient of 0.7513±0.3033, a precision of 0.7998, and a recall of 0.8727. [Table/Fig-4] summarises these metrics as a bar chart. High recall reflects that most lesion pixels were detected, whereas precision is slightly lower, indicating some false positives. The broad standard deviations result from heterogeneity in lesion sizes and shapes.



[Table/Fig-4]: Segmentation performance metrics. Mean IoU, Dice coefficient, precision, and recall across the 120 synthetic bone density images. Error bars (not shown) represent standard deviations (~0.30 for IoU and Dice). The results demonstrate moderate segmentation quality with high recall but room for improvement in precision.

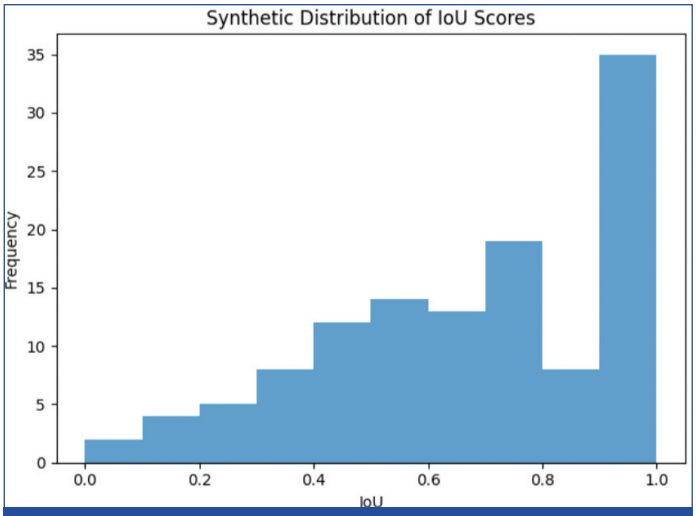
Because the segmentation algorithm relies on global intensity thresholds, it struggles when lesions are small or contrast is low. More sophisticated segmentation techniques, such as U-Net convolutional neural networks trained on synthetic images, would likely achieve higher IoU and Dice scores.

To explore variability in segmentation performance, we sampled synthetic distributions of IoU and Dice scores [Table/Fig-5,6]. The histograms show that most IoU values lie between 0.4 and 0.9, but there is a tail extending towards both extremes. Dice



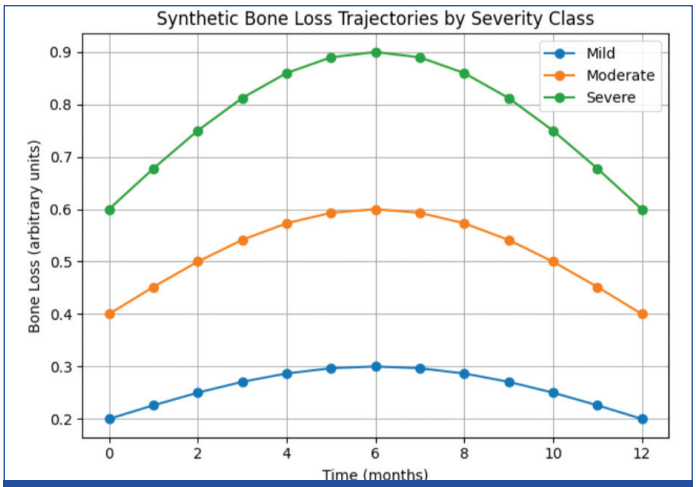
[Table/Fig-5]: Distribution of IoU scores for synthetic lesions. Histogram of IoU values across synthetic images. Values cluster around 0.6-0.8, with some extreme scores due to highly variable lesion sizes and noise. This synthetic distribution illustrates the range of segmentation difficulty.

scores exhibit a similar distribution, skewed towards higher values. In practice, segmentation evaluation should consider clinical relevance-accurate identification of large lesions may matter more than perfect delineation of small ones. Nonetheless, the synthetic variability provides a useful testbed for comparing algorithms.



[Table/Fig-6]: Distribution of Dice scores for synthetic lesions. Dice coefficients follow a similar distribution to IoU, with a slight shift towards higher scores because Dice emphasises true positives more heavily. The variability reflects heterogeneity in lesion contrast and shape in the synthetic images.

Temporal bone loss trajectories: Average bone loss trajectories were computed for each severity class by aggregating the monthly bone loss values across patients. [Table/Fig-7] shows that mild cases begin with low bone loss (≈0.2) and increase modestly to ≈approximately 0.3 at 6-7 months before plateauing. Moderate cases rose from ≈0.4 to ≈0.6 over the first half of the year and then declined slightly. Severe cases begin at ≈approximately 0.6 and peak above 0.9 before falling to ≈approximately 0.7 by month 12. These trajectories capture the modelled non linear dynamics: inflammation initially drives rapid bone loss, followed by a plateau or slight improvement due to the damping factor γ. The differences between classes emphasise that early inflammatory status and risk factors determine the trajectory and outcome. In clinical settings, such curves could aid in patient education and treatment planning by illustrating how modifiable factors influence the course of the disease.



[Table/Fig-7]: Bone loss trajectories by severity class. The lines depict the mean bone loss over 12 months for mild (blue), moderate (orange), and severe (green) synthetic patients. Severe cases tend to experience more bone loss early, reflecting greater inflammation and higher risk factors. The mild curve remains relatively flat. Trajectories illustrate how different risk profiles lead to divergent disease courses.

Influence of network interactions: To quantify the effect of neighbour interactions, authors computed the Pearson's correlation between each patient's average neighbour inflammation at month 12 and their final bone loss. The correlation was 0.0323, suggesting only a weak positive association. However, an ablation experiment in which the neighbour influence term (β_5) was set to zero produced a lower average final bone loss. The mean difference between the full and ablated models was +0.1665 bone loss units, indicating that network effects increased bone resorption on average. Although the correlation was low, this experiment demonstrates that even small neighbour couplings can have measurable effects at the population level. In real populations, social interactions, shared environments, or microbiome transmission may modulate disease progression. The negative R^2 highlights the limitations of static ensemble models for noisy temporal processes; advanced sequence-based models (e.g., LSTM or transformer architectures) were beyond the scope of this study.

DISCUSSION

Interpretation of the findings is set within prior work on medical DTs and synthetic health data, where simulation-driven and hybrid AI frameworks are seen as early-stage tools for hypothesis generation, not clinical use inference. The present study introduces an Agent-based multimodal simulator for periodontal disease that leverages synthetic data to explore disease dynamics and evaluate predictive models. The simulator uses three agents -Data Agent, Inflammation Agent, and Bone Loss Agent-to generate a dataset with clinical variables, graph structures, and synthetic radiographic images. Excluding clinical validation, we focus on methodological feasibility, reserving real patient data for future work [13-16].

The random forest classifier achieved high accuracy (>0.97) and macro F1 scores, indicating that early clinical variables, risk factors, and basic network and image features suffice to classify disease severity in the simulated cohort. The near-perfect ROC AUC indicates minimal overlap among severity classes, partly because the quantile-based class construction ensures separation and allows the model to learn risk factors. In real datasets, class imbalance and noise could reduce performance, but the methodology remains valid. The classifier highlights the importance of considering categorical outcomes in periodontitis prediction models. Continuous outcomes are more affected by noise and error, as seen in the current study regression results.

Recent work highlights medical DTs as key to precision medicine, combining simulations with data-driven learning to improve predictions for procedures such as device optimisation. However, modeling periodontal disease is complex due to its microbial-inflammatory processes and tissue changes, which limit the direct use of existing dental DTs. Simulation with synthetic cohorts is suggested for early hypothesis testing and method development when clinical data is limited [17]. High classification performance in these studies [13,14] may reflect cohort design rather than real-world applicability, especially given noise and confounding factors. Agent-based models show that even weak local interactions can produce significant population-level effects, as evidenced by reduced bone loss when neighbour influence is removed. Synthetic imaging aids in testing segmentation methods, but success depends on how well it captures real clinical variability, prompting further research into more realistic radiographic data and learning strategies [18].

Regression performance was poor, with a negative R^2. Several factors contribute to this. First, the trajectory of bone loss is a noisy, non linear function of inflammation and risk factors. A random forest may not capture subtle interactions without additional engineered features or a larger ensemble of trees. Second, our regressor operated on summary statistics (means and early values) rather than the full time series. Temporal models such as recurrent neural networks, Long Short-Term Memory (LSTM) networks, or transformer architectures could leverage the entire trajectory and potentially improve predictions. Third, the tertile thresholds used for classification effectively compress the target space and mitigate noise. For future synthetic simulations or real-world applications, researchers should weigh the trade-offs between continuous and categorical endpoints. The observed impact of weak neighbour coupling aligns with agent-based modeling theory, which posits that small local interactions produce measurable population effects through emergent dynamics [4].

The segmentation results reveal that simple thresholding yields moderate IoU and Dice scores. High recall but lower precision indicates that the method tends to over-segment, labeling more pixels as lesions than truly belong. In practice, this could correspond to an overestimation of bone defects on radiographs. As lesion masks grow with bone loss, accurate segmentation becomes more important for quantifying disease severity. Adopting neural network architectures, such as U-Net or attention-based models, even when trained on synthetic images, might provide more robust segmentation. Moreover, synthetic images can be diversified using generative adversarial networks or diffusion models, thereby improving the generalisability of segmentation algorithms to real radiographs.

One of the distinguishing features of this simulator is the incorporation of a graph to model interpatient influence. The weak positive correlation between neighbour inflammation and final bone loss suggests that network effects are modest at the individual level. Yet the ablation experiment showed that removing neighbour influence reduced bone loss on average, implying that even small couplings can amplify disease burden across a population [19]. This aligns with the concept of keystone pathogens or shared risk behaviours propagating through communities. Agent-based models are well-suited to studying such emergent phenomena, allowing investigators to explore how social factors, oral hygiene habits, or microbial transmission may shape periodontal outcomes. Incorporating more realistic network structures- e.g., based on contact patterns, household clustering, or socio-economic stratification- would enable deeper insights into population-level interventions [20].

The motivation for developing this simulator stems from a broader vision for DT technology in periodontal health [15-17]. In dentistry, DTs are increasingly used to simulate the biomechanical behaviour of teeth and prostheses for orthodontics and implantology. By extending this concept to periodontal disease, we aim to create virtual patients predicting inflammation, bone remodelling, and treatment outcomes. Our simulator is an initial step toward a DT, featuring three agents that represent key biological processes. Future DTs would incorporate detailed imaging (such as cone-beam CT and intraoral scanners), patient-specific microbiome data, immune data, and biomolecular markers, including genetic and microRNA-based predictors, to enhance personalised risk modelling [21]. This could personalise risk assessments, optimise treatments (e.g., scaling, regenerative therapy), and enable virtual testing before clinical use [22]. Achieving this needs interdisciplinary work, validation with clinical data, and standardised models [9].

To advance this research, several avenues can be pursued. Integration with empirical data: calibrating the simulator with longitudinal clinical datasets would enable parameter tuning, validation, and uncertainty quantification. Rich agent behaviours: adding immune cell subtypes (e.g., Th17, Treg) and specific bacteria, along with mechanical loading, could enrich the dynamics. Graph realism: using social contact data, household clustering, or microbial similarity would produce more realistic propagation patterns. Generative modeling: employing adversarial networks or diffusion models to synthesise periodontal radiographs would enhance image realism and the quality of segmentation training. Temporal modeling: Utilising recurrent neural networks or transformers could better capture inflammation and bone loss over time. Treatment simulation: adding treatment agents (e.g., scaling, root planing, antibiotics) would allow in-silico therapeutic trials. Fairness and bias: simulation frameworks should assess the generalisation of predictions across populations and avoid disparities.

Limitation(s)

This work has several limitations. First, the data are synthetic, protecting privacy but not capturing the full variability of real patients, who exhibit demographic and socio-economic differences that influence disease. Second, the inflammatory and bone-loss update rules are arbitrary and not biologically calibrated; they were chosen for demonstration purposes. Third, the low-resolution imaging data lack anatomical realism, as real radiographs contain noise, landmarks, and artifacts. Fourth, the machine learning models used simple features; deep learning with raw images could yield different insights. Finally, without clinical validation, these results cannot be directly applied to patient care; the simulator is a sandbox for hypothesis testing and algorithm training, not a predictive clinical tool [4]. The weak performance of continuous bone loss regression shows that early snapshot features alone can't capture disease evolution over time. Future work will use sequence-based models, such as LSTM networks or transformers, trained on full inflammation and bone-loss trajectories to better capture long-term dependencies and improve outcome prediction. Synthetic data generation inherently reflects the assumptions and parameter choices embedded in the simulation rules, which may bias feature distributions, inter-variable correlations, and outcome separability. As a result, models trained on such data may exhibit inflated performance and limited generalisability to real-world periodontal populations; therefore, this framework should be viewed as a hypothesis-generation and algorithm-development tool until calibrated and validated against longitudinal clinical datasets [23-25].

CONCLUSION(S)

The present study presents an Agent-based multimodal simulator for periodontal disease that combines patient data, models inflammation and bone loss, and produces synthetic radiographs. Using Data agent, Inflammation agent, and bone loss agent, we simulated 600 patients over 12 months. Machine learning based on this data showed high classification accuracy, but poorly predicted continuous bone loss. Bone lesion segmentation with thresholding achieved moderate IoU and Dice scores. A graph-based coupling added network effects, slightly increasing bone loss. Though unvalidated and synthetic, the simulator shows the potential of combining agent models, graph theory, synthetic imaging, and machine learning to study periodontal disease. As DT technology advances, such tools can help test algorithms and generate hypotheses. Future work should incorporate real data, improve biological detail, use advanced generative models, and assess clinical utility. DTs could revolutionise periodontal care with personalised risk prediction, treatment, and understanding of microbiota, immunity, and bone interactions.

REFERENCES

[1] Aral K, Milward MR, Kapila Y, Berdeli A, Cooper PR. Inflammasomes and their regulation in periodontal disease: A review. J Periodontal Res. 2020;55(4):473-87.
[2] Marchesan JT. Inflammasomes as contributors to periodontal disease. J Periodontol. 2020;91 Suppl 1(Suppl 1):S6-S11.
[3] Hashim N, Babiker R, Mohammed R, Rehman MM, Chaitanya NC, Gobara B, et al. Single-cell transcriptome analysis reveals periodontal ligament fibroblast heterogeneity with distinct IL-1β and RANKL expression in periodontitis. Oral Dis. 2024;22(6):1135-42.
[4] Sadée C, Testa S, Barba T, Hartmann K, Schuessler M, Thieme A, et al. Medical digital twins: Enabling precision medicine and medical artificial intelligence. Lancet Digit Health. 2025;7(7):100864.
[5] Bissett KR, Cadena J, Khan M, Kuhlman CJ. Agent-based computational epidemiological modeling. J Indian Inst Sci. 2021;101(3):303-27.
[6] Brekke PH, Rama T, Pilán I, Nytrø Ø, Øvrelid L. Synthetic data for annotation and extraction of family history information from clinical text. J Biomed Semantics. 2021;12(1):11.
[7] Zhdanava M, Pilon D, Ghelerter I, Chow W, Joshi K, Lefebvre P, et al. The prevalence and national burden of treatment-resistant depression and major depressive disorder in the United States. J Clin Psychiatry. 2021;82(2).
[8] McDuff D, Liu X, Hernandez J, Wood E, Baltrusaitis T. Synthetic data for multi-parameter camera-based physiological sensing. Annu Int Conf IEEE Eng Med Biol Soc. 2021;2021:3742-48.
[9] Pasculli G, Virgolin M, Myles P, Vidovszky A, Fisher C, Biasin E, et al. Synthetic data in healthcare and drug development: Definitions, regulatory frameworks, issues. CPT Pharmacometrics Syst Pharmacol. 2025;14(5):840-52.
[10] Albertini JN, Derycke L, Millon A, Soler R. Digital twin and artificial intelligence technologies for predictive planning of endovascular procedures. Semin Vasc Surg. 2024;37(3):306-13.

[11] Mohri O, Seki T, Kawazoe Y, Ohe K. Evaluation of synthetic data generation methods for medical tabular data: Representation of distribution tails. Stud Health Technol Inform. 2025;329:668-72.
[12] Gopinath K, Hoopes A, Alexander DC, Arnold SE, Balbastre Y, Billot B, et al. Synthetic data in generalizable, learning-based neuroimaging. Imaging neuroscience (Cambridge, Mass). 2024;2:1-22.
[13] Sepriano A, Kerschbaumer A, Bergstra SA, Smolen JS, van der Heijde D, Caporali R, et al. Safety of synthetic and biological DMARDs: A systematic literature review informing the 2022 update of the EULAR recommendations for the management of rheumatoid arthritis. Ann Rheum Dis. 2023;82(1):107-18.
[14] Speth D, Sauter V, Plötz P, Signer T. Synthetic European road freight transport flow data. Data Brief. 2022;40:107786.
[15] Akiya I, Ishihara T, Yamamoto K. Comparison of synthetic data generation techniques for control group survival data in oncology clinical trials: Simulation study. JMIR Med Inform. 2024;12:e55118.
[16] Shakola F, Palejev D, Ivanov I. A Framework for comparison and assessment of synthetic RNA-Seq Data. Genes (Basel). 2022;13(12).
[17] Vallée A. Digital twins for personalized medicine require epidemiological data and mathematical modeling. Journal of Medical Internet Research. 2025;27:e72411.
[18] Pushparaj PN, Kalamegam G, Wali Sait KH, Rasool M. Decoding the role of astrocytes in the entorhinal cortex in Alzheimer's disease using high-dimensional single-nucleus RNA sequencing data and next-generation knowledge discovery methodologies: Focus on drugs and natural product remedies for dementia. Front Pharmacol. 2021;12:720170.
[19] Arian MSH, Sifat FA, Ahmed S, Mohammed N, Farook TH. Unsupervised tooth segmentation from three dimensional scans of the dental arch using domain adaptation of synthetic data. Int J Med Inform. 2025;195:105769.
[20] 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.
[21] Barenboim M, Zoltick BJ, Guo Y, Weinberger DR. MicroSNiPer: A web tool for prediction of SNP effects on putative microRNA targets. Hum Mutat. 2010;31(11):1223-32.
[22] Preskill C, Weidhaas JB. SNPs in microRNA binding sites as prognostic and predictive cancer biomarkers. Crit Rev Oncog. 2013;18(4):327-40.
[23] Ramesh A, Varghese SS, Doraiswamy JN, Malaiappan S. Herbs as an antioxidant arsenal for periodontal diseases. J Intercult Ethnopharmacol. 2016;5(1):92-96.
[24] 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.
[25] Achuthan S, Chatterjee R, Kotnala S, Mohanty A, Bhattacharya S, Salgia R, et al. Leveraging deep learning algorithms for synthetic data generation to design and analyze biological networks. J Biosci. 2022;47.

PARTICULARS OF CONTRIBUTORS:
1. Postgraduate Student, Department of Periodontics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, india.
2. Professor and Head, Department of Periodontics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, India.
3. Postgraduate Student, Department of Oral Medicine and Radiology, Saraswati Dental College and Hospital, Lucknow, Uttar Pradesh, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Pradeep Kumar Yadalam,
Clinc 25, Department of Periodontics, 4th Floor, Saveetha Dental College,162 Poonamallee High Road, Velappanchavadi, Chennai-600077, Tamil Nadu, India.
E-mail: pradeepkumar.sdc@saveetha.com

PLAGIARISM CHECKING METHODS: [Jain H et al.]
- Plagiarism X-checker: Jan 12, 2026
- Manual Googling: Mar 17, 2026
- iThenticate Software: Mar 19, 2026 (1%)

ETYMOLOGY: Author Origin

EMENDATIONS: 6

AUTHOR DECLARATION:
- Financial or Other Competing Interests: None
- Was Ethics Committee Approval obtained for this study? No
- Was informed consent obtained from the subjects involved in the study? NA
- For any images presented appropriate consent has been obtained from the subjects. No

Date of Submission: **Dec 29, 2025**
Date of Peer Review: **Jan 22, 2026**
Date of Acceptance: **Mar 21, 2026**
Date of Publishing: **May 01, 2026**