Non-Invasive Periodontal Disease Classification Using Thermography and Machine Learning: A Clinical Decision Support Approach

Antony Morales-Cervantes, *Member, IEEE*, Gerardo M. Chávez-Campos, *Member, IEEE*, Adriana C. Téllez-Anguiano, *Member, IEEE*, Ricardo Martínez-Parrales, *Member, IEEE*, Mayra Yunuen Rincón-Pineda and Francisco J. González

***Abstract*—Periodontal diseases such as gingivitis and periodontitis are common oral health conditions that require timely and accurate detection. This study presents a non-invasive diagnostic support system that integrates infrared thermography with clinical data to classify periodontal health status. A cross-sectional study was conducted on 91 individuals, categorized as healthy, with gingivitis, or with periodontitis. Thermographic images from three facial perspectives were analyzed to extract gingival temperature features, which were combined with clinical parameters including plaque index, age, sex, smoking status, and the presence of systemic conditions. Several machine learning models were evaluated using ten-fold cross-validation, both with and without dimensionality reduction. A two-step classification approach yielded the best results: logistic regression was used to identify periodontitis, followed by gradient boosting to differentiate between healthy and gingivitis cases. The combined model achieved an accuracy of 94.51% and an F1-score of 94.49%, while models based solely on thermographic data reached an accuracy of 75.82%. These findings support the feasibility of using thermal imaging and artificial intelligence to improve the classification of periodontal disease. The proposed method offers a promising non-invasive solution to enhance diagnostic accuracy and inform personalized dental care.**

***Index Terms*—Dental screening, non-invasive diagnosis, Machine Learning, periodontal disease, infrared thermography, gingivitis, periodontitis.**

Legend for Review Markup

- ~~old text~~ — text marked for removal or replacement.
- new text — suggested new or added content.
- Orange boxes — reviewer comments or todos, added via the **\marx** LaTeX command.

A. Morales-Cervantes was with the Faculty of Science, Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico, during the completion of this work. He is now with the Instituto Tecnológico de Morelia, Tecnológico Nacional de México, Morelia, Mich. 58120, Mexico.

G. M. Chávez-Campos*, A. C. Téllez-Anguiano, and R. Martínez-Parrales are with the Instituto Tecnológico de Morelia, Tecnológico Nacional de México, Morelia, Mich. 58120, Mexico (*e-mail: gmarx_cc@itmorelia.edu.mx)

F. J. González is with the Faculty of Science, Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico.

Todo List

TODO LIST

include more references to support refs 1
Toda la introducción fue ligeramente modificada para incluir una redacción más formal, pero lo marcado en azul en la introducción lo modifiqué más, revisa por favor!! 2
Revisa esta parte Antony 3
¿El PCA fue aplicado a todo? 3
¿? 4
fix table 5

I. Introduction

PERIODONTAL diseases constitute a major global public health concern, particularly among older adults, with prevalence rates ranging from 60% to over 70% depending on the population studied [1]. Periodontal diseases are not only a leading cause of tooth loss but are also associated with chronic oral pain and systemic health complications [2]. Gingivitis and periodontitis, the primary manifestations of periodontal disease, arise from complex interactions between host genetic susceptibility and subgingival microbial communities, which together influence both the onset and progression of disease [3]. Gingivitis is typically characterized by reversible inflammation of the gingival tissues, whereas periodontitis involves chronic inflammation and irreversible destruction of the supporting structures of the teeth [4], [5]. Although gingivitis can often be effectively managed in its early stages, unchecked progression to periodontitis poses a serious threat to oral function and significantly diminishes patients' quality of life [6].

include more references to support refs

Beyond their localized effects on oral structures, periodontal diseases have been linked to systemic inflammation and the involvement of distant organs, primarily through the release of proinflammatory cytokines and bacterial by-products into the bloodstream [7]. In addition to their local impact within the oral cavity, these diseases are associated with systemic complications such as impaired mastication, nutritional deficiencies, and increased healthcare costs [8], [9]. Conventional diagnostic methods, including periodontal probing and intraoral radiography, present notable limitations in detecting early-

stage or active inflammation [10]. These limitations highlight the importance of early and precise diagnostic methods to halt the disease sequence.

Periodontal inflammation is characterized by immune cell infiltration, blood vessel dilation, and metabolic alterations that affect local tissue function, leading to observable changes in surface temperature [11]. Inflammation, a central process in periodontal disease, typically manifests with clinical signs such as redness, swelling, pain, and elevated tissue temperature [5]. Within this context, infrared thermography (IRT) has emerged as a promising non-invasive tool for detecting thermal variations in periodontal tissues associated with inflammation [12]. Thermographic imaging enables the assessment of both the temperature of affected areas and adjacent tissues, offering a comprehensive view of the inflammatory state [13], [14]. While recent studies suggest a correlation between gingival temperature and inflammatory activity, the lack of standardized acquisition and analysis protocols has limited the clinical integration of infrared thermography in periodontal diagnostics [15]. To address this gap, this study proposes a reproducible and clinically oriented thermographic methodology that integrates both image-based features and clinical parameters. This approach aims to establish a foundation for reliable temperature-based biomarkers that can enhance clinical decision-making in the early detection and classification of periodontal disease.

Machine learning (ML) methods have shown considerable promise in analyzing complex biomedical data, including subtle patterns indicative of underlying biological dysfunction [16]. In the field of periodontology, ML has emerged as a transformative tool for disease classification based on clinical variables and imaging data. Numerous studies have demonstrated that algorithms such as decision trees, support vector machines (SVMs), and artificial neural networks (ANNs) can accurately predict dental diseases using structured clinical data [17]. Specifically for periodontitis, deep learning approaches have been applied to classify disease stages in dental images, yielding promising results in terms of diagnostic sensitivity and specificity [18]. Furthermore, integrating clinical variables with imaging data has been shown to enhance diagnostic performance, suggesting that multimodal hybrid models may overcome the limitations of conventional diagnostic techniques [19]–[21].

Since IRT provides a non-invasive way to record tissue thermal activity, integrating it with machine learning models could improve the detection of periodontal diseases by recognizing thermal patterns linked to gingival inflammation. While thermography has demonstrated diagnostic value in other dental fields, its limited application in periodontics underscores the need for research to evaluate its potential in identifying thermal markers of periodontal disease [20]. As a diagnostic tool, thermography may indirectly reflect key physiopathological processes, including vascular hyperemia and metabolic activation, both of which are characteristic of chronic inflammation [22]. The combination of thermal imaging and advanced artificial intelligence methods offers a valuable opportunity to enhance early detection of periodontitis and distinguish it more accurately from gingivitis.

This study introduces an innovative machine learning approach for classifying periodontal health using infrared thermography. Multiple machine learning models are assessed to determine their ability to distinguish between healthy patients, those with gingivitis, and those with periodontitis. A significant contribution of this research lies in the development and assessment of a two-phase classification strategy designed to enhance diagnostic precision in the identification of periodontal disease. Besides using thermal features, the model adds complementary clinical variables to boost its robustness and clinical relevance. This integrated approach not only improves performance but also demonstrates the potential of combining physiological imaging with contextual patient data to support reliable, non-invasive diagnosis.

Finally, a 10-fold cross-validation is conducted to evaluate the system's stability and its usefulness in real clinical settings. By combining a functional thermal biomarker with computational modeling, this research connects the fields of clinical dentistry, biomedical imaging, and physiological signal analysis. It aims to help develop diagnostic tools that not only enable early detection but also offer insight into the underlying inflammatory processes of periodontal disease.

Toda la introducción fue ligeramente modificada para incluir una redacción más formal, pero lo marcado en azul en la introducción lo modifiqué más, revisa por favor!!

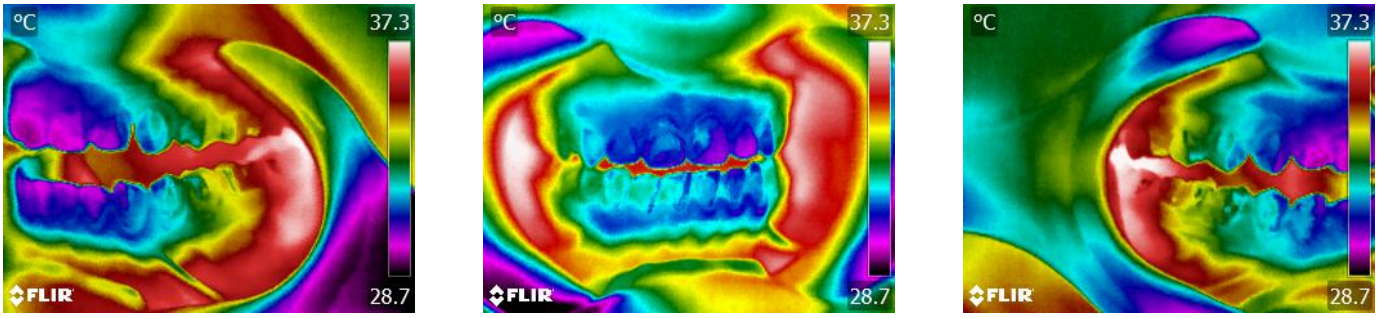
II. Materials and Methods

The study included a total of 91 participants, evenly divided into three groups: patients with periodontitis ($n = 30$), patients with gingivitis ($n = 30$), and a control group of healthy individuals ($n = 31$). Baseline characteristics of each group are presented in Table I. Each column group at the Table I includes mean age $\pm$ standard deviation (SD), number and percentage of male participants, smokers, and individuals with systemic diseases, as well as the O'Leary plaque index (mean $\pm$ SD). The participants' ages ranged from 18 to 60 years old, including both men and women, with dental structures from upper canines to lower canines. They were recruited from the clinic of the Master of Dental Sciences program at the "Universidad Autónoma de San Luis Potosí" and the periodontics clinic of the "Universidad Potosina". All participants signed an informed consent form, and the local ethics committee approved the study.

TABLE I: Baseline Characteristics of the Study Population.

Characteristics	N=30 Periodontitis	N=30 Gingivitis	N=31 Controls
Average age (years)	42.9 $\pm$ 10.3	26.0 $\pm$ 7.8	29.6 $\pm$ 6.9
Men (%)	16 (53%)	18 (54%)	19 (63.3%)
Smokers (%)	7 (23.3%)	5 (16.7%)	4 (13.3%)
Systemic disease (%)	11 (36.7%)	7 (23.3%)	0 (0%)
O'Leary index (%)	91.4 $\pm$ 9.7	85.8 $\pm$ 18.6	67.6 $\pm$ 15.4

Thermographic images were captured in a controlled environment with an average temperature of $23 \pm 3°C$, ensuring homogeneity in the temperature measurements. Images were obtained from three views: (a) left side, (b) front, and (c) right side for each patient, as shown in Figure 1. The images



(a) (b) (c)

Fig. 1: Thermographic images in three views per patient: (a) left side, (b) frontal, and (c) right side view.

were captured using a FLIR®T400 camera based on a focal plane array of microbolometers, with a thermal sensitivity of $< 0.05\,^{\circ}\mathrm{C}$ (50 mK), an accuracy of $\pm 2\%$, a spectral range of 7.5–13 μm, and a temperature range from -20 to $1200\,^{\circ}\mathrm{C}$. Emissivity was calibrated to 0.97 [23], and the camera distance was 0.3 m. To ensure measurement reliability, restrictions were set before image capture, forbidding alcohol, tobacco, and physical activity for at least 24 hours prior to the camera session.

Revisa esta parte Antony

Thermographic analysis was conducted by segmenting *regions of interest* (ROIs) in the gingival area to obtain representative temperature values for each region across the different captured views. Thermal measurements were manually extracted by an expert periodontist using the FLIR Tools software. The analysis specifically targeted the incisors and canines in all four dental quadrants, considering three distinct areas per tooth: mesial, medial, and distal.

Figure 2 illustrates an example of the temperature measurements obtained from a lower left canine. In this image, the markers indicate the specific ROIs: **Sp1** corresponds to the mesial region, **Sp2** to the medial region, and **Sp3** to the distal region.

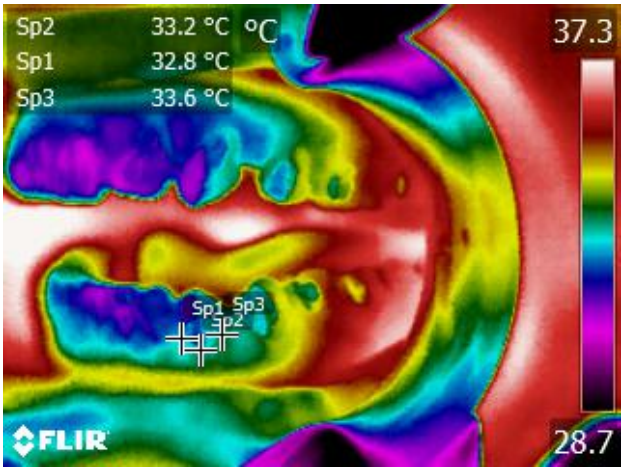


Fig. 2: Thermographic measurement on lower left canine: mesial, medial, and distal regions identified.

In addition to the thermographic analysis, clinical variables such as sex, age, O'Leary plaque index, Greene and Vermillion oral hygiene index, smoking status, number of cigarettes consumed, presence of systemic disease, dental crowding, and presence of dental calculus were also included. The combination of thermal and clinical features enabled the evaluation of the robustness of the classification model, comparing the performance of the analysis based solely on thermography with that of the model enriched with clinical data.

A. Evaluation of Machine Learning Models

Multiple ML models were evaluated to determine the most effective approach for classifying subjects as healthy, with gingivitis, or with periodontitis. The models tested included Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, Random Forests, Support Vector Machines (SVM), Neural Networks, XGBoost, and LightGBM. Due to the high dimensionality of the thermographic image features, Principal Component Analysis (PCA) was applied for dimensionality reduction. Two classification strategies were evaluated to assess the diagnostic potential of thermographic imaging: the first relied solely on thermographic features of the dental organs (DO) dataset, while the second dataset combined these features with additional clinical variables (DO + Clinics), as shown in Figure 3.

Principal Component Analysis (PCA) was applied to reduce feature dimensionality in both datasets: the one containing only thermographic features, and the one combining thermographic features with clinical variables, followed by the training and evaluation of various machine learning models using a 10-fold cross-validation approach. Performance metrics—including mean accuracy, AUC, F1-score, precision, recall, and confusion matrices—were computed and summarized in a comparative table to highlight the impact of including clinical data.

¿El PCA fue aplicado a todo?

All machine learning models and statistical analyses were implemented using Python (version 3.10.14) with the following modules: Scikit-learn, XGBoost, NumPy, Pandas, Matplotlib, and Seaborn.

B. ROC Curve Analysis and Two-Phase Classifier Decision

After identifying the best-performing models, a detailed analysis of the ROC curves in a One-vs-Rest (OvR) scheme

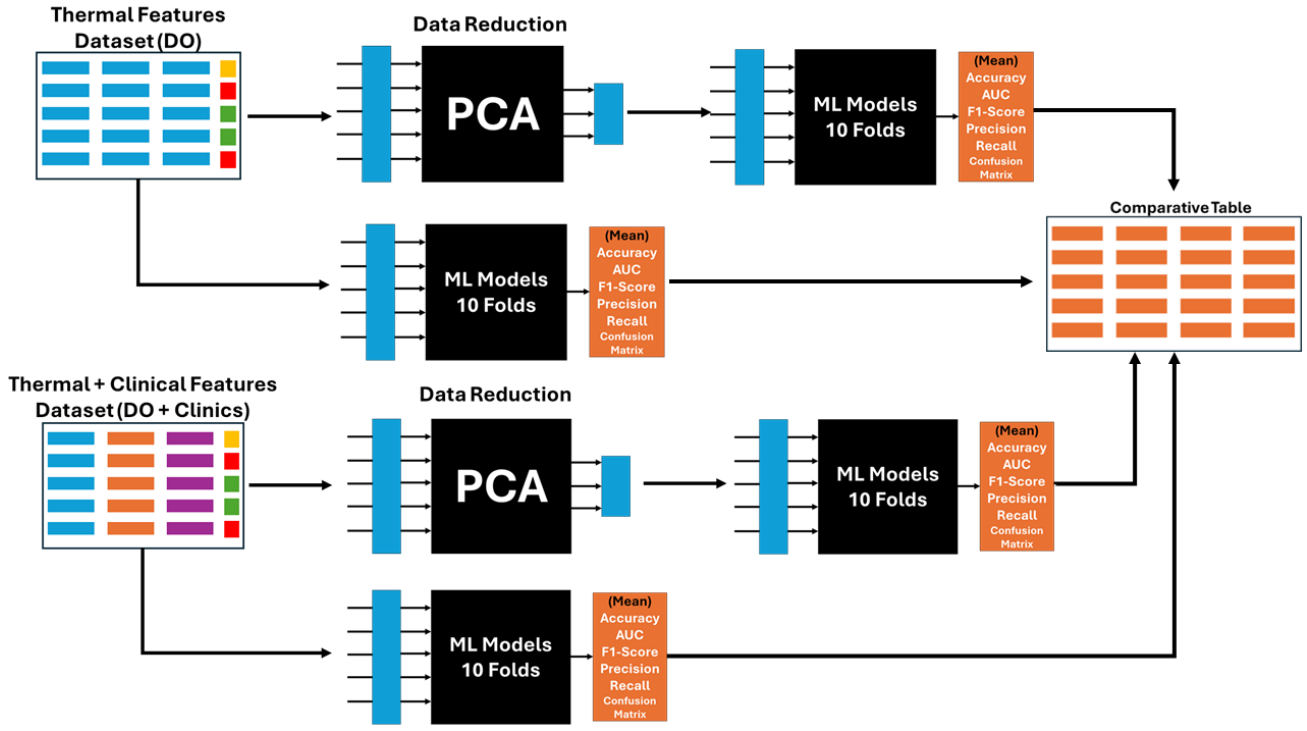


Fig. 3: Model evaluation with PCA and cross-validation using thermographic and clinical features for classification.

was performed to assess each model's ability to distinguish among the three classes. The ROC curve is a conceptual tool that plots the *true positive* rate against the *false positive* rate at various classification thresholds, providing insight into the model's discriminative performance. This analysis led to the development of a two-phase classification strategy. In Phase 1, Logistic Regression with a threshold optimized via the Youden index was used to identify periodontitis cases. In Phase 2, XGBoost was used to distinguish between gingivitis and healthy subjects. This sequential approach improved overall model precision and enhanced classification reliability in clinical settings. The entire workflow is illustrated in Figure 4.

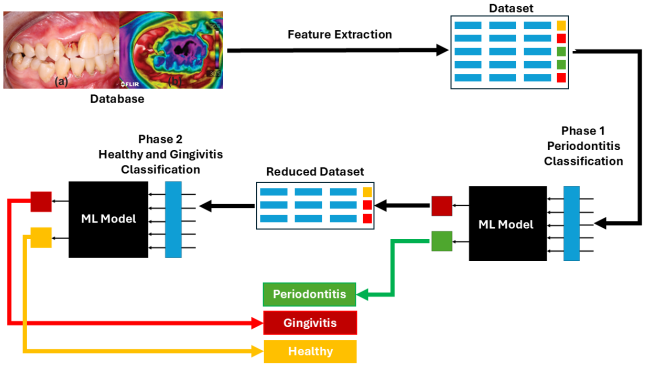


Fig. 4: Two-phase classification model: first detects periodontitis, then classifies gingivitis and healthy cases.

The final performance of the two-phase classification model was assessed using 10-fold stratified cross-validation; see Figure 5. Therefore, the dataset was divided into 10 balanced folds, with each iteration training on 90% of the data and testing on the remaining 10%. This process was repeated 10 times, and performance metrics (accuracy, AUC, F1-score, precision, recall, and confusion matrix) were averaged to assess model stability and reliability.

This approach preserves the class distribution (healthy, gingivitis, and periodontitis) across all folds, ensuring a balanced representation in both training and testing sets. ~~Stratification helps prevent bias from class imbalance and enables a more reliable evaluation of model stability across different data subsets.~~ ¿?

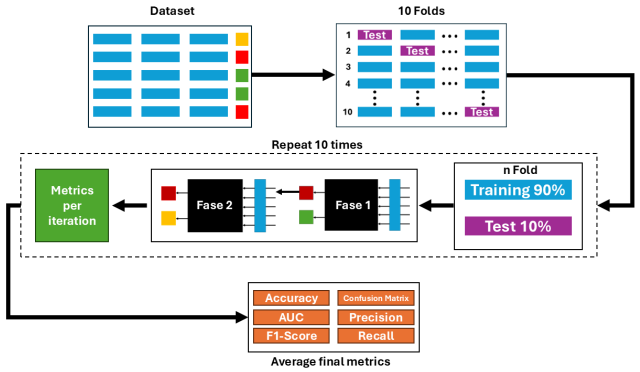


Fig. 5: Scheme of 10-fold cross-validation used to evaluate the two-phase periodontal classification system.

The metrics used to evaluate the model's performance include accuracy, area under the ROC curve (AUC), F1-score, precision, recall, and the confusion matrix. These metrics enable the analysis of various aspects of the model's performance

in classifying periodontitis, gingivitis, and periodontal health, and are defined and described in detail below.

Accuracy measures the proportion of correct predictions over the total number of samples and is defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

TP: True positives.
TN: True negatives.
FP: False positives.
FN: False negatives.

Recall indicates the model's ability to correctly identify positive cases, calculated as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

Precision measures the proportion of cases correctly classified as positive among all positive predictions made by the model:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

F1-score is the harmonic mean between precision and recall, providing a balance between the two metrics:

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Area under the ROC curve (AUC-ROC) measures the model's ability to discriminate between classes. It is calculated as the area under the curve that relates the true positive rate (*recall*) with the false positive rate (*1 - specificity*).

Finally, the **confusion matrix** for the multiclass classification is represented as:

$$\mathbf{M} = \begin{bmatrix} TP_h & FN_{h\to g} & FN_{h\to p} \\ FP_{g\to h} & TP_g & FN_{g\to p} \\ FP_{p\to h} & FP_{p\to g} & TP_p \end{bmatrix} \tag{5}$$

where:

- TP_h, TP_g, TP_p: True positives for healthy, gingivitis, and periodontitis.
- $FN_{h\to g}$, $FN_{h\to p}$: Healthy patients misclassified as gingivitis or periodontitis.
- $FN_{g\to p}$: Gingivitis patients misclassified as periodontitis.
- $FP_{g\to h}$, $FP_{p\to h}$: Gingivitis or periodontitis predicted as healthy.
- $FP_{p\to g}$: Periodontitis misclassified as gingivitis.

These metrics offer a comprehensive evaluation of the model, allowing for analysis of its performance in terms of overall accuracy, discriminative ability, and the balance between precision and recall in multiclass classification.

III. Results

This section presents the performance of multiple machine learning algorithms evaluated for classifying periodontal health status by comparing thermal and clinical features across different setups. To identify the most effective model for distinguishing between periodontitis, gingivitis, and healthy subjects, various machine learning algorithms were evaluated using 10-fold cross-validation. Models including Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Support Vector Machines (SVM), Artificial Neural Networks (ANN), XGBoost, and LightGBM were compared across four input configurations: (i) thermographic features of dental organs with dimensionality reduction via PCA; (ii) thermographic features without dimensionality reduction; (iii) a combination of thermographic and clinical features with PCA; and (iv) a combination of thermographic and clinical features without PCA.

Table II shows the average accuracy and area under the ROC curve (AUC) obtained for each model across the different configurations evaluated.

fix table

TABLE II: Performance of models using thermal and clinical features across configurations.

	(i) PCA (IRT)		(ii) **IRT only**	
Model	Accuracy	AUC	Accuracy	AUC
Logistic Regression	0.684	0.791	0.632	0.741
KNN	0.632	0.765	0.579	0.704
Decision Tree	0.421	0.516	0.579	0.672
Random Forest	0.789	0.799	0.632	0.750
SVM	0.789	0.851	0.579	0.793
ANN	0.526	0.634	0.526	0.780
XGBoost	0.842	0.897	0.684	0.785
LightGBM	0.789	0.850	0.684	0.775
	(iii) PCA (IRT + Clinical)		(iv) **IRT + Clinical**	
Model	Accuracy	AUC	Accuracy	AUC
Logistic Regression	0.737	0.770	0.834	0.874
KNN	0.684	0.744	0.718	0.859
Decision Tree	0.474	0.604	0.777	0.878
Random Forest	0.684	0.832	0.777	0.915
SVM	0.632	0.829	0.764	0.904
ANN	0.684	0.752	0.820	0.818
XGBoost	0.632	0.742	0.850	0.924
LightGBM	0.684	0.728	0.821	0.897

The results show that models trained only with thermographic IRT features had more inconsistent performance, with Accuracy ranging from 0.42 to 0.84 depending on the model. It was observed that using PCA for dimensionality reduction did not always enhance performance and, in some cases, reduced discrimination ability. In contrast, adding clinical variables improved model stability and boosted classification metrics across all configurations.

Figure 6 shows the distribution of classification accuracy for different machine learning models evaluated with the combined thermal and clinical features (IRT + Clinical) under 10-fold cross-validation, using the original feature set without PCA. While several models demonstrate acceptable performance, Logistic Regression and XGBoost stand out for their high median accuracy and consistent performance across folds. Specifically, Logistic Regression achieved the highest median accuracy of 0.86, whereas XGBoost exhibited the most compact distribution with fewer outliers, indicating greater stability. These results support the selection of both models as key components in the proposed two-phase classification framework.

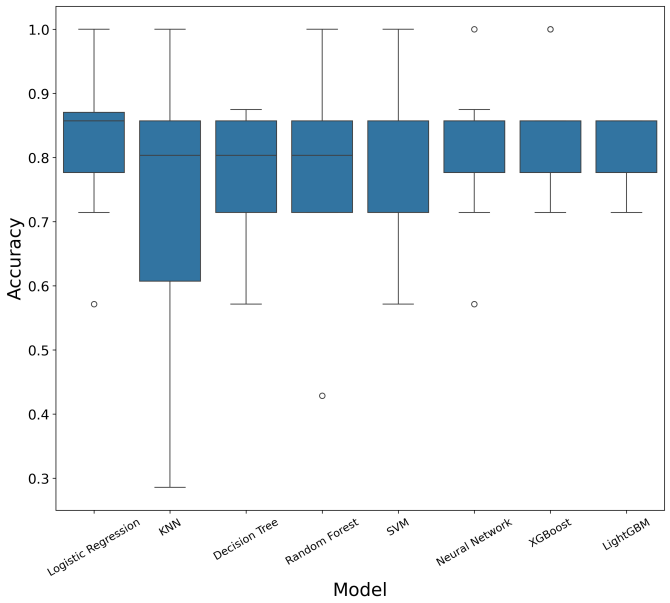


Fig. 6: Accuracy distribution across models with Thermal + Clinical Features (without PCA) in 10-fold cross-validation.

To assess class-specific performance, ROC curves were generated using a One-vs-Rest (OvR) strategy, in which a separate binary classifier is trained for each class by treating it as the positive class and all others as negative. This allows for evaluating the model's ability to distinguish each diagnostic category independently. Logistic Regression achieved the highest AUC (0.874) for periodontitis detection (Table II), while XGBoost showed the most stable overall performance across folds (Figure 6). These results triggered the implementation of a two-phase classification strategy: Logistic Regression was applied first to detect periodontitis cases, followed by XGBoost to classify the remaining subjects as gingivitis or healthy. This approach aimed to maximize diagnostic sensitivity while minimizing false positives.

Figure 7 shows the ROC curves produced by the Logistic Regression model using the complete feature set; (iv) IRT + clinical. The model achieved excellent class separation, with AUC values of 1.00 for healthy, 0.90 for gingivitis, and 0.87 for periodontitis, supporting its selection for the initial phase of the classification strategy.

A. Evaluation of the Two-Phase Classifier

To evaluate the performance of the proposed two-phase classification strategy, the model was first implemented using all available variables (IRT and clinical features). Ten-fold cross-validation was applied to assess its stability and generalization capacity. As shown in Table III, the classifier achieved an average accuracy of 94.51%, F1-score of 0.9449, recall of 0.9451, and precision of 0.9478. The corresponding confusion matrix is presented in Figure 8a, which shows that the classifier accurately identifies patients with periodontitis and effectively differentiates them from those with gingivitis and healthy subjects.

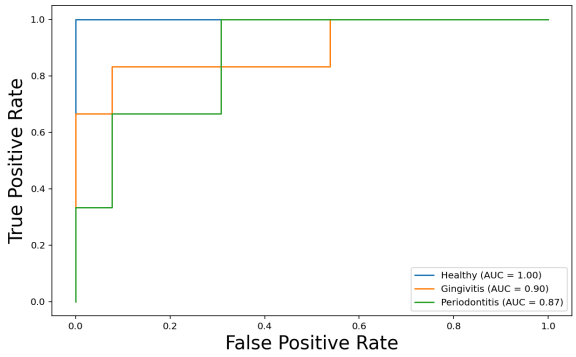


Fig. 7: Multiclass ROC curves (OvR) of Logistic Regression with IRT + Clinical Features.

Figure 8a shows the confusion matrix of the two-phase classifier when using both thermal and clinical features. The model demonstrates high classification accuracy across all classes: Healthy subjects were perfectly identified (31 out of 31); Gingivitis cases were mainly correctly classified (26 out of 30), with four misclassified as periodontitis; and Periodontitis cases were also well classified (29 out of 30), with only one misclassified as gingivitis. This matrix highlights the model's strong discriminative power when both types of features are used, particularly in separating healthy individuals and periodontitis cases.

TABLE III: Average results of the two-phase classifier using 10-fold cross-validation with and without clinical features.

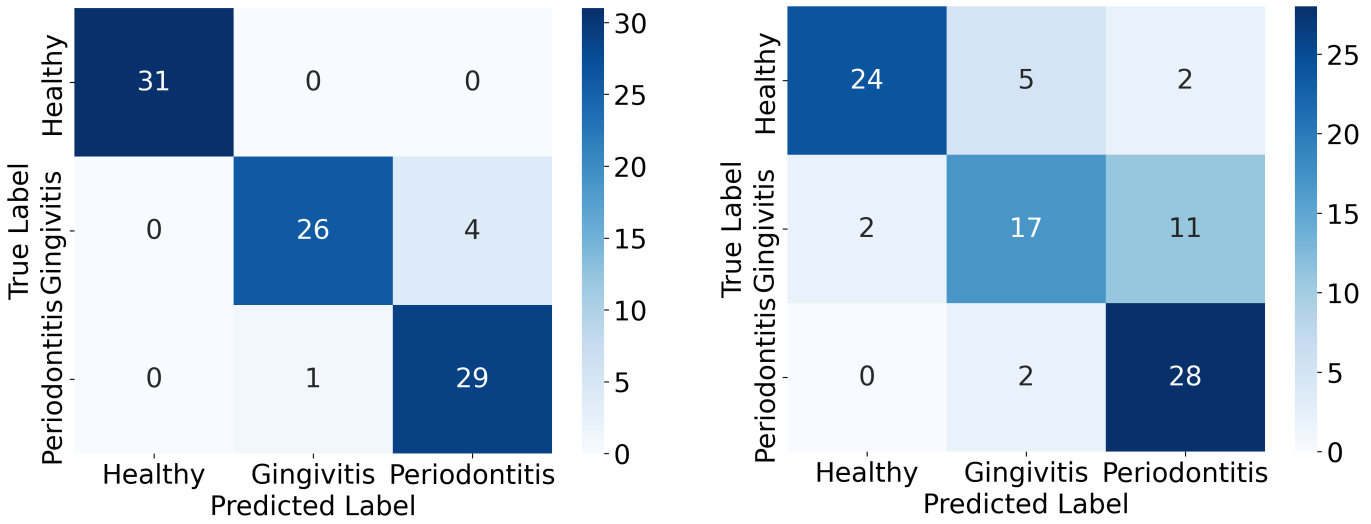
Metric	Thermal + Clinical Features	Thermal Features Only
Accuracy	0.9451	0.7582
F1-score	0.9449	0.7535
Recall	0.9451	0.7581
Precision	0.9478	0.7714

To further assess the robustness of the model, Figure 9 presents the evolution of performance across the 10 folds of cross-validation. Results indicate stable classification across most iterations, with minimal variability in one case.

After confirming the model's robustness using the complete feature set, we also evaluated its performance using only thermal features to explore the standalone diagnostic potential of infrared thermography. As reported in Table III, this configuration resulted in a decreased performance: accuracy of 75.82%, F1-score of 0.7535, recall of 0.7581, and precision of 0.7714. The corresponding confusion matrix is shown in Figure 8b, revealing notable challenges in distinguishing between gingivitis and periodontitis when clinical variables are excluded.

B. Discussion

To contextualize the findings, Table IV summarizes recent relevant studies. Overall, emerging technologies such as infrared thermography (IRT) and ML algorithms have shown promise as complementary tools to traditional methods of periodontal diagnosis. For example, Gunupati et al. [15] and



(a) (b)

Fig. 8: Confusion matrices of the two-phase classifier: (a) with clinical data, (b) without.

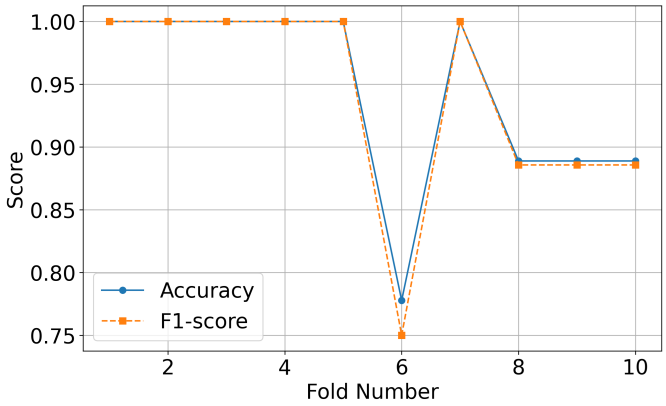


Fig. 9: Performance across 10 folds of the two-phase classifier with thermal and clinical features.

Rams and Slots [24] investigated gingival temperature as an indicator of inflammation. While Gunupati reported only a moderate correlation between temperature and disease activity, Rams found that periodontal pockets with elevated temperatures harbored significantly more pathogens, supporting the potential of temperature as a biomarker of active disease. These findings suggest that IRT may detect subtle thermal changes associated with periodontal inflammation that are not captured through conventional visual examination.

IRT has already demonstrated value in various areas of dentistry, including the diagnosis of temporomandibular disorders [25], lupus erythematosus [26], as well as applications in implantology and endodontics [27]. However, its use in identifying periodontal diseases such as periodontitis remains limited. Although some studies have reported its utility in detecting thermal differences in periapical inflammatory lesions [28], systematic evaluation of its diagnostic potential in the periodontal context, especially when combined with computational approaches, has been scarce.

More recently, machine learning models have been applied to the diagnosis of periodontitis using invasive clinical data and digital radiographs, often involving exposure to ionizing radiation [29], [30]. In contrast, the strategy proposed in this study employs a non-invasive approach by combining IRT with clinical variables, achieving an accuracy of 94.51% in classifying subjects with periodontitis, gingivitis, and healthy individuals. This aligns with previous findings and reinforces the validity of gingival surface temperature as a complementary biomarker. However, consistent with prior studies, our results also indicate that it should not replace traditional clinical diagnostic criteria.

As shown in the last row of Table IV, the proposed model integrates key elements identified in previous studies, combining clinical variables with emerging technologies. A multifactorial approach has been supported by the findings of Deng et al. and Beak et al., which suggest that the inclusion of complementary data, such as salivary biomarkers or systemic risk factors, enhances diagnostic accuracy compared to relying solely on a single source of information [31], [33]. In the present study, gingival temperature, clinical data, and machine learning algorithms were integrated, resulting in high diagnostic performance (ACC = 94.51%, AUC = 0.98), comparable to other recent AI-based methods (e.g., ACC = 89% in Deng et al. [31]; ACC = 92.9% in Shon et al. [32]).

Furthermore, unlike traditional visual assessments or intraoral radiography, which are subject to observer variability or involve exposure to ionizing radiation, a non-invasive strategy based on infrared thermography was employed. This methodological distinction gains relevance in light of the systematic analysis conducted by Radha et al. [30], which reported that most machine learning applications in periodontology rely on invasive clinical data or radiographic imaging. Consequently, the findings presented in this work highlight the need for safer and more accessible diagnostic alternatives.

The robust performance of the proposed model, reflected in

TABLE IV: Comparison of previous studies on conventional periodontal diagnosis vs. new technologies.

Reference (Year)	**Method / Main Metric**	**Sample**	**Key Findings**
Gunupati *et al.* (2019) [15]	Gingival Surface Temperature (GST) by thermometry	$n = 50$	AUC = 0.61 for gingival temperature; sensitivity of 72%, specificity of 64%, with no statistically significant differences.
Rams and Slots (2024) [24]	Subgingival thermometer + bacterial culture	$n = 8$ (32 sites)	Significant difference between *P. gingivalis* and elevated subgingival temperature (p = 0.030), and between red/orange complex pathogens and elevated subgingival temperature (p = 0.012)
Aboushady *et al.* (2021) [28]	Infrared thermography	$n = 80$	Significantly higher mean temperature in acute abscesses compared to chronic lesions and controls ($p < 0.001$); strong agreement with clinical diagnosis ($\kappa = 0.97$).
Bashir *et al.* (2022) [29]	10 ML models using clinical and demographic data	$n = 7{,}104$ (train), $n = 2{,}023$ (external test)	Internal AUC >0.95, accuracy = 96.7%; external AUC = 0.78, accuracy = 65.2%.
Deng *et al.* (2024) [31]	RF-based ML using salivary biomarkers and clinical questionnaire	$n = 408$	Multiclass classification into 3 and 6 periodontal health stages. Based on confusion matrices, global accuracy was $\sim$89.8% (3 classes) and $\sim$80.5% (6 classes).
Shon *et al.* (2022) [32]	Deep learning (U-Net + YOLOv5) on panoramic X-rays	140 teeth (10 images)	Accuracy = 0.929; mean recall = 0.805, mean precision = 0.732, mean F1-Score = 0.696.
Beak *et al.* (2024) [33]	5 ML models; 16 clinical risk factors selected by correlation	$n = 2229$	Best performance: AUC = 0.828, accuracy = 0.786, sensitivity = 0.394, specificity = 0.924, precision = 0.646
Chang *et al.* (2022) [34]	Deep learning (InceptionV3 multitask) on periapical radiographs	$n = 236$	Accuracy = 0.87, Sensitivity = 0.86, Specificity = 0.88
This work	Infrared thermography and clinical variables; 2-phase ML classifier (LR + XGBoost)	$n = 91$	Combined model: AUC = 0.98, ACC = 94.51%, F1 = 0.9449, Recall = 0.9451, Precision = 0.9478 (10-fold CV); outperformed previous studies in 3-class classification. Thermal-only model: AUC = 0.82, ACC = 75.82%, F1 = 0.7535, Recall = 0.7581, Precision = 0.7714

its high accuracy (ACC) and area under the curve (AUC), can be attributed to several design decisions grounded in existing literature. First, rigorous k-fold cross-validation and hyperparameter tuning were applied to minimize the risk of overfitting. This approach is critical, as prior studies have cautioned that models evaluated solely on the training cohort may overestimate their predictive capabilities [29]. For instance, Bashir et al. [29] reported marked reductions in accuracy when their models were tested on external populations, underscoring the necessity of independent validation. In the present study, the inclusion of both internal cross-validation and an independent test set enabled a more reliable estimation of generalization capacity.

Second, feature selection was guided by clinical relevance, incorporating variables such as systemic disease, temperature, and demographic data. This strategy aimed to reduce noise and improve model robustness, as recommended by Bashir et al. [29] and Beak et al. [33], who demonstrated that pruning irrelevant input variables enhances both predictive power and model stability. Furthermore, the integration of infrared thermography introduced a novel, physiologically grounded dimension. Given that active inflammation induces detectable thermal variations [15], [24], the model was capable of identifying subclinical gingival inflammation that traditional diagnostic methods may overlook.

This synergistic combination of multimodal inputs contributed to the model's high predictive performance, which in several instances surpassed that of approaches relying on single diagnostic modalities such as radiographic imaging or clinical examination alone. Overall, the findings support the hypothesis that a well-calibrated hybrid model, designed with methodological rigor and informed variable selection, can achieve an optimal balance between sensitivity and specificity, thereby justifying the observed results.

The findings of this study carry several practical implications for clinical practice and public health. Firstly, the incorporation of gingival infrared thermography into periodontal examinations may provide clinicians with a non-invasive adjunctive tool for detecting active inflammation. Unlike conventional radiography, thermography does not involve ionizing radiation, rendering it safer for repeated use and more suitable for vulnerable populations. This advantage suggests potential applications in community-based periodontal screening programs or tele-diagnosis frameworks, where rapid thermal imaging could be used to identify individuals who require further evaluation.

Additionally, the developed AI model demonstrated the capacity to automatically stratify periodontal disease severity, effectively distinguishing between gingivitis, incipient periodontitis, and advanced periodontitis. In clinical settings, this capability could be leveraged to prioritize treatment, enabling timely care for patients at greater risk and thereby optimizing the use of healthcare resources. By facilitating early detection and risk-based stratification, the proposed system may contribute to preventive strategies that reduce the overall burden of periodontal disease and its systemic consequences.

Notably, the multiclass classification accuracy achieved (ACC = 94.51%) is comparable to that reported for experienced periodontists conducting comprehensive clinical assessments [31], [32]. Thus, rather than replacing clinical judgment,

the AI system is intended to support and standardize diagnostic procedures, mitigating subjectivity and reducing dependence on individual expertise.

A further implication concerns therapeutic monitoring. As subgingival temperature has been associated with microbial load and inflammatory activity [24], serial thermographic measurements could be used to assess treatment response, such as reductions in localized "hot spots" following initial periodontal therapy. In summary, the integration of infrared thermography and artificial intelligence into the diagnostic workflow has the potential to enhance early detection, guide timely interventions, and improve long-term outcomes, including tooth preservation and control of associated systemic inflammation.

C. Limitations and Future Work

Despite the promising results, several inherent limitations of this study should be acknowledged. First, the sample size and geographic scope were limited. Although a sufficient number of subjects were included to demonstrate statistically significant differences, data were collected from a single center, which may restrict the generalizability of the model to populations with differing demographic profiles or disease patterns. As emphasized in previous research, the development of robust AI models in periodontology requires large, diverse cohorts [29]. Therefore, future work should involve external validation in independent populations, ideally through international, multicenter studies.

Second, the spatial resolution of infrared thermography may pose challenges in patients with dental crowding or fragile gingival tissues. In such cases, thermal signal overlap may hinder the precise attribution of temperature readings to specific periodontal sites. Additionally, limitations related to the machine learning model itself must be considered. While internal k-fold cross-validation was employed to mitigate overfitting, a risk remains that the model may be partially tailored to the specific characteristics of the training dataset. Although high accuracy was achieved on the local test set, identical performance cannot be assumed in other populations without recalibration. A similar limitation was observed by Beak et al., where a model trained in one clinical setting showed a marked decrease in performance when applied externally (AUC reduced to 0.65) [33].

Finally, the cross-sectional design of the study precludes any assessment of temporal causality. Specifically, the observation of elevated gingival temperature in association with periodontitis does not confirm whether such elevations precede and predict future tissue destruction. Longitudinal follow-up studies would be necessary to determine whether sites with elevated thermal readings are more likely to exhibit future attachment loss. Incorporating this dimension would enhance the biomarker's prognostic value and contribute to its clinical utility.

Due to the limitations mentioned earlier, several directions for future research are proposed. First, the implementation of multicenter clinical trials involving larger and more heterogeneous samples is recommended to refine and externally validate the proposed model. The inclusion of participants from diverse age groups, ethnic backgrounds, and systemic health conditions would enable a more comprehensive evaluation of the model's robustness in the face of biological variability. Additionally, direct performance comparisons between the AI system and experienced clinicians would be of significant value. For instance, concordance studies in which AI-generated diagnoses are compared with those of multiple calibrated periodontists, along with assessments of diagnostic time and inter-observer variability, could provide insights into clinical applicability and reliability.

Second, integrating additional data modalities may further enhance the model's diagnostic accuracy. Future studies could explore the combination of thermographic and clinical data with intraoral 3D imaging, subgingival microbiome profiles, or host genetic markers. As noted in the review by Shon et al. [32], the fusion of imaging and clinical information has been shown to significantly improve diagnostic performance, a finding that aligns with the present study's results.

Third, enhancing the interpretability of AI models will be crucial for their adoption in clinical settings. The development of explainable AI tools-such as visualization techniques that highlight the regions or variables contributing most to a given prediction-could foster greater trust among clinicians. For example, heat maps superimposed on gingival images may help identify areas of elevated risk, offering visual support for the AI's diagnostic output.

Finally, a critical avenue for future investigation involves assessing the real-world clinical impact of these technologies. Research is needed to determine whether large-scale implementation of AI-assisted thermographic screening improves population-level periodontal health outcomes or whether home-based monitoring of gingival temperature with portable devices can reduce the progression from gingivitis to periodontitis. Addressing these questions will require prospective interventional studies. Such efforts would not only enhance diagnostic accuracy but also support the development of personalized periodontal care, aligning with the principles of precision medicine.

IV. Conclusions

This study demonstrated the feasibility and diagnostic potential of infrared thermography as a complementary, non-invasive tool for detecting and classifying periodontal diseases. By integrating thermal imaging with clinically relevant variables, a machine learning model was developed that achieved a high classification accuracy (ACC = 94.51%) in distinguishing among healthy subjects, gingivitis, and periodontitis. These findings highlight the importance of integrating physiological imaging with computational intelligence to improve diagnostic accuracy in periodontal evaluation.

The implementation of a two-phase classifier further improved the discrimination between disease categories, illustrating the importance of adaptive classification strategies when dealing with conditions characterized by subtle thermal and clinical features. The findings add to the increasing evidence supporting artificial intelligence in digital dentistry, especially

in improving diagnostic methods that are safe, affordable, and scalable.

Notably, the study indicates that thermographic patterns of the gingiva could serve as viable biomarkers for early inflammatory activity, offering an alternative to traditional diagnostic methods such as probing and radiography, which are either invasive or rely on ionizing radiation. The integration of thermal and clinical data into machine learning algorithms has the potential to streamline screening processes, reduce subjectivity, and support early intervention strategies.

Nonetheless, the generalizability and clinical utility of the proposed model require further validation. Future research should involve larger, multicenter, and demographically diverse cohorts to assess the robustness of the system across different populations and settings. The influence of external factors on thermal image acquisition-such as ambient conditions, anatomical variability, and patient positioning-should also be systematically investigated to enhance model reliability and reproducibility.

In addition, efforts should be directed toward refining image preprocessing and thermal normalization techniques, exploring multimodal data integration (e.g., 3D imaging, salivary biomarkers, or genetic profiles), and improving model interpretability to facilitate clinical adoption. Longitudinal studies will be essential for evaluating the prognostic value of gingival thermal patterns and their correlation with disease progression and treatment outcomes.

In conclusion, this research represents a significant step forward in applying artificial intelligence to oral health, demonstrating that analyzing gingival surface temperature using machine learning can provide valuable insights into periodontal health. The integration of such technologies into routine dental workflows holds promise for enabling earlier, safer, and more personalized diagnosis and care, while fostering interdisciplinary collaboration in the management of oral-systemic health.

REFERENCES

[1] Al Subait AA, Saidi KM, Alsagoor HH, Alrasheedi NM, Alanazi AG, Alfattany RJ, Alshammari HM, Alshehri SA, Alfaifi AA, Alosaimi OO, Samargandi AA. Prevalence, incidence and outcome of periodontal diseases among the elderly. *Int J Community Med Public Health*. 2021;9(1):409–413. doi:10.18203/2394-6040.ijcmph20214835.

[2] American Academy of Periodontology. (2024). Gum disease risk factors.

[3] Zoheir N, Kurushima Y, Lin GH, Nibali L. Periodontal infectogenomics: A systematic review update of associations between host genetic variants and subgingival microbial detection. Clin Oral Investig. 2022;26(3):2209-21. doi:10.1007/s00784-021-04246-2

[4] Revilla-León, M., et al. (2023). Artificial intelligence models for diagnosing gingivitis and periodontal disease: A systematic review. *The Journal of Prosthetic Dentistry, 130*(6), 816-824.

[5] Morales, A., et al. (2016). Las enfermedades periodontales como enfermedades crónicas no transmisibles: Cambios en los paradigmas. *Revista Clínica Periodoncia, Implantología y Rehabilitación Oral, 9*(2), 203–207.

[6] Neurath, N., Kesting, M. (2024). Cytokines in gingivitis and periodontitis: from pathogenesis to therapeutic targets. *Frontiers in Immunology, 15*, 1435054.

[7] Martínez-García M, Hernández-Lemus E. Periodontal inflammation and systemic diseases: an overview. *Front Physiol*. 2021;12:709438. doi:10.3389/fphys.2021.709438.

[8] Varela-López A, Bullon B, Gallardo I, Quiles JL, Bullon P. Association of specific nutritional intake with periodontitis. BMC Oral Health. 2024;24(1):640. doi:10.1186/s12903-024-04384-6

[9] Chapple, I. L. C. (2014). Time to take periodontitis seriously. *BMJ, 348*, 1–2.

[10] Khader, Anas Abdul. "The Evolution of Periodontal Disease Classification." *Medical Forum Monthly*, vol. 35, no. 10, 2024.

[11] Celik D, Kantarci A. Vascular changes and hypoxia in periodontal disease as a link to systemic complications. *Pathogens*. 2021;10(10):1280. doi:10.3390/pathogens10101280.

[12] Nola, Iskra A., y Darko Kolanć. "Thermography in biomedicine." *2015 57th International Symposium ELMAR (ELMAR)*, IEEE, 2015.

[13] Ramirez-GarciaLuna, Jose L., Karla Rangel-Berridi, Robert Bartlett, Robert D.J. Fraser, and Mario A. Martinez-Jimenez. "Use of infrared thermal imaging for assessing acute inflammatory changes: a case series." *Cureus*, vol. 14, no. 9, 2022.

[14] Liu, Qian, Mingzhu Li, Wenping Wang, Shengbo Jin, Haozhe Piao, Yuxin Jiang, Ningxin Li, and Huini Yao. "Infrared thermography in clinical practice: a literature review." *European Journal of Medical Research*, vol. 30, no. 1, 2025, pp. 1–16.

[15] Gunupati, Sumanth, Hasya Sappiti, Sreenivas Nagarakanti, B. V. Ramesh Reddy, and Vijay Kumar Chava. "Validating gingival surface temperature as an alternative tool in the diagnosis of periodontal disease activity: An observational clinical trial." *Journal of Dental Research, Dental Clinics, Dental Prospects*, vol. 13, no. 2, 2019, pp. 123–127

[16] Wu Y, Wu M. Biomedical data mining and machine learning for disease diagnosis and health informatics. *Bioengineering*. 2024;11(4):364. doi:10.3390/bioengineering11040364.

[17] Rimi, Iffat Firozy, Md Ariful Islam Arif, Sharmin Akter, Md Riazur Rahman, AHM Saiful Islam, and Md Tarek Habib. "Machine learning techniques for dental disease prediction." *Iran Journal of Computer Science*, vol. 5, no. 3, 2022, pp. 187–195.

[18] Li, Xin, Dan Zhao, Jinxuan Xie, Hao Wen, Chunhua Liu, Yajie Li, Wenbin Li, and Songlin Wang. "Deep learning for classifying the stages of periodontitis on dental images: a systematic review and meta-analysis." *BMC Oral Health*, vol. 23, no. 1, 2023, pp. 1017.

[19] Chang, Hyuk-Joon, Sang-Jeong Lee, Tae-Hoon Yong, Nan-Young Shin, Bong-Geun Jang, Jo-Eun Kim, Kyung-Hoe Huh et al. "Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis." *Scientific Reports*, vol. 10, no. 1, 2020, p. 7531.

[20] Zhao, Dan, Morteza Homayounfar, Zhe Zhen, Mei-Zhen Wu, Shuk Yin Yu, Kai-Hang Yiu, Varut Vardhanabhuti, George Pelekos, Lijian Jin, and Mohamad Koohi-Moghadam. "A multimodal deep learning approach to predicting systemic diseases from oral conditions." *Diagnostics*, vol. 12, no. 12, 2022, p. 3192.

[21] Ferrara, Elisabetta, Biagio Rapone, and Alessandro D'Albenzio. "Applications of deep learning in periodontal disease diagnosis and management: a systematic review and critical appraisal." *Journal of Medical Artificial Intelligence*, vol. 8, 2025.

[22] Całkosiński I, Dobrzyński M, Rosińczuk J, Dudek K, Chrószcz A, Fita K, Dymarek R. The use of infrared thermography as a rapid, quantitative, and noninvasive method for evaluation of inflammation response in different anatomical regions of rats. *Biomed Res Int*. 2015;2015:972535. doi:10.1155/2015/972535.

[23] Jones, Bryan F. "A reappraisal of the use of infrared thermal image analysis in medicine." *IEEE Transactions on Medical Imaging*, vol. 17, no. 6, 1998, pp. 1019–1027.

[24] Rams, T.E.; Slots, J. (2024). Elevated subgingival temperature infers high bacterial pathogen counts in severe periodontitis. Clin. Exp. Dent. Res. 10(3):e891. DOI: 10.1002/cre2.891.

[25] Diniz de Lima, Elisa, José Alberto Souza Paulino, Ana Priscila Lira de Farias Freitas, José Eraldo Viana Ferreira, Jussara da Silva Barbosa, Diego Filipe Bezerra Silva, Patrícia Meira Bento, Ana Marly Araújo Maia Amorim, and Daniela Pita Melo. "Artificial intelligence and infrared thermography as auxiliary tools in the diagnosis of temporomandibular disorder." *Dentomaxillofacial Radiology*, vol. 51, no. 2, 2022

[26] Rahmayani, L., Yahya, M., Soraya, C., & Syahreza, S. (2020). Thermal condition of muscle area around the temporomandibular joint in patient with systemic lupus erythematosus using infrared thermography application: A case report. *Journal of International Society of Preventive and Community Dentistry*, **10**(5), 674–679.

[27] López, Hans Kepler Becerra, Midian Clara Castillo Pedraza, and Jorge Homero Wilches Visbal. "Principios físicos y avances recientes de la termografía infrarroja en odontología." *Revista Cubana de Medicina Militar*, vol. 51, no. 1, 2022. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0138-65572022000100007

[28] Aboushady, M. Atef, Wael Talaat, Zaid Hamdoon, Tarek M. Elshazly, Nivin Ragy, Christoph Bourauel, and Sameh Talaat. Thermography as a non-ionizing quantitative tool for diagnosing periapical inflammatory

lesions. *BMC Oral Health*, vol. 21, no. 1, 2021, p. 260. https://doi.org/10.1186/s12903-021-01618-9

[29] Bashir, Nasir Z., Zahid Rahman, and Sam Li-Sheng Chen. "Systematic comparison of machine learning algorithms to develop and validate predictive models for periodontitis." *Journal of Clinical Periodontology*, vol. 49, no. 10, 2022, pp. 958–969. https://doi.org/10.1111/jcpe.13692

[30] Radha, R. C., B. S. Raghavendra, B. V. Subhash, Jeny Rajan, and A. V. Narasimhadhan. "Machine learning techniques for periodontitis and dental caries detection: A narrative review." *International Journal of Medical Informatics*, vol. 178, 2023, 105170. https://doi.org/10.1016/j.ijmedinf.2023.105170

[31] Deng, K.; Zonta, F.; Yang, H.; Pelekos, G.; Tonetti, M.S. (2024). Development of a machine learning multiclass screening tool for periodontal health status based on non-clinical parameters and salivary biomarkers. J. Clin. Periodontol. 51(12):1547–1560.

[32] Shon, H.S.; Kong, V.; Park, J.S. (2022). Deep Learning Model for Classifying Periodontitis Stages on Dental Panoramic Radiography. Applied Sciences, 12(17):8500.

[33] Beak, W.; Park, J.; Ji, S. (2024). Data-driven prediction model for periodontal disease based on correlational feature analysis and clinical validation. Heliyon, 10(6):e32496.

[34] Chang, Jennifer, Ming-Feng Chang, Nikola Angelov, Chih-Yu Hsu, Hsiu-Wan Meng, Sally Sheng, Aaron Glick, *et al.* (2022). Application of deep machine learning for the radiographic diagnosis of periodontitis. *Clinical Oral Investigations*, **26**(11), 6629–6637.