Revista Mexicana de
Ingeniería Biomédica

RESEARCH ARTICLE

Point-Based Infrared Thermography and Two-Phase Machine Learning for Non-Invasive Periodontal Disease Classification

Clasificación no invasiva de la enfermedad periodontal mediante termografía infrarroja basada en mediciones puntuales y aprendizaje automático en dos fases.

Antony Morales-Cervantes[1,*], Gerardo Marx Chávez-Campos[1], Adriana del Carmen Téllez-Anguiano[1], Ricardo Martínez-Parrales[1], Mayra Yunuen Rincón-Pineda[1], Francisco Javier González[2]

[1] *Departamento de Estudios de Posgrado e Investigación, TecNM Instituto Tecnológico de Morelia, Av. Tecnológico #1500, colonia Lomas de Santiaguito, C.P. 58120, Morelia, Michoacán, México*

[2] *Coordinación para la Innovación y Aplicación de la Ciencia y la Tecnología, Universidad Autónoma de San Luis Potosí, Av. Sierra Leona #550-2ª, Lomas de San Luis, C.P. 78210 San Luis Potosí, México*

[*] Corresponding author(s); contact in the panel below.

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 XGBoost 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 localized infrared thermographic measurements 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.

Resumen. Las enfermedades periodontales, como la gingivitis y la periodontitis, son afecciones comunes de salud bucal que requieren una detección oportuna y precisa. Este estudio presenta un sistema de apoyo al diagnóstico no invasivo que integra la termografía infrarroja con datos clínicos para clasificar el estado de salud periodontal. Se realizó un estudio transversal en 91 individuos, categorizados como sanos, con gingivitis o con periodontitis. Se analizaron imágenes termográficas desde tres perspectivas faciales para extraer características de la temperatura gingival, las cuales se combinaron con parámetros clínicos, incluyendo el índice de placa, edad, sexo, tabaquismo y la presencia de afecciones sistémicas. Se evaluaron varios modelos de aprendizaje automático utilizando validación cruzada estratificada de 10 particiones, tanto con cómo sin reducción de dimensionalidad. Un enfoque de clasificación en dos etapas arrojó los mejores resultados: se utilizó regresión logística para identificar la periodontitis seguida de XGBoost para diferenciar entre los casos sanos y con gingivitis. El modelo combinado alcanzó una exactitud del 94.51% y una puntuación F1 del 94.49%, mientras que los modelos basados únicamente en datos termográficos alcanzaron una exactitud del 75.82%. Estos hallazgos respaldan la viabilidad de utilizar imágenes térmicas e inteligencia artificial para mejorar la clasificación de la enfermedad periodontal. El método propuesto ofrece una solución no invasiva prometedora para mejorar la precisión diagnóstica y orientar una atención odontológica personalizada.

Keywords: Artificial intelligence, infrared thermography, machine learning, periodontal disease. *(Palabras clave: Inteligencia artificial, termografía infrarroja, aprendizaje automático, enfermedad periodontal.)*

Revista Mexicana de Ingeniería Biomédica
ISSN 2395-9126 • SOMIB

Vol./No./Year	47/2/2026
Article no.	e2026-1666
DOI	10.17488/RMIB.47.2.1666
Received	10 March 2026
Accepted	23 June 2026
Published	05 July 2026
Corresponding	antony.mc@morelia.tecnm.mx

1. 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]. 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.

Recent advances in infrared thermography (IRT) have highlighted its potential as a non-invasive adjunctive tool for evaluating inflammatory oral conditions [11]. The physiological basis of periodontal thermography is associated with local vascular and metabolic alterations triggered by the inflammatory response, which promotes vasodilation, increased vascular permeability and perfusion, and localized thermal changes in gingival tissues [11], [12]. Consequently, inflammatory periodontal conditions may exhibit higher surface temperatures compared to healthy tissues, supporting the use of thermal biomarkers for disease assessment [12]. Recent studies have also explored the integration of artificial intelligence with thermal imaging for gingival inflammation analysis, reinforcing the growing interest in computational approaches for periodontal diagnostics [13].

Despite these advantages, reliable oral thermographic assessment requires rigorous acquisition standardization due to the high sensitivity of infrared imaging to environmental and procedural variations [14]. International thermographic guidelines emphasize the importance of patient acclimatization, controlled ambient temperature, airflow minimization, emissivity calibration, reproducible positioning, and standardized image acquisition protocols to reduce thermal artifacts and improve measurement reproducibility [15]. These considerations are particularly relevant in machine learning applications, where model performance strongly depends on the consistency and quality of the acquired thermal data.

Infrared thermography enables the assessment of both the temperature of affected areas and adjacent tissues, offering a comprehensive view of the inflammatory state [16], [17]. 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 [18]. To address this gap, this study proposes a reproducible and clinically oriented thermographic methodology that integrates localized thermographic measurements 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 [19]. 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 [20]. 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 [21]. 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 [22], [23], [24].

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 [23]. 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 [25]. 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.

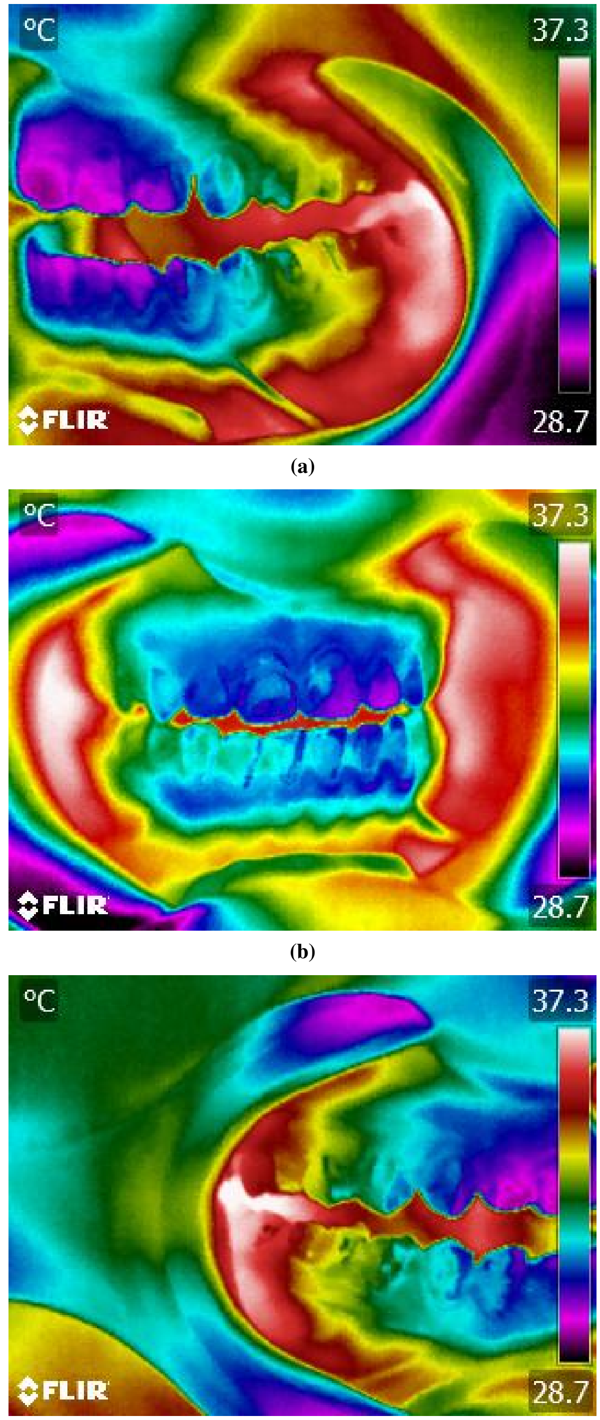
Finally, a 10-fold cross-validation is conducted to evaluate the system's stability and its usefulness in real clinical settings. By combining a functional thermal biomarker with computational modeling, this research connects the fields of clinical dentistry, biomedical imaging, and physiological signal 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.

2. 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 1. Each column group in Table 1 includes mean age ± standard deviation (SD), number and percentage of male participants, smokers, and individuals with systemic diseases, as well as the O'Leary plaque index (mean ± SD). The participants' ages ranged from 18 to 60 years old and included 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 [Institution redacted for blind review]. The study protocol was approved by the Institutional Ethics Committee (Approval Number 018/2025). All procedures involving human participants were conducted in accordance with the ethical standards of the institutional research committee and with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to their inclusion in the study.

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 were captured using a FLIR® T400 camera based on a focal plane array of microbolometers, with a thermal sensitivity of $< 0.05\,°C$ (50 mK), an accuracy of $\pm 2\%$, a spectral range of 7.5–13 μm, and a temperature range from -20 to $1200\,°C$. Emissivity was calibrated to 0.97 [26], and the camera distance was 0.3 m. To ensure consistency during image acquisition, all patients were evaluated in a standardized supine position on a dental chair. The working distance was initially calibrated using a reference ruler and subsequently maintained throughout image acquisition by the operator. For lateral perspectives, the camera position was adjusted relative to the patient's sagittal plane to obtain direct orthogonal views of the lateral gingival regions, minimizing perspective-related thermal distortion and maintaining consistent acquisition conditions across subjects. 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.

Thermographic analysis was performed through anatomically guided point-based thermal extraction using FLIR Tools software. Thermal measurements were manually obtained by a trained dental expert (Master's candidate in Dental Sciences) and subsequently reviewed by a senior periodontal specialist to ensure consistency in point placement and data extraction. The analysis focused on the gingival mucosa surrounding the incisors and canines across all four dental quadrants. Rather than extracting average temperatures from large segmented regions, localized temperature measurements were obtained



(a) (b) (c)

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

Table 1. Baseline Characteristics of the Study Population.

Characteristics	Periodontitis (N=30)	Gingivitis (N=30)	Controls (N=31)
Average age (years)	42.9 ± 10.3	26.0 ± 7.8	29.6 ± 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 ± 9.7	85.8 ± 18.6	67.6 ± 15.4

from three standardized anatomical points per tooth: the mesial papilla (Sp1), the medial gingival margin (Sp2), and the distal papilla (Sp3), as illustrated in Figure 2. These measurements were manually positioned directly over the visible gingival tissue in each thermographic view.

For each participant, thermal measurements were collected from frontal, left lateral, and right lateral views, generating a total of 36 continuous thermal variables per subject. Consequently, the thermographic images were transformed into a structured low-dimensional tabular dataset used as input for the machine learning pipeline. Importantly, the raw thermographic images themselves were not directly used for model training, feature extraction through deep learning, or pixel-wise image analysis.

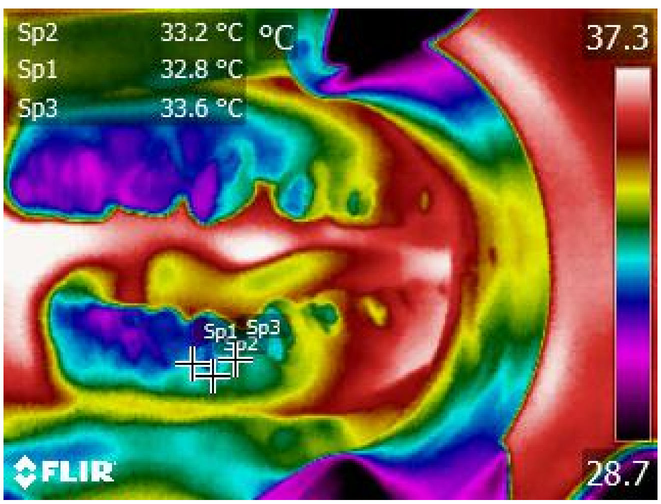


Figure 2. Thermographic measurement on lower left canine: mesial, medial, and distal regions identified.

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 integration of localized thermal measurements with clinical variables enabled evaluation of the complementary diagnostic contribution of thermographic biomarkers within the proposed classification framework.

2.1 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. Principal Component Analysis (PCA) was applied for dimensionality reduction of the extracted thermographic and clinical feature set. 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.

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. To reduce the risk of overfitting during this exploratory study, exhaustive hyperparameter optimization strategies were not performed. Instead, the evaluated classification algorithms were implemented primarily using standard baseline configurations to ensure consistent comparative evaluation across models. Limited parameter adjustments were introduced only when necessary to ensure stable convergence and model operation (e.g., Logistic Regression with max_iter=500). Consequently, the reported performance metrics reflect comparative model behavior under a consistent 10-fold cross-validation framework.

2.2 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 was performed to assess each model's ability to distinguish among the three classes. The decision to implement a two-phase classification strategy was driven by preliminary statistical analyses and multiclass model evaluations. Initial ANOVA and post-hoc Tukey analyses suggested overlapping distributions between gingivitis and periodontitis for several thermal variables, indicating increased classification complexity between these disease stages. Subsequent multiclass ROC curve anal-

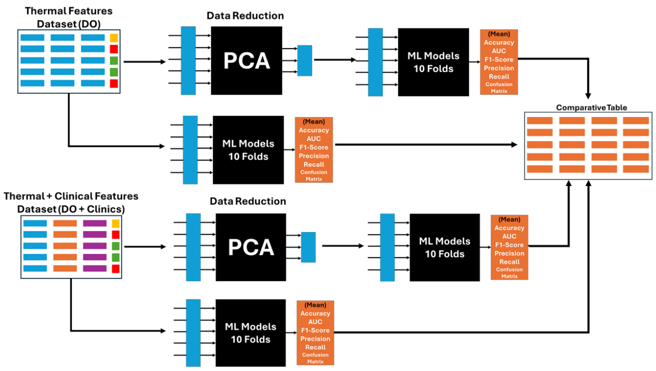


Figure 3. Model evaluation with PCA and cross-validation using thermographic and clinical features for classification.

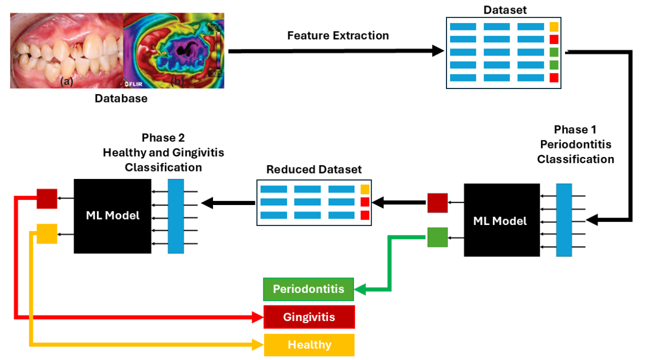


Figure 4. Two-phase classification model: first detects periodontitis, then classifies gingivitis and healthy cases.

yses in a One-vs-Rest (OvR) scheme further supported this observation by revealing distinct class-specific discriminative behaviors across the evaluated models. 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. The Area Under the Curve (AUC) values from the ROC analysis are presented in Table 2. Specifically, simpler linear models like Logistic Regression achieved the highest performance in isolating Periodontitis cases from the rest of the dataset (AUC = 0.900), outperforming complex ensembles for this specific task. Conversely, distinguishing between healthy subjects and those with mild inflammation (Gingivitis) required algorithms capable of capturing non-linear relationships, a task well-suited for tree-based ensemble methods like XGBoost, which demonstrated optimal performance in this regard (AUC = 0.992 for the Healthy class).

Therefore, rather than using a single-phase multiclass classifier that compromises overall performance (as shown in Table 4), a sequential architecture was designed to take advantage of the complementary classification behavior observed across models. The sequential architecture was not designed a priori, but emerged from the observed class-specific behavior of the evaluated multiclass models during preliminary experiments. 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.

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.

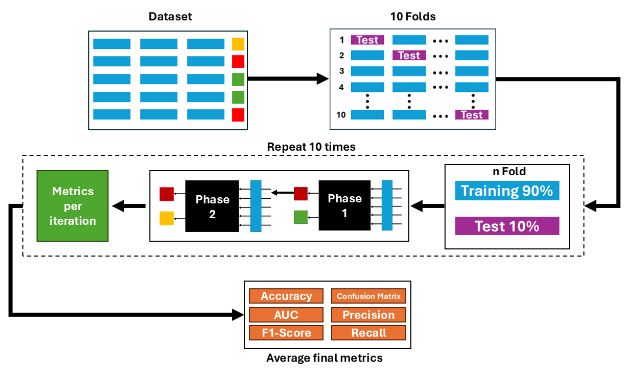


Figure 5. Scheme of 10-fold cross-validation used to evaluate the two-phase periodontal classification system.

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 standard classification metrics provide a comprehensive evaluation of the model's overall accuracy, discriminative ability, and the balance between precision and recall, and were calculated following established machine learning methodologies in dental research [20].

The specific confusion matrix for our multiclass classification problem (Healthy, Gingivitis, and Periodontitis) is represented as:

$$M = \begin{bmatrix} TP_h & FN_{h\to g} & FN_{h\to p} \\ FP_{g\to h} & TP_g & FN_{g\to p} \\ FP_{p\to h} & FP_{p\to g} & TP_p \end{bmatrix} \tag{1}$$

where:

- $\mathrm{TP_h}$, $\mathrm{TP_g}$, $\mathrm{TP_p}$: True positives for healthy, gingivitis, and periodontitis.
- $\mathrm{FN_{hg}}$, $\mathrm{FN_{hp}}$: Healthy patients misclassified as gingivitis or periodontitis.

Table 2. One-vs-Rest (OvR) ROC-AUC scores for single-phase multiclass model evaluation.

Model	Healthy (AUC)	Gingivitis (AUC)	Periodontitis (AUC)
Logistic Regression	0.989	0.862	0.900
KNN	0.974	0.797	0.792
Decision Tree	0.992	0.793	0.793
Random Forest	0.991	0.883	0.890
SVM	0.976	0.855	0.893
Neural Network	0.985	0.848	0.862
XGBoost	0.992	0.847	0.886
LightGBM	0.985	0.858	0.897

- FN_{gp}: Gingivitis patients misclassified as periodontitis.
- FP_{gh}, FP_{ph}: Gingivitis or periodontitis predicted as healthy.
- FP_{pg}: Periodontitis misclassified as gingivitis.

3. Results and Discussion

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 3 and Table 4 show the average accuracy and area under the ROC curve (AUC) obtained for each model across the different configurations evaluated.

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.

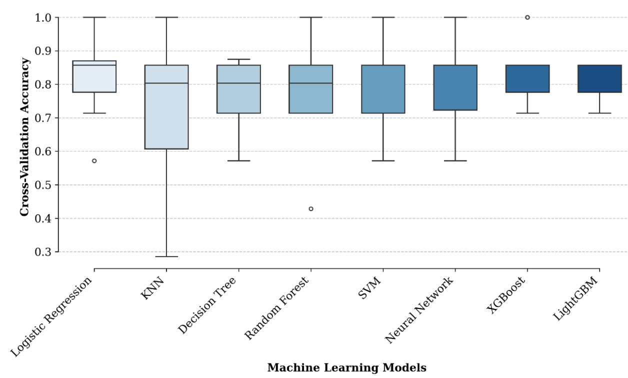


Figure 6. Accuracy distribution across models with Thermal + Clinical Features (without PCA) in 10-fold cross-validation.

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 (0.900) for periodontitis detection (Table 2), 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.

3.1 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 5, 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 7(a), which shows that the classifier accurately identifies patients with periodontitis and effectively differentiates them from those with gingivitis and healthy subjects.

Table 3. Performance of models using thermal and clinical features for input (i) and (ii) configurations

Model	(i) PCA (IRT) Accuracy	(i) PCA (IRT) AUC	(ii) IRT only Accuracy	(ii) IRT only 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

Table 4. Performance of models using thermal and clinical features for input (iii) and (iv) configurations.

Model	(iii) PCA (IRT + Clinical) Accuracy	(iii) PCA (IRT + Clinical) AUC	(iv) IRT + Clinical Accuracy	(iv) IRT + Clinical 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

Table 5. 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

Figure 7 (a) 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.

To further assess the robustness of the model, Figure 8 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.

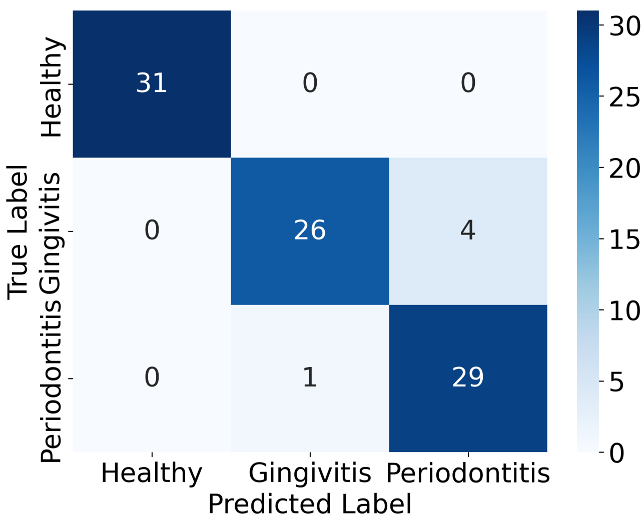
After confirming the model's robustness using the complete feature set, we also evaluated its performance using only thermal features to explore the standalone diagnostic potential of infrared thermography. As reported in Table 5, this configuration resulted in a decreased performance: accuracy of 75.82%, F1-score of 0.7535, recall of 0.7581, and precision of 0.7714. The corresponding confusion matrix is shown in Figure 7 (b), revealing notable challenges in distinguishing between gingivitis and periodontitis when clinical variables are excluded.

To further evaluate the complementary contribution of infrared thermography, an additional baseline experiment was conducted using only the clinical variables (age, sex, smoking status, systemic disease, and O'Leary index) under the same 10-fold cross-validation framework (Table 6). Models trained exclusively on clinical data achieved strong diagnostic performance, with LightGBM yielding the highest accuracy (88.00%) and F1-score (0.8762), followed closely by Random Forest and Logistic Regression models. However, this performance remained below that achieved by the proposed multimodal two-phase classifier integrating both thermographic and clinical information (Accuracy = 94.51%, F1-score = 0.9449). In contrast, the thermal-only configuration achieved an accuracy of 75.82%, indicating that neither modality independently captured the full complexity of periodontal inflammation. These findings support the complementary nature of thermal and clinical information, where localized gingival temperature measurements provide physiologically relevant biomarkers that enhance discrimination between subtle inflammatory stages when combined with contextual clinical variables.

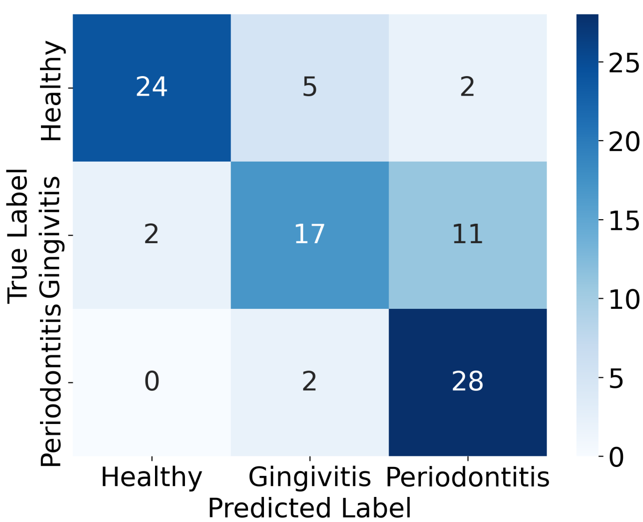
To explicitly assess the potential confounding effect of age, an additional age-exclusion analysis was performed by removing age from both the clinical-only and multimodal datasets while preserving the same 10-fold cross-validation framework. This analysis was motivated by the age imbalance observed among groups, particularly the higher mean age of the periodontitis group. When age was excluded, the best clinical-only model was LightGBM, with an accuracy of 77.00%, an AUC of 0.8778, and an F1-score of 0.7502, compared with 88.00% accuracy and an F1-score of 0.8762 when age was included. This reduction confirms that age contributed substantially to

Table 6. Average performance of machine learning models using only clinical variables under 10-fold cross-validation.

Model	Accuracy	F1-score	Recall	Precision
Logistic Regression	0.8555	0.8468	0.8555	0.8761
KNN	0.77	0.7451	0.77	0.7913
Decision Tree	0.8478	0.8403	0.8478	0.865
Random Forest	0.8789	0.8723	0.8789	0.8958
SVM	0.8233	0.8078	0.8233	0.8348
ANN	0.5811	0.4904	0.5811	0.4754
LightGBM	0.88	0.8762	0.88	0.8883
XGBoost	0.8467	0.8424	0.8467	0.8661



(a)



(b)

Figure 7. Confusion matrices of the two-phase classifier: (a) with clinical data, (b) without.

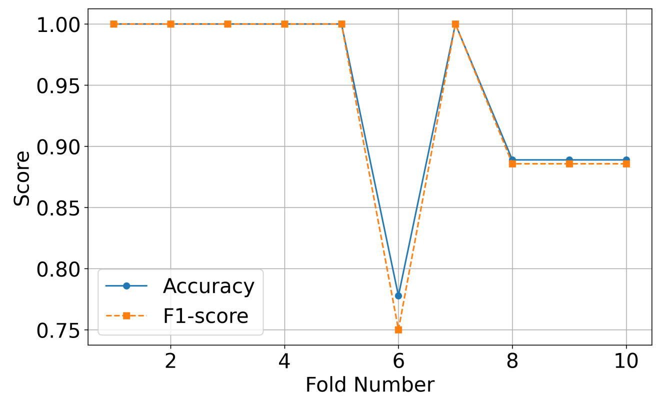


Figure 8. Performance across 10 folds of the two-phase classifier with thermal and clinical features.

the clinical-only baseline performance and should therefore be considered a relevant covariate and potential confounding factor in the present cohort.

Importantly, the proposed multimodal two-phase model was also evaluated after excluding age, achieving an accuracy of 83.52%, an F1-score of 0.8263, a recall of 0.8352, and a precision of 0.8841. Although lower than the original age-included multimodal model, this performance remained above the age-excluded clinical-only baseline. The performance gap between the clinical-only baseline and the thermography-assisted model remained consistent regardless of whether age was included (88.00% vs. 94.51%) or excluded (77.00% vs. 83.52%). These findings suggest that localized gingival thermographic measurements provide complementary predictive information beyond the non-age clinical variables, while also indicating that age-balanced external validation is necessary before generalizing the magnitude of the reported performance.

3.2 Comparison

To contextualize the findings, Table 7 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. [18] and Rams and Slots [27] 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 [28], lupus erythematosus [29], as well as applications in implantology and endodontics [30]. 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 [31], systematic evaluation of its diagnostic potential in the periodontal context, especially when combined with computational approaches, has been scarce.

As shown in the last row of Table 7, 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 [33], [35]. 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. [33]; ACC = 92.9% in Shon et al. [34]).

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. [37], 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 its high accuracy (ACC) and area under the curve (AUC), can be attributed to the integration of clinically relevant variables with localized thermographic measurements and to the use of a consistent internal validation strategy. In the present study, model performance was evaluated using stratified 10-fold cross-validation to reduce dependence on a single train-test partition and provide a more stable estimate of classification behavior. However, exhaustive hyperparameter optimization and external validation were not performed; therefore, the reported results should be interpreted as exploratory internal validation estimates rather than as definitive evidence of generalizability. This consideration is important because previous studies have shown that models evaluated only on internal cohorts may overestimate their predictive performance when applied to external populations [32].

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. [32] and Beak et al. [35], who demonstrated that pruning irrelevant input variables enhances both predictive power and model stability. Furthermore, the integration of localized thermographic biomarkers introduced a novel, physiologically grounded dimension. Given that active inflammation induces detectable thermal variations [18], [27], 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 [33], [34]. 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 [27], 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

Table 7. Comparison of previous studies on conventional periodontal diagnosis vs. new technologies.

Reference (Year)	Method / Main Metric	Sample	Key Findings
Gunupati et al. (2019) [18]	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) [27]	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) [31]	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 (κ=0.97).
Bashir et al. (2022) [32]	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) [33]	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 ∼89.8% (3 classes) and ∼80.5% (6 classes).
Shon et al. (2022) [34]	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) [35]	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) [36]	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); demonstrated competitive performance compared with previous multiclass periodontal classification studies. Thermal-only model: AUC = 0.82, ACC = 75.82%, F1 = 0.7535, recall = 0.7581, precision = 0.7714.

potential to enhance early detection, guide timely interventions, and improve long-term outcomes, including tooth preservation and control of associated systemic inflammation.

3.3 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 [32]. 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) [35].

Although the sequential architecture improved classification performance, potential error propagation between Phase 1 and Phase 2 remains an inherent limitation of multi-stage classifiers. Analysis of the aggregated confusion matrices across the 10-fold cross-validation experiments revealed that the primary misclassifications occurred between Gingivitis and Periodontitis cases, likely reflecting the biological overlap between adjacent inflammatory stages. Out of 91 instances, Phase 1 exhibited only one False Negative, allowing a single Periodontitis case to erroneously propagate to Phase 2, where it was classified as Gingivitis. In contrast, Phase 1 produced four False Positives, classifying true Gingivitis cases as Periodontitis; these cases were terminally classified and effectively bypassed Phase 2. Remarkably, within Phase 2, the model made zero internal misclassifications among the true Healthy and Gingivitis cases it received. Clinically, this error distribution indicates that the two-phase framework possesses a slight tendency toward over-triage rather than under-triage, a conservative behavior that is generally favorable for early clinical screening tools. Furthermore, a methodological limitation exists regarding the structural design of the two-phase classification architecture. Because the sequential framework, composed of Logistic Regression followed by XGBoost, was formulated post-hoc based on the One-vs-Rest (OvR) performance of single-phase models evaluated on the same dataset, there is a risk of architecture-selection bias. Although 10-fold cross-validation provided an internal estimate of model performance and reduced the dependence on a single train-test split, it does not fully eliminate the possibility that the pipeline design was influenced by dataset-specific patterns. Therefore, the two-phase framework should be interpreted as an exploratory architecture generated from the present cohort rather than as a fully optimized or externally validated diagnostic pipeline. Future studies should validate this architecture prospectively using independent cohorts, ideally with the classification strategy predefined before model evaluation.

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. Furthermore, age may represent a potential confounding factor in thermographic periodontal assessment, since physiological aging processes can influence vascularization, inflammatory response, and localized tissue temperature. Although age was incorporated as a clinical variable within the proposed multimodal framework, the present cohort size did not allow a dedicated age-stratified analysis to independently quantify its effect on gingival thermal measurements. 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. [34], 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

4. 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 localized gingival thermal measurements 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 supporting the development of accessible and scalable diagnostic support methodologies.

Notably, the study indicates that localized thermographic gingival measurements could represent complementary thermographic biomarkers associated with inflammatory activity, potentially supporting conventional periodontal assessment 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 may contribute to improving screening consistency and supporting future decision-support strategies in periodontal evaluation.

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. Finally, the current methodology relies on manual extraction of localized thermal measurements under expert clinical supervision. Although this approach ensured anatomically consistent point placement during this proof-of-concept study, it may introduce observer-dependent variability and limit scalability in larger clinical settings. Future work will therefore focus on developing automated thermographic landmark detection and extraction methods using computer vision and deep learning architectures, such as YOLO or U-Net, to improve reproducibility and facilitate future clinical implementation.

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 demonstrates the feasibility of combining localized gingival thermographic measurements with machine learning approaches for periodontal disease classification. Although additional multicenter validation and methodological standardization remain necessary, the proposed framework may contribute to the future development of non-invasive decision-support tools for periodontal assessment, while fostering interdisciplinary collaboration in the management of oral-systemic health.

5. Funding

This work was supported by the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) through a postdoctoral fellowship awarded to Antony Morales-Cervantes under Grant CVU 418502.

6. Acknowledgements

We sincerely thank the Faculty of Sciences and the Faculty of Dentistry at the Universidad Autónoma de San Luis Potosí (UASLP) for their support in the collection and management of the database used in this study. This database was originally generated within the framework of the Ph.D. thesis in Electronic Engineering by Dr. Antony Morales-Cervantes and the Master's thesis in Dental Sciences by MSc. Xiomara Osorio, respectively, and is utilized independently for the scope of this work. The authors particularly extend their gratitude to Dr. José Obed García for facilitating the collaboration between both faculties during that initial stage.

7. Author Contributions (CRediT)

Morales-Cervantes, A. : Conceptualization; Methodology; Software; Formal analysis; Writing – original draft. **Chávez-Campos, G. M.** : Supervision; Validation; Writing – review & editing. **Téllez-Anguiano, A. C., Martínez-Parrales, R**.: Investigation; Validation; Data curation; Writing – review & editing. **Rincón-Pineda, M. Y**. : Investigation; Validation; Writing – review & editing**. González, F. J.** : Resources; Supervision.

8. Conflict of Interest

The authors declare no conflict of interest.

9. Data Availability Statement

The datasets generated and analyzed during the current study are not publicly available due to patient privacy and confidentiality restrictions, but are available from the corresponding author upon reasonable request.

10. Bioethics

The study was approved by the Institutional Ethics Committee of the Instituto Tecnológico de Morelia (protocol number 018/2025). All procedures involving human participants were conducted in accordance with the ethical standards of the institutional research committee and with the Declaration of Helsinki and its subsequent revisions. Written informed consent was obtained from all participants prior to inclusion in the study.

11. Artificial Intelligence Use Disclosure

The authors declare that no generative AI tools were used in the preparation of this manuscript.

12. References

[1] A. A. Al Subait *et al.*, "Prevalence, incidence and outcome of periodontal diseases among the elderly," *Int J Community Med Public Health*, vol. 9, no. 1, pp. 409–413, 2021, doi: 10.18203/2394-6040.ijcmph20214835.

[2] American Academy of Periodontology, "Gum disease risk factors." `https://www.perio.org/` 2024.

[3] N. Zoheir, Y. Kurushima, G. H. Lin, and L. Nibali, "Periodontal infectogenomics: A systematic review update of associations between host genetic variants and subgingival microbial detection," *Clin Oral Investig*, vol. 26, no. 3, pp. 2209–2221, 2022, doi: 10.1007/s00784-021-04246-2.

[4] M. Revilla-León *et al.*, "Artificial intelligence models for diagnosing gingivitis and periodontal disease: A systematic review," *The Journal of Prosthetic Dentistry*, vol. 130, no. 6, pp. 816–824, 2023.

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

[6] N. Neurath and M. Kesting, "Cytokines in gingivitis and periodontitis: From pathogenesis to therapeutic targets," *Frontiers in Immunology*, vol. 15, p. 1435054, 2024.

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

[8] A. Varela-López, B. Bullon, I. Gallardo, J. L. Quiles, and P. Bullon, “Association of specific nutritional intake with periodontitis,” *BMC Oral Health*, vol. 24, no. 1, p. 640, 2024, doi: 10.1186/s12903-024-04384-6.

[9] I. L. C. Chapple, “Time to take periodontitis seriously,” *BMJ*, vol. 348, p. g2645, 2014.

[10] A. A. Khader, “The evolution of periodontal disease classification,” *Medical Forum Monthly*, vol. 35, no. 10, 2024.

[11] S. N. Elboray, O. M. Ezzatt, A. Salah, A. Mohamed, and M. M. Abdalwahab, “Evaluation of the accuracy of infrared thermographic imaging for the diagnosis of periodontal diseases: A cross-sectional study,” Journal of Clinical Periodontology, 2025, doi: 10.1111/jcpe.70066.

[12] H. N. Guetatlia et al., “Thermography and infrared spectroscopy in the detection of periodontal inflammation in vivo: A systematic review,” Diagnostics, vol. 16, no. 222, 2026, doi: 10.3390/diagnostics16020222.

[13] Z. T. Çankaya, A. Koyuncu, and S. Gürbüz, “Artificial intelligence assisted thermal imaging for gingival inflammation assessment: A novel approach,” Journal of Esthetic and Restorative Dentistry, 2026, doi: 10.1111/jerd.70045.

[14] L. K. M. O. Lopes et al., “The influence of different cheek and lip retractors and emissivity on intraoral infrared thermography,” Dentomaxillofacial Radiology, vol. 53, no. 6, pp. 417–422, 2024, doi: 10.1093/dmfr/twae025.

[15] R. G. Schwartz et al., “Guidelines for dental-oral and systemic health infrared thermography,” Pan American Journal of Medical Thermology, vol. 5, pp. 41–55, 2019.

[16] J. L. Ramirez-GarciaLuna, K. Rangel-Berridi, R. Bartlett, R. D. J. Fraser, and M. A. Martinez-Jimenez, “Use of infrared thermal imaging for assessing acute inflammatory changes: A case series,” *Cureus*, vol. 14, no. 9, 2022.

[17] Q. Liu *et al.*, “Infrared thermography in clinical practice: A literature review,” *European Journal of Medical Research*, vol. 30, no. 1, pp. 1–16, 2025.

[18] S. Gunupati, H. Sappiti, S. Nagarakanti, B. V. R. Reddy, and V. K. 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, pp. 123–127, 2019.

[19] Y. Wu and M. Wu, “Biomedical data mining and machine learning for disease diagnosis and health informatics,” *Bioengineering*, vol. 11, no. 4, p. 364, 2024, doi: 10.3390/bioengineering11040364.

[20] I. F. Rimi, M. A. I. Arif, S. Akter, M. R. Rahman, A. S. Islam, and M. T. Habib, “Machine learning techniques for dental disease prediction,” *Iran Journal of Computer Science*, vol. 5, no. 3, pp. 187–195, 2022.

[21] X. Li *et al.*, “Deep learning for classifying the stages of periodontitis on dental images: A systematic review and meta-analysis,” *BMC Oral Health*, vol. 23, no. 1, p. 1017, 2023.

[22] H.-J. Chang *et al.*, “Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis,” *Scientific Reports*, vol. 10, no. 1, p. 7531, 2020.

[23] D. Zhao *et al.*, “A multimodal deep learning approach to predicting systemic diseases from oral conditions,” *Diagnostics*, vol. 12, no. 12, p. 3192, 2022.

[24] E. Ferrara, B. Rapone, and A. 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.

[25] I. Całkosiński *et al.*, “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*, vol. 2015, p. 972535, 2015, doi: 10.1155/2015/972535.

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

[27] T. E. Rams and J. Slots, “Elevated subgingival temperature infers high bacterial pathogen counts in severe periodontitis,” *Clin. Exp. Dent. Res.*, vol. 10, no. 3, p. e891, 2024, doi: 10.1002/cre2.891.

[28] E. Diniz de Lima *et al.*, “Artificial intelligence and infrared thermography as auxiliary tools in the diagnosis of temporomandibular disorder,” *Dentomaxillofacial Radiology*, vol. 51, no. 2, 2022.

[29] L. Rahmayani, M. Yahya, C. Soraya, and S. Syahreza, “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*, vol. 10, no. 5, pp. 674–679, 2020.

[30] H. K. B. López, M. C. Castillo Pedraza, and J. H. Wilches Visbal, “Principios físicos y avances recientes de la termografía infrarroja en odontología,” *Revista Cubana de Medicina Militar*, vol. 51, no. 1, 2022, Available: `http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0138-65572022000100007`

[31] M. A. Aboushady *et al.*, “Thermography as a non-ionizing quantitative tool for diagnosing periapical inflammatory lesions,” *BMC Oral Health*, vol. 21, no. 1, p. 260, 2021, Available: `https://doi.org/10.1186/s12903-021-01618-9`

[32] N. Z. Bashir, Z. Rahman, and S. L.-S. Chen, “Systematic comparison of machine learning algorithms to develop and validate predictive models for periodontitis,” *Journal of Clinical Periodontology*, vol. 49, no. 10, pp. 958–969, 2022, Available: `https://doi.org/10.1111/jcpe.13692`

[33] 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.*, vol. 51, no. 12, pp. 1547–1560, 2024.

[34] H. S. Shon, V. Kong, and J. S. Park, “Deep learning model for classifying periodontitis stages on dental panoramic radiography,” *Applied Sciences*, vol. 12, no. 17, p. 8500, 2022.

[35] W. Beak, J. Park, and S. Ji, “Data-driven prediction model for periodontal disease based on correlational feature analysis and clinical validation,” *Heliyon*, vol. 10, no. 6, p. e32496, 2024.

[36] J. Chang *et al.*, “Application of deep machine learning for the radiographic diagnosis of periodontitis,” *Clinical Oral Investigations*, vol. 26, no. 11, pp. 6629–6637, 2022.

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