Vol. 8, No. 4, August 2026

I S I Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Machine Learning Models for Early Gingival Health Screening Using Routine Clinical and Behavioral Data

Novia Lestari[1], **Ozzy Secio Riza**[2]

[1,2]Information Systems Study Program, Universitas Islam Negeri Imam Bonjol Padang, Padang, Indonesia

Received:
February 4, 2026
Revised:
March 10, 2026
Accepted:
June 24, 2026
Published:
August 22, 2026

Corresponding Author:
Author Name*:
Ozzy Secio Riza
Email*:
ozzysecioriza@uinib.ac.id

DOI:
10.63158/journalisi.v8i4.1640

Abstract. This study addresses the high prevalence of periodontal disease in Indonesia by developing an accessible, early screening tool that overcomes the limitations of existing AI approaches, which heavily rely on costly radiographic imaging. We developed a machine learning framework utilizing demographic, clinical, behavioral, and medical-history data from 500 patients at the Maninjau Community Health Centre. Four algorithms—Random Forest, Logistic Regression, Support Vector Machine (SVM), and XGBoost—were evaluated using stratified 5-fold cross-validation and ROC curve assessment. The results demonstrate that Random Forest achieved the highest overall classification performance, attaining an accuracy of 95.2%, an AUC of 0.983, and an F1-score of 0.9521, with superior capability across all categories (Normal, Gingivitis, and Periodontitis). To demonstrate practical applicability, this model was deployed as a Streamlit-based prototype decision-support application for real-time clinical use. This research provides significant value by demonstrating that routinely collected, non-invasive clinical and behavioral data can serve as potent predictors for periodontal screening, eliminating the dependency on expensive imaging. This offers a cost-effective digital health framework suitable for primary healthcare settings. Future research will focus on validating this model using larger, multi-center datasets to ensure broader generalizability.

Keywords: Gingival Health Screening; Machine Learning; Random Forest; Multiclass Classification; Clinical Decision Support; Periodontal Disease

1. INTRODUCTION

Gingival health is essential for maintaining oral function, mastication, speech, aesthetics, and long-term tooth retention while also contributing to overall systemic health [1]. Pathological changes in the gingiva commonly manifest as edema, bleeding, and attachment loss, which may progress to periodontal disease if left untreated [2]. Current periodontal classification systems distinguish between periodontal health, gingivitis, and periodontitis using standardized clinical criteria. The 2018 Classification of Periodontal and Peri-Implant Diseases and Conditions developed jointly by the American Academy of Periodontology (AAP) and the European Federation of Periodontology (EFP) provides an internationally accepted framework for periodontal diagnosis and classification. This framework incorporates clinical indicators such as gingival inflammation, bleeding on probing, periodontal pocket depth, and clinical attachment loss to differentiate periodontal health states. These standardized classifications provide an important clinical foundation for the development and evaluation of machine-learning-based screening and risk-stratification systems [3], [4]. These standardized classifications provide an important clinical foundation for the development of machine-learning-based screening and risk-stratification tools.

Beyond its impact on the oral cavity, periodontal disease has been associated with several systemic conditions, including cardiovascular disease, diabetes mellitus, and adverse pregnancy outcomes. The high prevalence of periodontal disease in Indonesia, reported at 74.1% in the 2018 Basic Health Research (Riskesdas) survey [5], highlights the urgent need for effective preventive strategies and early screening approaches. Delayed detection, limited public awareness of early symptoms [6], and restricted access to professional dental care frequently result in disease progression before treatment is initiated. Consequently, there is a growing need for accessible and practical screening approaches capable of identifying gingival health problems at an early stage, particularly in community healthcare settings where access to specialized periodontal assessment may be limited [7].

In recent decades, the integration of advanced computational methods, particularly machine learning (ML), has revolutionized medical diagnostics [8], A machine learning model is a mathematical representation or algorithm designed to learn patterns or relationships in data and make predictions based on that data. These models are created through a training process using data, where the model "learns" to recognize patterns and adjust its internal parameters to produce appropriate outputs [9].

The field of dentistry has seen a growing number of studies employing ML for tasks such as radiographic analysis and intraoral image classification to identify periodontal disease [9]. While these approaches have shown promise, they are fundamentally constrained by the need for specialized and expensive imaging equipment. This reliance on radiographic imaging and specialized diagnostic infrastructure may limit scalability and hinder widespread implementation in primary healthcare settings, particularly in resource-constrained environments [10], [11], [12].

In contrast, the demographic, clinical, behavioural, and medical-history variables used in this study can be collected during routine dental examinations and patient interviews without requiring radiographic imaging, laboratory biomarkers, or specialized diagnostic equipment. These variables were selected based on their documented associations with periodontal health and disease progression, as previous clinical and epidemiological studies have identified factors such as age, smoking, diabetes status, oral hygiene behaviors, and clinical periodontal indicators as important determinants of periodontal risk and severity [9], [13]. The primary focus of this study, however, is predictive model performance and deployment feasibility rather than quantifying the relative contribution of individual predictors.

Previous research has highlighted the growing application of artificial intelligence (AI) and machine learning (ML) in healthcare [8], particularly in medical image-based diagnostics. For example, [13] developed a machine learning model utilizing age information, self-reported data, gingival bleeding during tooth brushing, and levels of the aMMP-8 enzyme. [14] reported that a Naive Bayes Classifier achieved 92.32% accuracy with a kappa value of 0.833 in classifying dental diagnosis risks. A deep learning method based on CNN for periodontal disease detection from radiographs was presented by [15]. Despite the growing adoption of machine learning for periodontal disease detection, most existing approaches rely on radiographic images, intraoral photographs, or specialized clinical measurements, limiting their applicability in primary healthcare environments. This study addresses this gap by developing a multiclass gingival health classification framework based on routinely collected demographic, clinical, behavioral, and medical-history data that can be obtained during standard dental examinations. The novelty of this research lies in the comparative evaluation of four machine learning algorithms—Random Forest, Logistic Regression, Support Vector Machine (SVM), and XGBoost—using stratified 5-fold cross-validation on a real-world community health-center dataset, together with the development of a Streamlit-based prototype decision-support application. The proposed framework is intended to support preliminary gingival health screening and periodontal-risk stratification rather than replace comprehensive periodontal examination. Definitive diagnosis and treatment decisions remain the responsibility of qualified dental professionals. Accordingly, this study aims to evaluate the predictive performance of these algorithms for multiclass gingival health classification and to assess the feasibility of implementing the best-performing model as a Streamlit-based screening-support application. The findings contribute to the growing field of oral health informatics by demonstrating the potential of routine clinical and behavioral data for supporting accessible and cost-effective gingival health screening in community healthcare settings.

2. METHODS

This research was conducted using the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework [14], [16] a well-established approach for data-centric initiatives. The study followed a structured workflow consisting of six comprehensive phases of the CRISP-DM framework: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. Grounding the technical workflow in all six phases ensures a clear transition from data exploration to practical clinical application.

2.1. Business Understanding

The central business objective is to enhance the efficacy of early detection of periodontal disease at the primary healthcare level. The problem identified is the high national prevalence of the disease, largely due to late diagnosis stemming from a lack of public awareness regarding subtle initial symptoms. The goal is to develop an automated screening model using machine learning that can effectively analyze clinical and behavioral risk factors. This model must be accurate, non-invasive, and scalable, enabling primary health centers to identify individuals who may benefit from further periodontal

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

assessment, referral, and preventive care, thereby reducing the progression to severe periodontal stages and mitigating associated systemic health risks.

2.2. Data Understanding and Collection

This study employed a retrospective observational design using secondary clinical records obtained from the Maninjau Community Health Centre. The dataset comprised patient records collected between January and July 2025 during routine dental examinations and direct patient interviews conducted by a licensed dentist. The reference diagnosis (ground truth) was assigned during the clinical assessment and subsequently extracted from the medical records for machine-learning model development. Initially, 512 patient records were screened for eligibility. After applying the inclusion and exclusion criteria and removing records with incomplete or inconsistent information, 500 patient records were retained for the final analysis.

Inclusion criteria :

1) Patients who underwent routine dental examinations at the Maninjau Community Health Centre between January and July 2025.
2) Patients with a documented periodontal diagnosis (Normal Gingiva, Gingivitis, or Periodontitis) assigned by a licensed dentist.
3) Patients with complete records for the variables included in the study, including demographic, clinical, behavioral, medical-history, and healthcare-access information.

Exclusion criteria:

1) Records with missing periodontal diagnosis labels.
2) Records with incomplete information for one or more study variables.
3) Duplicate patient records.
4) Records containing inconsistent, contradictory, or invalid data entries.

Missing values were handled through record exclusion. Records containing missing information in one or more variables required for analysis were removed during the data-screening process. No imputation techniques were applied because the number of incomplete records was relatively small. Initially, 512 patient records were screened, and

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

12 records were excluded due to incomplete or inconsistent information, resulting in a final dataset of 500 patient records for analysis.

The dataset included demographic variables (age and gender), clinical variables (dental calculus, loose teeth, and frequent gum bleeding), behavioral variables (smoking habits and tooth-brushing frequency), medical-history variables (diabetes mellitus), and healthcare-access variables (access to healthcare facilities and regular dental check-ups). A detailed description of the study variables is presented in Table 1.

Table 1. Classification Of Research Variables For Gingival Health Screening

Category	Variables	Description
Demographic	Age Gender	Basic characteristics of respondents that describe the population.
Clinical	Presence of dental calculus Loose teeth Frequent gum bleeding	Clinical conditions that can be directly observed or examined by dental professionals.
Behavioral	Smoking habits Tooth brushing frequency	Lifestyle factors that influence oral and periodontal health.
Medical History	Diabetes Mellitus	A systemic disease that increases susceptibility to periodontal disease.
Healthcare Access	Access to healthcare facilities Regular dental check-ups	External factors related to the availability and utilization of healthcare services.

The target variable for gingival health classification consisted of three categories: Normal Gingiva, Gingivitis, and Periodontitis. The reference diagnosis (ground truth) was determined through clinical periodontal examinations performed by a licensed dentist and recorded in the patient medical records. Based on the clinical assessment, individuals were classified into three categories: Normal Gingiva, Gingivitis, and Periodontitis. The machine learning models were subsequently trained using routinely collected demographic, clinical, behavioral, medical-history, and healthcare-access variables, while the dentist-assigned diagnosis served as the reference classification label.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2.3. Data Preparation

Prior to model development, the dataset underwent comprehensive cleaning and preprocessing to ensure data integrity and algorithmic compatibility. The proportion of missing values was negligible, and records with incomplete entries were excluded. Categorical variables were transformed into numerical representations using the LabelEncoder() function from the Scikit-learn library. This procedure was applied to variables including gender, smoking habits, tooth-brushing frequency, diabetes status, dental calculus, loose teeth, frequent gum bleeding, healthcare access, and regular dental check-ups. The continuous feature Age was normalized using Min-Max scaling to a range of 0–1, thereby minimizing scale-related bias in distance-sensitive algorithms.

To prevent information leakage, preprocessing operations were incorporated within the cross-validation workflow. Feature scaling was performed separately within each training fold and subsequently applied to the corresponding validation fold during model evaluation. Feature selection analysis indicated that all ten identified risk factors possessed substantial clinical relevance and were therefore retained for model training. The final dataset comprised 500 patient records, including 188 Normal Gingiva cases, 238 Gingivitis cases, and 74 Periodontitis cases.

2.4. Modelling

Predictive models were implemented in Python using the scikit-learn and XGBoost libraries. The software environment consisted of Python version 3.12.13, Scikit-learn version 1.6.1, XGBoost version 3.2.0, Streamlit version 1.58.0, Pandas version 2.2.2, and NumPy version 2.0.2. Model performance was evaluated through stratified 5-fold cross-validation to obtain reliable generalization estimates and minimize overfitting. To ensure reproducibility, a fixed random seed of 42 was used throughout the experiments, including model initialization and the stratified cross-validation procedure. For each algorithm, baseline hyperparameters were initialized based on established benchmarking practices. Default parameter values were retained unless explicitly specified in the model configuration. Hyperparameter tuning was not performed in this study. Instead, baseline model configurations were adopted to facilitate a consistent comparison of algorithm performance under comparable experimental conditions. The specific hyperparameter settings used for each machine learning algorithm are presented in Table 2. Default parameter values were retained unless explicitly specified in the model configuration.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 2. Hyperparameter Settings Used for Model Development

No.	Algorithm	Hyperparameter Settings
1	Random Forest	class_weight='balanced', random_state=42
2	Logistic Regression	max_iter=1000, class_weight='balanced', random_state=42
3	Support Vector Machine (SVM)	probability=True, class_weight='balanced', random_state=42
4	XGBoost	use_label_encoder=False, eval_metric='mlogloss', scale_pos_weight=2

The classification outcomes presented in the confusion matrices and performance tables correspond to cumulative predictions aggregated across the five validation folds. This approach reflects the complete evaluation of all 500 patient samples and provides a more reliable estimate of model performance. Four proven supervised learning algorithms are then used for comparative evaluation:

1) Random Forest: a highly reliable ensemble method. It works by collecting prediction results from various individual decision trees. This strategy inherently improves classification accuracy while strengthening the overall stability and robustness of the model [17], [18], [19].
2) Support Vector Machine (SVM): which identifies the best hyperplane to separate data points into various classifications [20].
3) XG-Boost: a gradient boosting algorithm recognized for its effectiveness and scalability in handling structured data [21], [22]
4) Logistic Regression: a linear model widely adopted as a baseline in both binary and multiclass classification problems [23], [24].

Each algorithm was trained using the preprocessed dataset to identify the model with the best predictive capability for gingival health classification.

2.5. Evaluation

The performance of each classification model was evaluated using both quantitative and graphical assessment techniques. A confusion matrix was employed to summarize prediction outcomes for each gingival health category (Normal Gingiva, Gingivitis, and Periodontitis). The confusion matrix records the number of correctly and incorrectly classified instances, thereby enabling detailed assessment of model strengths and

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

weaknesses [25]. Based on the confusion matrix, several performance metrics were calculated, including accuracy, precision, recall, and F1-score. Accuracy represents the proportion of correctly classified instances among all evaluated cases. Precision measures the proportion of positive predictions that are truly positive, whereas recall reflects the model's ability to identify all actual positive instances. The F1-score, defined as the harmonic mean of precision and recall, provides a balanced measure of classification performance. Because the dataset exhibits class imbalance across the three gingival health categories, precision, recall, and F1-score were calculated using weighted averaging.

This approach accounts for the relative frequency of each class and provides an overall performance estimate that more accurately reflects the class distribution of the dataset. The Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) were used to evaluate the discriminative ability of each model. The ROC curve illustrates the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across different classification thresholds. An AUC value closer to 1.0 indicates stronger discrimination capability between classes. Together, these metrics provide a comprehensive evaluation of model performance and predictive capability.

2.5. Deployment

To facilitate the translation of machine learning research into practical clinical applications, the best-performing model identified during the evaluation stage was implemented as a functional prototype. After comparative evaluation, Random Forest was selected as the best-performing model based on its overall classification performance. The selected model was subsequently retrained using the complete dataset of 500 patient records to maximize the use of available information. Stratified 5-fold cross-validation was used solely for model evaluation and selection, whereas the deployed model was trained using all available records. The final trained model was then serialized and integrated into a Streamlit-based prototype application developed to support gingival health screening, periodontal-risk stratification, and referral prioritization. The prototype is intended to assist the early identification of individuals who may require further periodontal assessment and is not designed to provide autonomous diagnosis, treatment recommendations, or substitute professional clinical

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

judgment. Definitive diagnosis and treatment planning remain the responsibility of qualified dental professionals.

A web-based graphical user interface (GUI) was developed using the Streamlit framework, allowing healthcare practitioners to enter patient data interactively and obtain real-time gingival health classification results. This lightweight deployment approach supports local implementation and enhances accessibility in clinical settings. The deployed application is designed as a screening-support and decision-support tool. Its outputs should be interpreted as preliminary classification results and should not be considered a substitute for professional periodontal examination or clinical diagnosis. The current prototype displays only the predicted gingival health classification category and does not present prediction probabilities or confidence scores to end users. Consequently, calibration analysis was not performed because probability-based decision support was beyond the scope of the present implementation. Future versions incorporating probability outputs may require dedicated calibration assessment. To provide a comprehensive overview of how these operational steps integrate seamlessly, the complete end-to-end methodological framework spanning from the initial business understanding to the final application deployment is systematically illustrated in Figure 1.

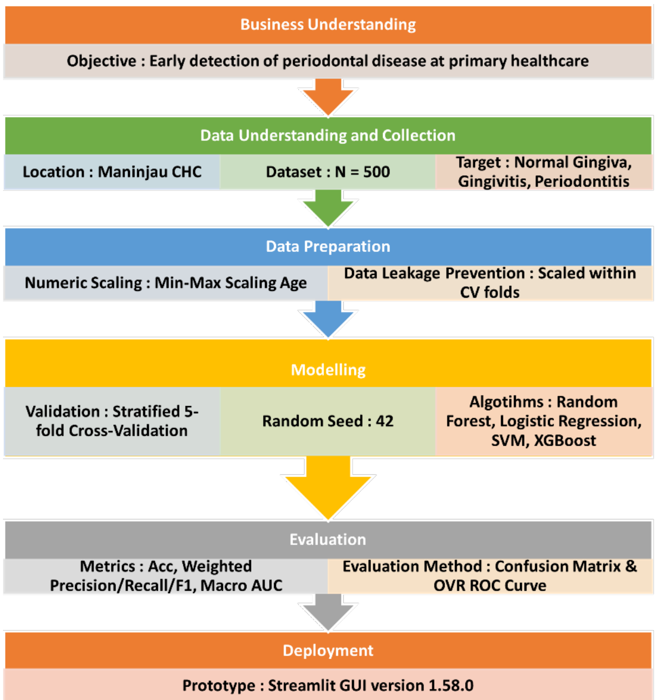

Figure 1. Research Method Flow

As delineated in Figure 1, the procedural flow guarantees that data preparation and iterative modeling transitions are explicitly aligned with the core screening objectives. Guided by this structured workflow, a preliminary evaluation of practical applicability, the prototype was reviewed by a licensed dentist to assess the clinical relevance of the selected input variables, classification outputs, and referral-support recommendations. The review indicated that the screening workflow and generated recommendations were consistent with routine periodontal screening practices. In addition, the prototype was informally tested by several users to assess functionality and ease of use. Feedback from these users suggested that the interface was intuitive and that the application could be operated without substantial technical difficulty. However, formal usability studies involving larger groups of healthcare professionals and end users have not yet been conducted and remain an important direction for future work.

2.6. Ethical Considerations and Data Governance

Formal research approval was obtained from the Head of the Maninjau Community Health Centre prior to data acquisition. The dataset consisted of anonymized secondary clinical records provided by the attending dentist following a formal research data request and approval process. Patient privacy was strictly maintained throughout the study, and all personally identifiable information (PII), including names, addresses, and patient registration numbers, was removed before data transfer, extraction, and analysis. No personally identifiable patient information was available to the researchers at any stage of the study. Because the research involved retrospective analysis of de-identified records and did not involve direct patient contact, intervention, or collection of new patient information, individual informed consent was not required. Access to the dataset was restricted to the research team and granted solely for academic research purposes. The records were stored in password-protected digital files and handled in accordance with data confidentiality principles to ensure the privacy, security, and protection of patient information.

3. RESULTS AND DISCUSSION

3.1. Performance Evaluation

The results presented in this section were obtained from the Modeling, Evaluation, and Deployment phases of the CRISP-DM methodology. A comparative analysis was

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

conducted to assess the predictive performance of four machine learning algorithms and their potential utility in automated gingival health classification. Following data preprocessing and stratified 5-fold cross-validation, Random Forest, Logistic Regression, Support Vector Machine (SVM), and XGBoost models were systematically developed and evaluated using a standardized experimental framework. To provide a more comprehensive assessment of multiclass classification performance, additional evaluation metrics were calculated. The results are presented in Table 3.

Table 3. Additional Multiclass Performance Metrics

Model	Macro Precision	Macro Recall	Weighted Precision	Weighted Recall	Balanced Accuracy
Random Forest	0.9619	0.9494	0.9532	0.9520	0.9494
Logistic Regression	0.9462	0.9386	0.9427	0.9400	0.9386
Support Vector Machine (SVM)	0.9547	0.9408	0.9466	0.9420	0.9408
XGBoost	0.9511	0.9354	0.9397	0.9380	0.9354

Table 3 presents additional multiclass evaluation metrics, including macro precision, macro recall, weighted precision, weighted recall, and balanced accuracy. These metrics were included to provide a more comprehensive assessment of model performance across the three gingival health categories and to account for the unequal class distribution within the dataset.

Model performance was comprehensively evaluated using the confusion matrices presented in Figure 2, which illustrate the distribution of correctly classified and misclassified cases across the three gingival health categories. To provide a complete representation of the 500-patient dataset, the confusion matrices were generated from aggregated out-of-fold predictions obtained during stratified 5-fold cross-validation. Each prediction corresponds to a validation sample that was not included in the training data for the respective fold. Consequently, the confusion matrices summarize model performance on unseen data across all folds and provide an unbiased estimate of classification performance. In addition, because the study addressed a multi-class

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

classification task involving Normal Gingiva, Gingivitis, and Periodontitis, the Receiver Operating Characteristic (ROC) curves shown in Figure 3 were derived using a One-vs-Rest (OvR) approach, with performance summarized through macro-averaged Area Under the Curve (AUC) values. The strong concentration of the ROC curves toward the upper-left region, as reflected in Table 3, indicates that routinely collected clinical and behavioral variables provide substantial predictive information for supporting early periodontal screening. For consistency throughout the evaluation process, class labels were encoded in the following order: Gingivitis, Normal Gingiva, and Periodontitis. This ordering was applied to both the confusion matrices and the One-vs-Rest ROC analyses.

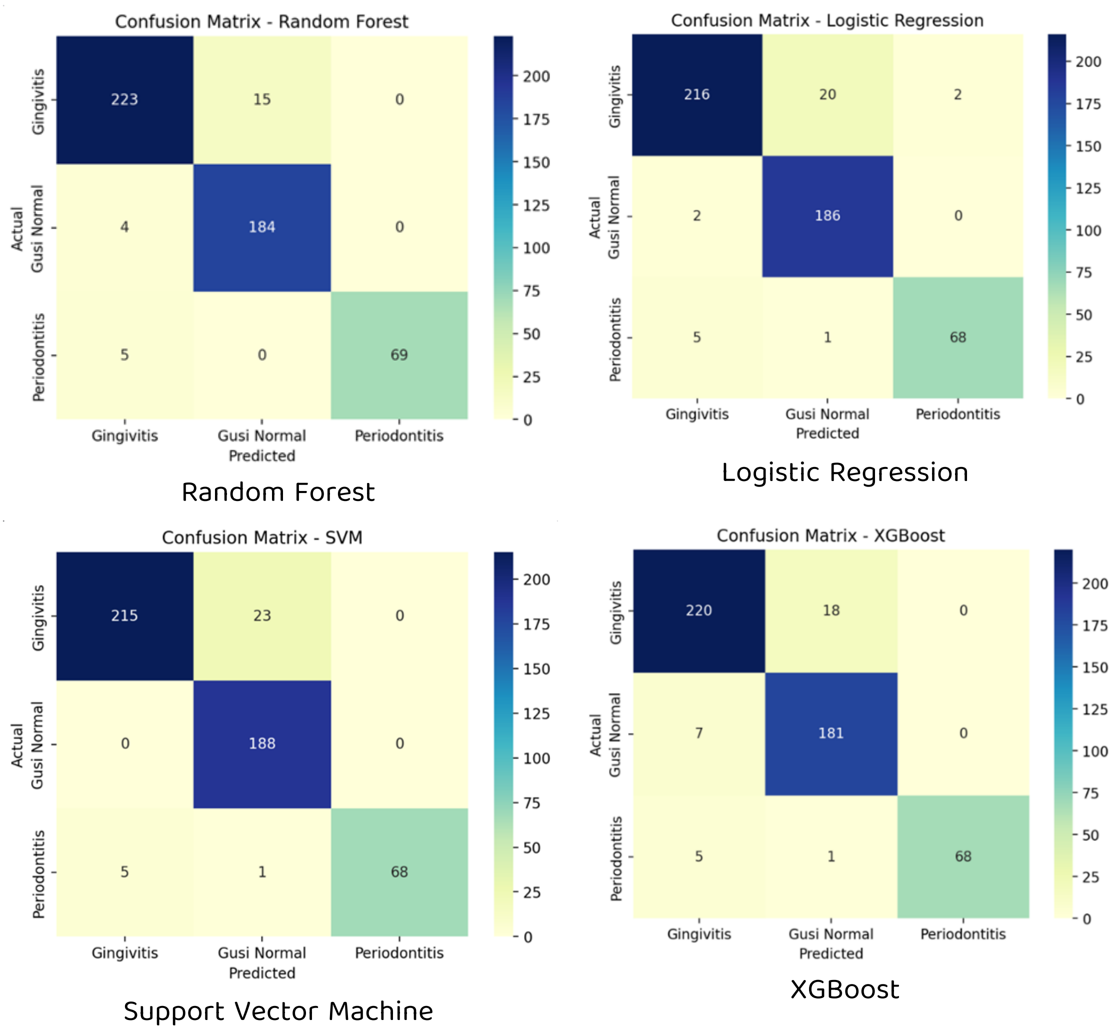

Figure 2. Confusion Matrix Visualization (class order: Gingivitis, Normal Gingiva, and Periodontitis).

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The confusion matrix results in Figure 2 indicate that all four models achieved strong classification performance with relatively few misclassifications. In this study, overall accuracy was computed as the proportion of correctly classified cases represented by the diagonal elements of the multi-class confusion matrix relative to the total number of evaluated samples. Based on this calculation, Random Forest correctly classified 476 of 500 cases (95.2%), representing the highest overall classification accuracy among the evaluated models. Logistic Regression correctly classified 470 cases (216+186+68). Based on these results, the overall accuracy of the model was 94.0% (470/500). The Support Vector Machine model correctly classified 471 cases (215+188+68), resulting in an overall accuracy of 94.2% (471/500). The SVM model also demonstrated strong classification capability, achieving an overall accuracy of 94.2%, although its performance remained slightly below that of Random Forest.

The XGBoost model achieved strong overall accuracy. With a total of 500 test data, the number of correct predictions was 469 cases (220+181+68). Based on the total correct predictions divided by the total data (469/500), the accuracy of this model is 93.8%. This performance shows that XGBoost is a highly effective in analyzing clinical data and providing reliable predictions for gingival health screening. In addition to overall accuracy, class-specific performance metrics were analyzed to evaluate model performance for each gingival health category. The results are summarized in Table 4.

Table 4. Class-Specific Performance Metrics

Model	Class	Precision	Recallll	F1-Score
Random Forest	Gingivitis	0.9612	0.9370	0.9489
	Normal Gingiva	0.9246	0.9787	0.9509
	Periodontitis	**1.0000**	**0.9324**	**0.9650**
Logistic Regression	Gingivitis	0.9686	0.9076	0.9371
	Normal Gingiva	0.8986	0.9894	0.9418
	Periodontitis	0.9714	0.9189	0.9444
Support Vector Machine (SVM)	Gingivitis	0.9773	0.9034	0.9389
	Normal Gingiva	0.8868	**1.0000**	0.9400
	Periodontitis	**1.0000**	0.9189	0.9577

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Model	Class	Precision	Recallll	F1-Score
XGBoost	Gingivitis	0.9483	0.9244	0.9362
	Normal Gingiva	0.9050	0.9628	0.9330
	Periodontitis	**1.0000**	0.9189	0.9577

Table 4 presents the class-specific precision, recall, and F1-score values for each evaluated algorithm. Across all models, classification performance remained consistently strong for the three gingival health categories. Random Forest demonstrated the most balanced class-level performance, achieving the highest F1-score for the Periodontitis category (0.9650) and the highest Periodontitis recall (93.2%). These findings are particularly important because Periodontitis represents the most clinically significant condition and the smallest class in the dataset. Logistic Regression and SVM exhibited strong performance in identifying Normal Gingiva, achieving recall values of 98.9% and 100%, respectively. Meanwhile, XGBoost achieved competitive performance across all categories, with class-specific metrics comparable to those of the other evaluated algorithms. Random Forest achieved the highest Periodontitis recall and F1-score among the evaluated models.

To further assess the discriminatory capability of the evaluated models across all gingival health categories, Receiver Operating Characteristic (ROC) analysis was performed. The ROC curves presented in Figure 3 illustrate the ability of each algorithm to distinguish among Normal Gingiva, Gingivitis, and Periodontitis across different classification thresholds. As shown in Figure 3, all evaluated models achieved ROC curves that were concentrated near the upper-left corner of the plot, indicating strong discrimination capability across the three gingival health categories. Based on the comparative evaluation, Random Forest demonstrated the strongest overall classification performance in this dataset, achieving the highest accuracy and F1-score among the evaluated algorithms.It exhibited the best overall performance with a precision of 95.2% and an F1-Score of 0.9521. Table 5 presents the overall performance comparison of the evaluated machine-learning algorithms. Precision, recall, and F1-score are reported as weighted-average metrics.

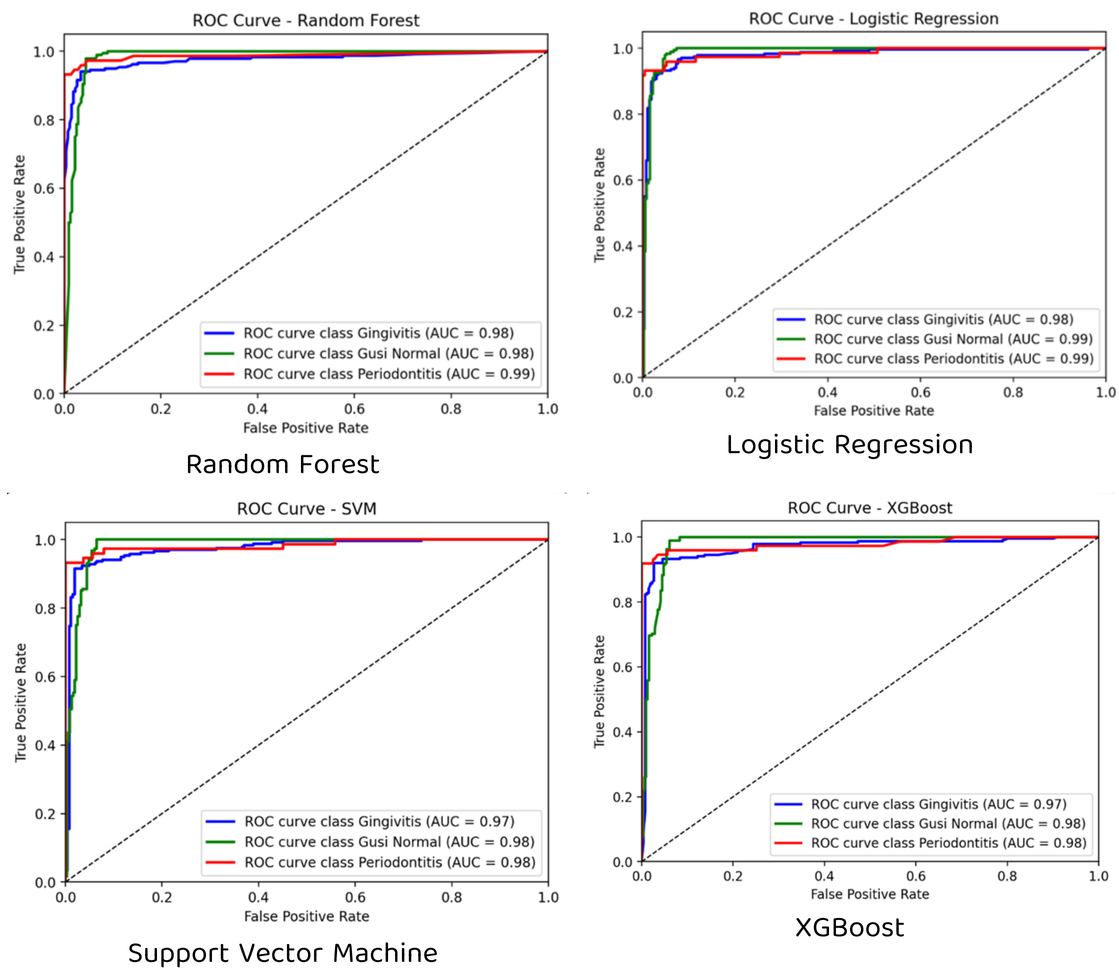

Figure 3. One-vs-Rest (OvR) ROC Curves for Gingivitis, Normal Gingiva, and Periodontitis.

Table 5. Model Performance Comparison (weighted-average F1-score)

No.	Model	Accuracy	ROC AUC	F1 Score
1	Random Forest	95.2%	0.983	0.9521
2	Logistic Regression	94.0%	0.986	0.9399
3	Support Vector Machine (SVM)	94.2%	0.976	0.9421
4	XGBoost	93.8%	0.976	0.9382

Although Random Forest achieved the highest overall accuracy (95.2%) and F1-score (0.9521), Logistic Regression achieved the highest ROC-AUC value (0.986), compared with 0.983 for Random Forest, 0.976 for SVM, and 0.976 for XGBoost. Logistic Regression may achieve a higher ROC-AUC because the model produces smoother probability estimates and stronger ranking performance across classification thresholds. Although its final class assignments resulted in slightly lower accuracy than Random Forest, the model

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

maintained excellent discrimination capability, which is reflected in its superior ROC-AUC value. Although the comparative evaluation identified Random Forest as the most suitable model for deployment, additional analysis was conducted to assess the stability and robustness of its performance across different validation folds. Therefore, fold-level performance metrics and confidence intervals were examined using stratified 5-fold cross-validation. The results are presented in Table 6.

Table 6 shows that Random Forest demonstrated consistently strong performance with a mean accuracy of 96.4% (SD = 0.0089), a mean macro F1-score of 0.9370 (SD = 0.0156), and a mean AUC of 0.9893 (SD = 0.0108). The fold-level metrics reported in Table 6 represent the average performance across the five validation folds and therefore may differ slightly from the aggregated out-of-fold results reported in Table 5. The corresponding 95% confidence intervals were 0.9529–0.9751 for accuracy, 0.9175–0.9564 for macro F1-score, and 0.9760–1.0000 for AUC. The relatively small standard deviations and narrow confidence intervals indicate stable performance across validation folds and support the robustness of the proposed Random Forest model. Having established the stability of the Random Forest model across validation folds, the next analysis aimed to identify which predictor variables contributed most strongly to the observed performance.

Table 6. Fold-Level Performance of the Selected Random Forest Model

Fold	**Accuracy**	**Macro F1**	**AUC**
1	0.970	0.9514	0.9971
2	0.970	0.9540	0.9909
3	0.960	0.9198	0.9975
4	0.950	0.9234	0.9898
5	0.970	0.9361	0.9711
Mean ± SD	**0.9640 ± 0.0089**	**0.9370 ± 0.0156**	**0.9893 ± 0.0108**
95% CI	**0.9529–0.9751**	**0.9175–0.9564**	**0.9760–1.0000**

To further investigate the factors underlying the superior performance of Random Forest, a feature-importance analysis was conducted using the final model trained on

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

the complete dataset. This analysis provides insight into the relative contribution of each predictor variable to the classification process and helps identify the routine clinical and behavioral factors that most strongly influenced gingival health classification. The results are presented in Figure 4.

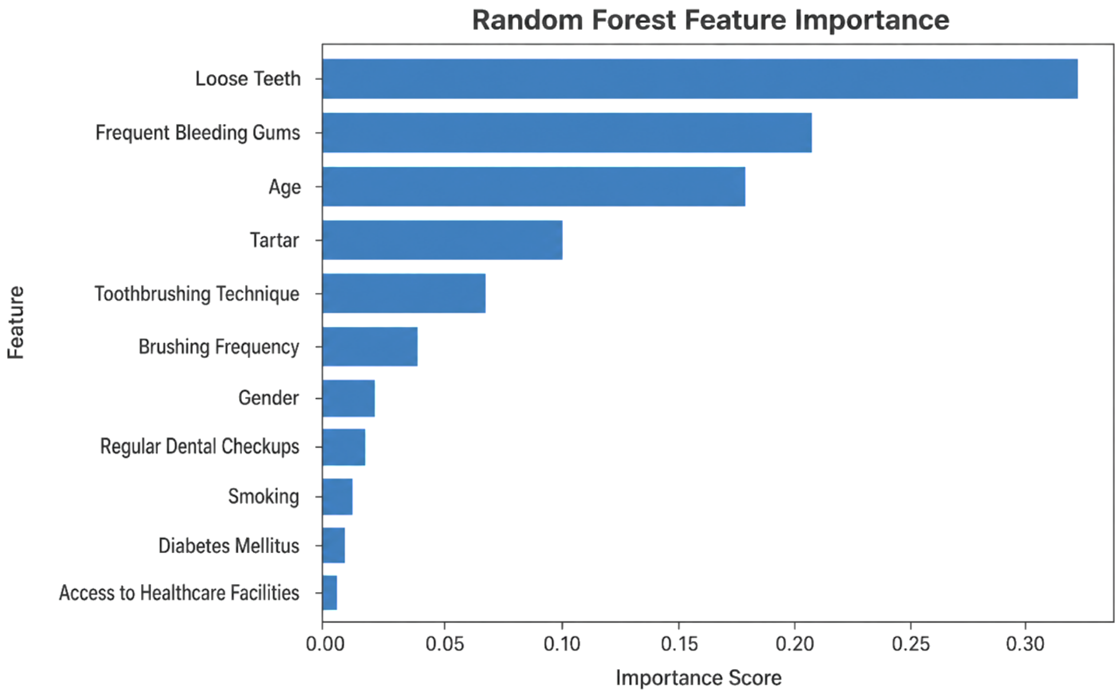

Figure 4. Random Forest Feature Importance

Figure 4 presents the feature-importance scores derived from the final Random Forest model. Among all evaluated predictors, tooth mobility (*gigi goyang*) emerged as the most influential variable (importance = 0.3247), followed by gingival bleeding (*gusi sering berdarah*) (0.2040), age (0.1811), and dental calculus (*karang gigi*) (0.0990). These variables collectively accounted for a substantial proportion of the model's predictive capability, indicating that clinically observable periodontal signs were the primary determinants of gingival health classification.Behavioral variables, including tooth-brushing technique (0.0670) and tooth-brushing frequency (0.0440), also contributed to prediction performance, although their influence was lower than that of the principal clinical indicators.

Although behavioral variables contributed less than the principal clinical indicators, their inclusion suggests that daily oral-health practices remain relevant predictors of gingival health status and may provide complementary information for screening applications. While feature-importance analysis identifies the relative contribution of individual predictors, it does not quantify the predictive value of broader feature groups. Therefore, an ablation analysis was conducted using the Random Forest model to evaluate the extent to which clinical, behavioral, and demographic-healthcare variables contributed to overall model performance. The results are presented in Table 7.

Table 7. Ablation Analysis Results

Feature Group	Accuracy (Mean ± SD)	Macro F1 (Mean ± SD)	AUC (Mean ± SD)
Clinical Only	0.912 ± 0.037	0.910 ± 0.036	0.960 ± 0.016
Behavioral Only	0.590 ± 0.032	0.516 ± 0.041	0.786 ± 0.017
Demographic + Healthcare	0.551 ± 0.066	0.503 ± 0.076	0.689 ± 0.027
Full Feature Set	0.902 ± 0.030	0.909 ± 0.022	0.966 ± 0.018

Interestingly, the full feature set achieved only a modest improvement in AUC (0.966) compared with the clinical-only model (0.960), while maintaining comparable classification performance. These findings indicate that routinely collected clinical indicators account for most of the predictive information used by the model. Nevertheless, the inclusion of behavioral, demographic, and healthcare-access variables may provide complementary information that improves model robustness and generalizability across different patient populations. The slight reduction in accuracy observed in the full-feature model may reflect the introduction of additional variability from less informative predictors, although these variables contributed to improved robustness and broader contextual representation of patient characteristics. Overall, the findings from the performance evaluation, feature-importance analysis, and ablation study consistently demonstrate the effectiveness of the proposed machine-learning framework for gingival health classification. Together, these analyses indicate that clinically relevant variables provide substantial predictive information and support the selection of Random Forest as the most suitable model for prototype implementation in this study.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

To demonstrate the practical applicability of the selected Random Forest model, the final trained model was integrated into a Streamlit-based prototype application designed to support gingival health screening and referral prioritization. The graphical user interface of the developed application is shown in Figure 5.

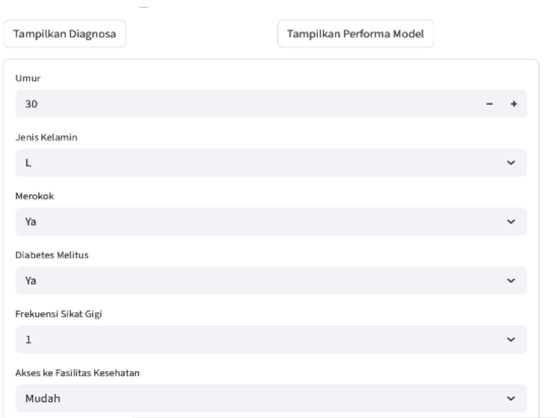

Figure 5. Streamlit-Based Interface for Gingival Health Screening

Figure 6. Example of Prediction Output for a New Patient

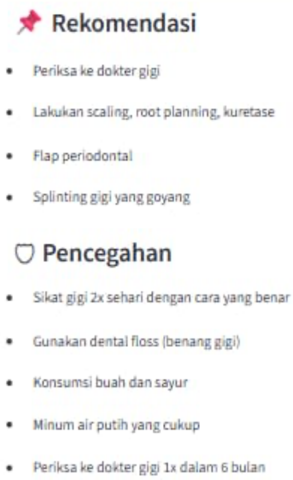

Figure 7. Recommendation and Prevention Features Based on Prediction Results

Once the data are submitted, the system utilizes the trained model to generate real-time screening outputs for gingival health classification (Normal Gingiva, Gingivitis, or Periodontitis) and provides supportive educational guidance associated with each predicted category. This implementation demonstrates the practical feasibility of integrating machine-learning-based decision support into routine periodontal screening workflows. Figure 6 presents an example of the prediction output generated by the

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Streamlit prototype for a new patient, illustrating how classification results and supporting information are communicated to end users.

As shown in Figure 6, the system classifies the submitted case as Periodontitis based on the provided clinical and behavioral information. This example demonstrates the ability of the proposed framework to transform routinely collected patient data into real-time screening outputs that support gingival health classification and referral prioritization. The generated classification should be interpreted as screening-support information and does not replace comprehensive periodontal examination, diagnosis, or treatment planning by a qualified dental professional.

Beyond generating classification results, the prototype also provides educational and preventive guidance tailored to the predicted gingival health category. These supporting outputs are intended to improve user awareness and encourage appropriate follow-up actions based on the screening result. An example of the recommendation and prevention module is presented in Figure 7. As illustrated in Figure 7, the application includes an educational guidance module that provides general information associated with each gingival health classification category. For cases classified as Periodontitis, the system presents educational information regarding commonly used periodontal care approaches and encourages consultation with qualified dental professionals for comprehensive clinical assessment and individualized treatment planning. The information provided by the application is intended solely for educational and screening-support purposes and should not be interpreted as a clinical diagnosis or treatment recommendation. This feature enhances user awareness while maintaining the central role of dental professionals in diagnosis, treatment planning, and patient management.

The educational recommendations and preventive guidance incorporated into the prototype were reviewed by a licensed dentist to ensure consistency with commonly accepted periodontal care practices. These recommendations are intended solely for educational and screening-support purposes and do not constitute individualized treatment advice. Final diagnosis, treatment planning, and patient management remain the responsibility of qualified dental professionals.

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Although the proposed prototype demonstrated promising classification performance and practical feasibility, it should be regarded as a preliminary screening-support tool rather than a system ready for routine healthcare deployment. Additional external validation, multicenter testing, usability evaluation, and prospective clinical studies are necessary to establish its generalizability, reliability, and real-world effectiveness before broader implementation can be considered.

3.2. Discussion

The comparative results indicate that no single model was uniformly superior across all evaluation criteria. Random Forest achieved the highest overall performance, obtaining the highest macro precision, macro recall, weighted precision, weighted recall, balanced accuracy, accuracy, and F1-score. These results suggest that Random Forest provided the most balanced classification performance across the three gingival health categories. In contrast, Logistic Regression achieved the highest ROC-AUC, indicating stronger discrimination capability across classification thresholds. SVM demonstrated competitive performance, particularly in identifying Normal Gingiva, while XGBoost produced stable and consistent results across multiple evaluation metrics. Collectively, these findings highlight the importance of considering model trade-offs rather than relying on a single performance indicator when selecting machine-learning models for periodontal screening applications.

The findings of the present study are consistent with previous machine-learning investigations in periodontal screening and oral-health classification. [26] reported that machine-learning approaches, including Random Forest and Logistic Regression, demonstrated strong capability for multiclass periodontal disease screening using routinely collected clinical variables. . Similarly, [27] identified Random Forest as the optimal classifier for periodontal health classification because of its robustness, resistance to overfitting, and ability to capture complex non-linear relationships among periodontal predictors. Despite differences in study populations, predictor sets, and disease definitions, the consistent superiority of ensemble-learning approaches across studies suggests that their capacity to integrate multiple clinical and behavioral risk factors provides a substantial advantage for periodontal screening applications. These observations support the findings of the present study and further strengthen the rationale for selecting Random Forest as the primary classification model.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Particular attention should be given to the Periodontitis category because it represents both the most clinically significant condition and the smallest class in the dataset (n = 74). In screening applications, recall is especially important because it reflects the model's ability to correctly identify patients who may require further periodontal assessment. Among the evaluated algorithms, Random Forest achieved the highest Periodontitis recall (93.2%) and the highest Periodontitis F1-score (0.9650), correctly identifying 69 of 74 Periodontitis cases. These findings suggest that Random Forest provides the most reliable balance between sensitivity and predictive accuracy for detecting individuals at increased risk of periodontal disease. Consequently, the model demonstrates strong potential for supporting early screening, referral prioritization, and preventive periodontal care. The strong performance of Random Forest may be attributable to its ability to handle heterogeneous clinical datasets and capture complex non-linear relationships among demographic, behavioral, and clinical predictors.

The superior performance of the Random Forest model, which achieved the highest accuracy (95.2%) and F1-score (0.9521), is likely attributable to its ensemble-based architecture. By combining predictions from multiple independently generated decision trees through a bagging strategy, Random Forest is well suited for handling heterogeneous tabular datasets containing both categorical and clinical variables. In this study, the model effectively utilized behavioral factors, such as smoking habits and tooth-brushing frequency, together with clinical indicators, including calculus and gingival bleeding. Compared with Logistic Regression (94.0% accuracy) and Support Vector Machine (94.2% accuracy), Random Forest is capable of modeling complex non-linear relationships and interactions among variables without relying on strong distributional assumptions. These characteristics contribute to greater model stability and improved generalization, particularly when classifying less represented categories such as Periodontitis within the 500-patient dataset.

In contrast, demographic and healthcare-access variables, including gender, regular dental check-ups, smoking status, diabetes mellitus, and access to healthcare facilities, exhibited relatively smaller importance scores within the present dataset. The prominence of tooth mobility, gingival bleeding, and dental calculus is clinically plausible because these variables are well-established indicators of periodontal tissue destruction and disease progression. Likewise, age is recognized as an important risk factor due to the cumulative effects of plaque accumulation, chronic inflammation, and periodontal attachment loss over time. The feature-importance results therefore support the clinical validity of the proposed machine-learning framework and suggest that the Random Forest model primarily relied on predictors that are biologically and clinically relevant to periodontal health assessment.

The feature-importance analysis also provides an explanation for the superior performance of Random Forest observed in this study. Because Random Forest can model complex non-linear interactions among multiple clinical and behavioral variables, it is capable of effectively integrating information from both direct periodontal indicators and supporting risk factors. This capability likely contributed to its higher accuracy and F1-score compared with the other evaluated algorithms.

From a screening perspective, the identified high-importance variables are routinely collected during basic periodontal examinations and can be assessed without advanced diagnostic equipment. This finding suggests that the proposed screening framework could be implemented in primary healthcare settings where comprehensive periodontal assessments may not always be available. By prioritizing clinically observable indicators such as tooth mobility, gingival bleeding, and dental calculus, the model may assist healthcare providers in identifying patients who require further periodontal evaluation and referral.

The ablation analysis further supports the feature-importance findings. Clinical variables alone achieved predictive performance comparable to that of the complete feature set, indicating that direct periodontal indicators constitute the primary source of information for gingival health classification. This finding is consistent with the feature-importance analysis, which identified tooth mobility, gingival bleeding, and dental calculus as the most influential predictors. Conversely, behavioral and demographic-healthcare variables exhibited substantially lower standalone predictive performance, suggesting that their primary role is to complement rather than replace clinical observations. These results reinforce the importance of incorporating routine periodontal examination findings into machine-learning-based screening systems.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

A limitation of this study is that the dataset was collected from a single healthcare center. Patient characteristics, oral-health behaviors, socioeconomic conditions, and disease prevalence may differ across regions and healthcare settings. Consequently, the generalizability of the proposed model to broader populations requires further validation using multicenter datasets. The class distribution may also have influenced model performance. Although stratified cross-validation was applied to preserve class proportions during model evaluation, the Periodontitis category represented the smallest class in the dataset (n = 74), whereas Gingivitis and Normal Gingiva contained substantially more observations. Models that are more robust to class imbalance, such as Random Forest, may therefore have gained an advantage in identifying minority-class cases.

Although stratified 5-fold cross-validation was employed to reduce the risk of overfitting and provide a robust internal evaluation framework, the absence of an independent external validation dataset may result in optimistic performance estimates. Consequently, the reported performance metrics should be interpreted as internal validation results. Further evaluation using external datasets collected from different institutions is required to confirm the generalizability and robustness of the proposed framework.

Future research should focus on prospective validation studies [28], multicenter testing, and larger population-based datasets to further assess model generalizability. Additional usability evaluations involving healthcare practitioners and end users are also warranted to examine workflow integration and practical adoption. Furthermore, fairness assessment across different demographic groups and probability calibration analysis may provide additional insights if future versions of the prototype incorporate confidence scores or probability-based decision support.

4. CONCLUSION

This study highlights the potential of multi-class machine learning models combined with routinely collected, non-invasive clinical and behavioral data for supporting early gingival health screening. Among the evaluated algorithms, Random Forest delivered the strongest overall performance, achieving an accuracy of 95.2%, an F1-score of 0.9521, and a macro-averaged AUC of 0.983 under a stratified 5-fold cross-validation scheme.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Furthermore, the successful deployment of the model within a Streamlit-based web application demonstrates its practicality as an accessible and cost-effective decision support tool for primary healthcare environments with limited resources. Nevertheless, the scope of the study is limited by the use of data from a single healthcare center. Therefore, additional validation is required before broader implementation can be recommended. Future research should focus on multicenter validation studies and prospective clinical assessments to evaluate model performance in diverse populations and to determine the real-world effectiveness of the deployed system in improving periodontal screening and patient care outcomes.

REFERENCES

[1] M. Hung *et al.*, "Development of a recommender system for dental care using machine learning," *SN Appl. Sci.*, vol. 1, no. 7, Jul. 2019, doi: 10.1007/s42452-019-0795-7.

[2] D. M. Alalharith *et al.*, "A deep learning-based approach for the detection of early signs of gingivitis in orthodontic patients using faster region-based convolutional neural networks," *Int. J. Environ. Res. Public Health*, vol. 17, no. 22, pp. 1–10, 2020, doi: 10.3390/ijerph17228447.

[3] P. N. Papapanou et al., "Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions," J, Periodontol., vol. 89, December 2017, pp. 173–182, 2018, doi: 10.1002/JPER.17-0721.

[4] J. G. Caton, G. Armitage, T. Berglundh, I. L. C. Chapple, S. Jepsen, and K. S. Kornman., "A new classification scheme for periodontal and peri-implant diseases and conditions – Introduction and key changes from the 1999 classification," *J. Periodontol.*, vol. 89, no. S1, Jun. 2018, doi: 10.1002/JPER.18-0157.

[5] P. Wahyuni, A. Rahardjo, and H. Novrinda, "Determinan Status Periodontal pada Remaja di Indonesia: Analisis data Riskesdas 2018 Determinants of Periodontal Status in Adolescents in Indonesia: Analysis of 2018 Riskesdas Data," *Cakradonya Dent. J.*, 2024.

[6] F. Mashaq, S. Herlina, and P. Latipa, "Expert System To Diagnose Periodontal Disease Using Certainty Factor Method Sistem Pakar untuk Mendiagnosa Penyakit Periodontal Menggunakan Metode Certainty Factor," *J. Komputer, Inf. dan Teknol.*, vol. 1, no. 1, pp. 134–140, 2021, doi: 10.53697/jkomitek.v1i1.

[7] Muhammad Hilmy Haidar Aly, "Klasifikasi Diabetes Menggunakan Algoritma Support Vector Machine Radial Basis Function," *J. Tek. Inform. dan Teknol. Inf.*, vol. 4, no. 1, pp. 28–38, Jan. 2024, doi: 10.55606/jutiti.v4i1.3420.

[8] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nat. Med.*, vol. 28, no. 1, pp. 31–38, Jan. 2022, doi: 10.1038/s41591-021-01614-0.

[9] M. A. Nazir, "Prevalence of periodontal disease , its association with systemic diseases and prevention," vol. 1, no. 2, 2017.

[10] B. Li, B. Aljabri, D. Beaton, L. Al-omran, M. A. Hussain, and D. S. Lee, "Predicting outcomes following endovascular aortoiliac revascularization using machine learning," 2025.

[11] A. A. Alshehri *et al.*, "A decision support system based on classification algorithms for the diagnosis of periodontal disease," *J. Med. Artif. Intell.*, vol. 8, no. 3, pp. 181–192, 2024, doi: 10.4103/sjoralsci.sjoralsci_50_24.

[12] C. A. Șalgău *et al.*, "Applications of Machine Learning in Periodontology and Implantology: A Comprehensive Review," Ann Biomed Eng 52, 2348–2371 (2024). https://doi.org/10.1007/s10439-024-03559-0.

[13] H. I. Fageeh *et al.*, "Accuracy in Diagnosing Periodontitis Using the AAP / EFP 2017 Classification," *Int. Dent. J.*, vol. 75, no. 6, p. 103907, 2025, doi: 10.1016/j.identj.2025.103907.

[14] E. Nurlia, M. Jajuli, and I. Purnamasari, "Penerapan Naïve Bayes Untuk Klasifikasi Tingkat Risiko Diagnosis Gigi Di Uptd Puskesmas Cingambul," *J. Inform dan Komputer*, vol. 4, no. 2, 2021, doi: 10.33387/jiko.

[15] N. H. Danial and D. Setiawati, "Convolutional Neural Network (Cnn) Based On Artificial Intelligence In Periodontal Diseases Diagnosis," *Interdental J. Kedokt. Gigi*, vol. 20, no. 1, pp. 139–148, Apr. 2024, doi: 10.46862/interdental.v20i1.8641.

[16] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.

[17] C. Z. V. Junus, T. Tarno, and P. Kartikasari, "Klasifikasi Menggunakan Metode Support Vector Machine Dan Random Forest Untuk Deteksi Awal Risiko Diabetes Melitus," *J. Gaussian*, vol. 11, no. 3, pp. 386–396, Jan. 2023, doi: 10.14710/j.gauss.11.3.386-396.

[18] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.

[19] L. A. Celi *et al.*, "Leveraging Data Science for Global Health", Springer Cham, 1st ed., 2020. [E-book] Available : Springer Nature Link.

[20] Y. Altork, "Comparative analysis of machine learning models for wind speed forecasting: Support vector machines, fine tree, and linear regression approaches," *Int. J. Thermofluids*, vol. 27, April 2025, p. 101217, 2025, doi: 10.1016/j.ijft.2025.101217.

[21] N. Bhuvaneswari, M. Sowmiya, A. K. Nandhana, and K. Preethi, "A High-Accuracy Heart Disease Prediction Model Using XG-Boost: A Machine Learning Approach for Clinical Decision Support," in *2026 IEEE Madhya Pradesh Section Conference (MPCON)*, IEEE, Mar. 2026, pp. 221–225. doi: 10.1109/MPCON69668.2026.11508271.

[22] P. Lestari, I. Tahyudin, A. Tikaningsih, and A. Nurhopipah, "Optimization of the XG-Boost Algorithm for Predicting Stroke Patient Care Outcomes," *Telematika*, vol. 18, no. 1, pp. 13–33, 2025, doi: 10.35671/telematika.v18i1.2817

[23] T. Denœux, "Logistic regression, neural networks and Dempster–Shafer theory: A new perspective," *Knowledge-Based Syst.*, vol. 176, pp. 54–67, Jul. 2019, doi: 10.1016/j.knosys.2019.03.030.

[24] S. B. D. Widyawati, P. Purwadi, and I. R. Yunita, "Comparative Analysis of Random Forest, Logistic Regression and SVM for Stunting Prediction Using Anthropometric Data," *J. Inf. Syst. Informatics*, vol. 7, no. 4, pp. 4271–4293, 2025, doi: 10.63158/journalisi.v7i4.1387.

[25] J. H. Lee, D. H. Kim, S. N. Jeong, and S. H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *J. Dent.*, vol. 77, July, pp. 106–111, 2018, doi: 10.1016/j.jdent.2018.07.015.

[26] K. Deng, F. Zonta, H. Yang, G. Pelekos, and M. S. Tonetti, "Development of a machine learning multiclass screening tool for periodontal health status based on non-clinical parameters and salivary biomarkers," *J. Clin. Periodontol.*, July 2023, pp. 1547–1560, 2023, doi: 10.1111/jcpe.13856.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[27] S. S. Yadadi, R. M. Shetty, and V. Sowmya, "Machine learning assisted classification of periodontal health and periodontitis using alveolar bone loss measurements on bitewing radiographs," May, 2026, doi: 10.3389/froh.2026.1814025.

[28] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement," *Circulation*, vol. 131, no. 2, pp. 211–219, 2015, doi: 10.1161/CIRCULATIONAHA.114.014508.