



Deep Learning Technique for Non-Invasive Periodontitis Classification from Oral Thermal Images Using Vision Transformer



Deepa Balaraman^{1*}, Jayashree Kannappan²

¹ Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai 602105, India

² Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai 600123, India

Corresponding Author Email: deepa.b@rajalakshmi.edu.in

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.430226>

ABSTRACT

Received: 18 January 2026

Revised: 16 April 2026

Accepted: 23 April 2026

Available online: 30 April 2026

Keywords:

thermal imaging, oral diseases, Pelican Optimization Algorithm, Linear Binary Pattern, Convolutional Neural Network, Vision Transformer, deep learning

Maintaining personal hygiene and a healthy lifestyle is mandatory for having a good interaction with society. Currently, the prevalence of oral diseases is rising significantly due to the modernized lifestyle of people. Gingivitis and Periodontitis are the most pervasive dental disease that exists across the world, irrespective of age. These gum diseases prevail due to the uncontrolled plaque condition that leads to several systemic disorders. The initial stages of plaque have a soft deposit over the teeth, and it progresses to form a hard deposit called calculus. It is difficult to maintain or control the plaque condition without periodic dental visits. The advances of Artificial Intelligence (AI) have provided commendable contributions in the domain of medical image processing using Deep Learning. Many existing research works have used visible images for classifying the gingivitis and Periodontitis through machine learning and deep learning algorithms. However, the misclassification occurs due to distorted images, interruption of noise signals, and poor pixel quality with external environmental lighting conditions. To address the above issues, the proposed research work classifies the coronal plaque, gingival, and Periodontitis using thermal images. Thermal images are pre-processed using Contrast Limited Adaptive Histogram Equalization (CLAHE) and the gamma correction method. Thermal noises are eradicated using the Gaussian filter by tuning its mean value. The features are extracted through multi-task learning using the Pelican Optimization Algorithm (POA) and the Tuned Linear Binary Pattern (TLBP) algorithm. The proposed Tuned Convolutional Neural Network (CNN) with Vision information Transformer algorithm (CNN-ViT) classifies the oral diseases (Coronal Plaque, Gingivitis, and Periodontitis) with an accuracy of 98.99% comparing the existing models.

1. INTRODUCTION

Nearly 3.5 billion people are suffering from oral gum health diseases [1, 2] according to World Health Organization (WHO) records. Recently, many people, irrespective of age, are affected by gum disease due to unhealthy food habits [3, 4]. Unhealthy food eating and improper maintenance of teeth lead to gum diseases starting from mild infectious gum and heart strokes [5], diabetes [6, 7]. Teeth are the first organ to be infected due to eating junk foods. The tooth infection arises due to unhealthy activities such as smoking, poor nutrition, and hormonal changes [8-10]. Tiny food particles deposit between gums and teeth, causing bacterial infections and leading to plaque formation [11]. The texture of plaque is a sticky and soft layer, which causes bacterial infections on the teeth. This infection should be treated immediately; otherwise, it leads to the calculus stage. Calculus, also known as tartar, is a hard texture deposited over the teeth due to the chemical reaction caused by the minerals and saliva inside the mouth. Calcium and phosphate are the main reasons for the hardened texture in teeth [12, 13]. Tartar is very rough and porous due to the chemical components such as Calcium Phosphate

$((Ca_3PO_4)_2)$ Calcium Carbonate $CaCO_3$ Magnesium Phosphate $(Mg_3(PO_4)_2)$ and Hydroxyapatite $(Ca_{10}(PO_4)_6(OH)_2)$. The above components are formed due to the process of crystallization with saliva and form a hard layer in the teeth. The calculus is of two types, such as Supragingival (above the gum) and subgingival (below the gum), as depicted in Figure 1.

Normally, supra-gingival calculus occurs near the openings of salivary glands (i.e., behind the lower teeth or the upper molars). The sediment appears yellow or white. In contrast, subgingival calculus is located at the root of the gums, where there is interaction with blood. It is often in dark brown or black. This condition results in swelling or irritation of the gums, medically termed as gum inflammation or gingivitis. This gum inflammation damages the soft tissues of the teeth and creates problems such as reddish, swollen gums [14], bleeding gums [15], pain while chewing, halitosis [16] (bad breath), bad taste [17], deep pockets between teeth and gums [18], receding gums [19], and tooth loss [20-22]. These sustained conditions lead to Periodontitis [23-25], which is a chronic inflammatory disease. The accumulation of sticky thin film damages the bones and tissues surrounding the teeth and

gums. Periodontitis is the starting stage of tooth bone rupture and tissue damage. It is reversible with the removal of the contaminated particles from the teeth. The appearance of the teeth changes at different stages of infection, as shown in Figure 2. Through continuous dental checkups and the scaling process, the above condition can be reversed without proceeding to further growth of dental disease.



Figure 1. Differences between supragingival and subgingival



Figure 2. Comparison of healthy teeth versus different stages of infectious teeth

The infection worsens if it is not controlled at the early stages, affecting the roots of the teeth and leading to the loss of teeth. The growth of periodontitis disease is not consistent among individuals; it can be managed with proper self-care and periodic dental checkups [26-28]. While dental checkups can be expensive, they are recommended for severely affected patients to reinforce personal hygiene and plaque control.

Untreated Periodontitis not only affects the teeth and gums but also causes severe respiratory syndromes [29], heart strokes [30], low blood pressure [31], and diabetes [32, 33]. Hence, continuous monitoring and an effective treatment plan are essential. Some common treatments for gingivitis and Periodontitis include scaling, antibiotics, root planning, and surgical options. Surgical operations depend on the severity of the infection. The early stages of tooth decay are corrected with surgical procedures such as fillings, replacing crowns, and root canal therapy. The gingivitis disease is reversed through professional cleaning and continuous maintenance of oral hygiene. The Periodontitis is corrected with scaling, the utilization of antibiotics, and root canal surgery. The worst case of Periodontitis is irreversible and leads to the loss of a tooth. Predicting the plaque and gingivitis at an earlier stage helps to treat the infection effectively, and their respective symptoms are listed in Figure 3.

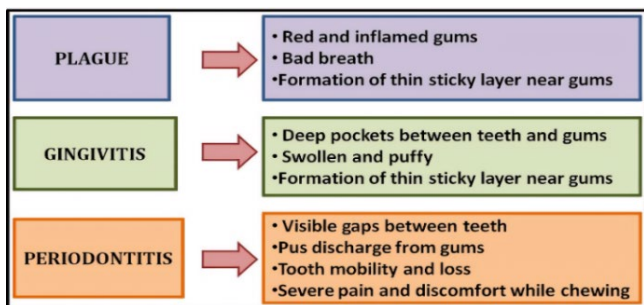


Figure 3. Common symptoms of oral disease

2. STUDY POPULATION

Currently, many researchers have explored innovative methodologies for detecting plaque, gingivitis, and Periodontitis at early stages. Recently, deep learning and machine learning algorithms have been used for the prediction and prognosis of any oral disease in the medical field. Nowadays, infrared (IR) and visible images are used for predicting oral diseases. A detailed study of the classification of oral diseases using machine learning and deep learning is discussed. The deep learning networks are based on various architectures such as Alexnet [6, 8], VGG [6, 9], Google Net [4, 6], and ResNet [2, 4, 6, 9]. The network addresses the vanishing gradient problem and helps to increase the learning rate. The minute details in the image are enhanced due to the rapid increase in the model's learning rate [1, 2]. Chau et al. [1] designed a hybrid model combining a DeepLabv3+ neural network, Xception, and MobileNetV2, predicting the periodontal condition. A Faster Region-Based Convolutional Neural Network (R-CNN) using ResNet-50 network [2] is applied to 567 intraoral photographs from the University of Hong Kong Dental Clinic for oral disease identification. The performance of the model is evaluated using the specificity (0.92) and sensitivity (0.94). Similarly, 134 new images from orthodontic patients were collected and classified. The performance of any deep learning model is measured using mean Average Precision (mAP). In this model, gum disease is predicted with an accuracy of 77.12%. Low image contrast or illumination noise degrades the performance of the model, resulting in a low recall value of 41.75%. Aroonratana et al. [3] classify the periodontal disease using a logistic regression (LR) model, which achieves a decent accuracy that reduces the Periodontitis prevalence to 44%. Its accuracy rate is low due to an imbalanced dataset, which leads to a poor learning rate.

Gingivitis, calculus, and soft deposits in the teeth are also classified using smartphones [4, 5]. The dataset is collected from 625 patients from Nanjing Stomatological Hospital in China. Convolutional Neural Network (CNN) classifies the oral diseases, and its performance is validated using Area Under Curve (AUC). The AUC value for the gingivitis condition is 87.11%, while calculus is nearly 80.11%. Soft deposits have an AUC rate of approximately 78.5%, which is very low compared to the other two diseases. The low contrast in images is due to improper pre-processing techniques and low-resolution image quality. The early gingivitis [5] signs are detected with ResNet-50 using Faster R-CNN. The teeth of orthodontic and periodontal patients are classified with an accuracy of 77.12%, a precision rate of 88.02%, and a recall rate of 41.75%. The performance of the deep learning model is evaluated using the mean Average Precision (mAP) of 68.19%. Deep learning models such as AlexNet [6], VGG [6], GoogLeNet [6], and ResNet [2, 4, 6, 25] are used for classifying the severity of oral diseases, and performance is validated using the accuracy rate. The AUC value of ResNet is high, with 97%, while GoogLeNet has 94%. The AlexNet and VGG nets have an accuracy of 92% and 89%, respectively. The gingivitis dataset [6] is created using smartphones such as iPhones, Samsung Galaxy, and Canon digital cameras. The images are captured from patients with their ages ranging between 14 and 60 years. These images are scaled to dimensions of 224 × 224 pixels. After performing a series of data augmentation processes such as random rotation, flipping, zooming, and shear transformation, a valid dataset with 638 images was created. Ensemble learning improves a

model's performance, but it is computationally expensive. The severe Periodontitis led to bone fractures and bone loss in the teeth, and so it is medically investigated using the periapical radiographs. Periapical radiograph (PER) is a diagnostic tool used for observing periodontal issues. Despite its usage, it is very difficult to identify small changes in bone density. The dentist's empirical manual interpretation and uncertainties in bone changes make it difficult to detect disease at an early stage. Issues in the radiographs-based analysis [7] developed due to the manual interpretation. To address the above problem, Kurt-Bayrakdar et al. [7] developed a classification model that combines CNN and U-Net architecture. This model identifies the alveolar, vertical, and horizontal bone loss due to periodontal diseases. The performance of the model is justified with the F1-score, accuracy, and AUC. This classification model achieves the highest accuracy rate (0.951) for predicting alveolar bone and the lowest accuracy rate (0.733) for vertical bone loss. Horizontal bone loss has a

moderate accuracy rate of about 0.910.

The Grad-CAM heat map visualization technique [8] is combined with AlexNet and Random Forest models for predicting Periodontitis, and obtained an accuracy of 96.8%. Apart from the deep learning techniques, this work also utilizes the Support Vector Machine (SVM), and a comparison is made between these two techniques. The SVM model's accuracy rate is 95.6%, which is slightly less than that of the CNN model. DeepLabv3+ neural network [9] is used for the detection of gingival inflammations. MobileNetV2 and Xception are used as the backbone models to classify the 567 intraoral images. The anatomical plaque and gingival tender tissues are diagnosed through the hybrid deep learning models [14-16], and obtained the sensitivity of 92% and specificity of 94%, respectively. The early detection of plaque [29, 30] will avoid unnecessary gum inflammation and tooth loss. It is mandatory to predict the earlier stages of tooth destruction and prevent the patients from destructive diseases.

Table 1. Comparing the methodologies utilized in the existing systems for predicting oral diseases

References	Oral Disease Type	Image Type/Count	Algorithm	Statistical Parameter
[1]	Periodontitis	Visible – 567 images	DeepLabv3+ neural network, MobileNetV2	Sensitivity (0.92) Specificity (0.94) Mean Intersection over Union (0.60)
[2]	Gingivitis	Visible – 134 images	Faster Region-Based Convolutional Neural Network (FR-CNN) with ResNet-50 backbone	Mean Average Precision (mAP) – 68.19%
[3]	Periodontitis	Visible – images from 1,743 Thai patients	Logistic Regression (LR)	Accuracy – 77.12% Odds Ratio (OR) – 95% Confidence Interval (CI) Area Under Curve (AUC)
[4]	Gingivitis, dental calculus and soft deposit	Visible – 3,932 oral images from 625 patients.	Convolutional Neural Network (CNN)	Gingivitis – 87.11% Dental Calculus – 80.11% Soft Deposits - 78.57%
[5]	Early gingivitis	Visible - 134 intraoral images	Faster R-CNN with ResNet-50 backbone	Accuracy -77.12% Precision – 88.02% Recall – 41.75% mAP – 68.19% Area Under Curve (AUC)
[6]	Gingivitis	Visible – 683 images	CNN with AlexNet, GoogLeNet, ResNet-50, Visual Geometry Group (VGG)	ResNet – 97% GoogLeNet – 94% AlexNet – 92% VGG – 89% Accuracy – 99.4%
[7]	Periodontal bone loss (Alveolar bone) and furcation	Gray Scale – 1121 panoramic radiographs	CNN combined with the U-Net architecture	AUC Alveolar bone loss – 0.951 Horizontal bone loss – 0.910 Vertical bone loss – 0.733
[8]	Periodontitis	Gray Scale – Peripheral Radiographs (PER)	CNN combined with AlexNet, Random Forest (RF) Support Vector Machine (SVM)	Accuracy AlexNet – 0.872 Random Forest – 96.8% SVM – 93.45%
PROPOSED METHOD	Coronal plaque, Gingivitis and Periodontitis	Fluke Thermal Image	Tuned Convolutional Neural Network with Vision Information Transformer (CNN-ViT) algorithm	Accuracy – 98.99% Precision – 97.65% Recall – 96.89%

2.1 Inference from literature survey

In the field of medical image processing, the existing research work uses hybrid classification models [2, 4, 8, 20, 24, 25]. Innovative hybrid deep learning methodologies predict the severity of the diseases. The X-ray and Computed Tomography (CT) images [7, 20, 27, 28] from the Electronic Health Records (EHR) are analyzed using the Artificial Intelligence (AI) models [13-16, 22, 23, 25, 28, 30] for classifying the tumors, fractures, ocular disease, aural disease, oral diseases, and cancer cells. Apart from disease prediction, deep learning models are used for predictive analysis, drug discovery, and robotic surgery [28, 29, 31]. The behavioral patterns of the patients, such as facial expressions, depression, and anxiety levels, are investigated through AI models [17, 19, 21, 28, 33]. The classification models, such as CNN [2, 4, 6, 8, 22, 23, 25], Support Vector Machine (SVM) [8, 13, 19], Linear and Logistic Regression models [3], are used for the classification of oral diseases. Moreover, ensemble learning is used for tuning the layers of the deep learning model, such as AlexNet [6], GoogLeNet [6], and ResNet [2, 4, 6]. The performance of the classification model is evaluated using the statistical parameters such as accuracy, Precision, Recall, F1-score, AUC, sensitivity, and specificity, which are tabulated in Table 1. Improper image quality leads to the ineffective segmentation of edges and contours of teeth and gums. The precise prediction of the deep learning classification models is low due to noise and artifacts in the RGB image. Henceforth, a detailed analysis of the existing machine learning and deep learning methodologies is conducted. The methodologies, advantages, and limitations of the proposed work are discussed in further sections.

2.2 Problem statement

The majority of the existing DL models utilize the RGB oral image dataset [1, 2, 4-6, 9] for identifying the gingivitis [10, 16, 24, 25, 28] and Periodontitis condition [3, 7-9, 12, 13, 17, 26, 28]. It is observed that the accuracy rate of a few existing classification models is poor, due to low contrast image quality [6, 9, 15, 17, 18]. Despite several pre-processing techniques, there is still an issue in improving the accuracy rate. The quality of RGB images is distorted with factors such as lighting conditions [17, 18], device calibration [4, 8], and interruption of noise signals [13, 16, 22, 25]. Few researchers have used thermal images for oral disease identification [18], and a valid thermal dataset for oral disease is yet to be formulated. It is inferred that there is a need for effective pre-processing techniques to enhance the pixel quality, which improves the performance of the model. Hence, there is a need for formulating an efficient framework for predicting oral diseases at an early stage.

2.3 Contribution

- To identify oral diseases such as coronal plaque, gingivitis, and Periodontitis at an earlier stage using the created teeth thermal image dataset.
- To reduce the thermal noise using the Gaussian filter and to enhance the pixel quality using tuned Contrast Limited Adaptive Histogram Equalization (CLAHE) and gamma correction, which are applied for teeth and gum region enhancement.
- To extract the features from pre-processed dental thermal images, using the proposed Tuned Linear Binary Pattern

(TLBP), Pelican Optimized based De-noising Convolutional Neural Network (PO-DnCNN), and Hippopotamus Optimization based De-noising Convolutional Neural Network (HO-DnCNN).

- To predict and classify coronal plaque, gingivitis, and Periodontitis oral diseases, using the proposed Tuned Convolutional Neural Network with Vision Information Transformers (CNN-ViT) algorithm. Based on the severity level of gum diseases, the prognosis is suggested to the patients.

3. METHODOLOGY

This proposed methodology classifies oral diseases such as coronal plaque, gingivitis, i.e., supragingival and subgingival, periodontitis, and healthy teeth using a created thermal image dataset. Thermal image differentiates the degradation of the gums precisely by comparing the visible images. The heat contours of the affected teeth are used to identify the severity of the oral infection.

3.1 Data collection

A total of 500 thermal images were collected for experimental purposes from Kaviya dental clinic, and they were converted to thermal images using the HIKIMICRO thermal camera. Among 500 images, 400 images are used for training, and 100 images are used for testing purposes. The teeth thermal images are captured from the predefined datasets that are available online. Apart from this, nearly 300 visible and thermal images have been collected from Mendeley [20] and Kaggle [27] online dataset is used for training the classification model. The overall process and flow of this research methodology are depicted in Figure 4.

For teeth thermal data collection and for validation of the proposed model, Dr. R.S. Kaviya, B.D.S., Dental Surgeon, Reg.No: 22049, has helped with clinical diagnosis and extended her full support for this research. As a part of the data collection totally 30 patients were subjected to testing, and they were interviewed. The patients were informed prior to, and their consent was obtained as per the ethical guidelines. This proposed study has been granted ethical clearance under the Reference Number 366/IEC/2024 by the Institutional Ethics Committee of Kaviya Dental Clinic, Chennai, Tamil Nadu, India. This proposed study utilized a content analysis approach to categorize the clinical and patient data, later used for classification of the gingivitis, periodontics, and plaque with the proposed classification model. The interviews were conducted by the treating doctor, with whom patients had an existing clinical relationship.

Participants were briefed about the research proposal, and proper consent was obtained. The interviewer (doctor) is a qualified medical professional with experience of 8 years in dental patient's care and clinical research. This doctor had a clinical interest in improving the health status of the patients. Participants were selected using consecutive sampling, where all the eligible patients visited during the clinical study period were included with their consent. Since the doctor – patient relationship was smooth due to the trust, all the patients agreed to participate in this study after signing the consent form. To ensure privacy and confidentiality, the interview was conducted exclusively between the doctor and patients without any intervention of a third party. The dataset has been collected through face-to-face, semi-structured interviews,

which were carried out in a clinic.

Among 30 patients, 5 were healthy, 10 were affected with plaque, 10 were infected with gingivitis, and 5 were affected with periodontitis. There were 15 men and 15 women in the age group between 18 and 60 years. A semi-structured interview with predefined sample questions was created with the joint effort of the authors and the clinical doctor. Each patient was interviewed (10–15 minutes) once during their clinical visit, and their consent was obtained for only picturing their teeth with a thermal camera. The verbal conversation was not recorded, and only the thermal images were acquired for our research study. We observed data saturation by the 25th patient, where no new symptoms were identified. This proposed study used the collected thermal images for earlier diagnosis of gum diseases using the proposed classification CNN - ViT model.

3.2 Teeth thermal image pre-processing

The acquired thermal images are pre-processed to enhance the pixel quality and perform the analysis. Pre-processing of teeth thermal images eradicates the noisy pixels and improves the proposed model’s classification accuracy rate. In acquired thermal images, the histogram equalization is applied and highlights the subtle features around the teeth and gums. Thermal image pixels are visualized in terms of red (hot regions) and blue (cold regions) colors. There are errors in the distribution of pixels due to the sensors. To rectify it, thermal images are subjected to the CLAHE process that tunes the image contrast as shown in Figure 5. The unevenly distributed pixels are focused and neutralized to improve the visibility of

the image features. Further to observe the degradation of teeth and gums, the proposed pre-processing CLAHE method is applied to thermal images and corrected using gamma correction. It equalizes the contrast and brightness level of the pixels and improves the quality of the teeth thermal image. The medical images have noise signals due to the patient’s movement and electronic interference.

The variation of the image pixels between the original thermal images (Figure 6(a)) and the noisy thermal image is as in Figure 6(b). and their corresponding histogram are shown in Figure 6. Each bin of the histogram represents a particular range of pixel intensity, and its height represents how many times that particular range of pixel are present in the teeth thermal image. When comparing these three histograms, it is clear that the noisy thermal image has an uneven distribution of pixels compared to the original and de-noised image, as shown in Figure 6(c). To analyze the pixel distribution, the gum regions are chosen across three images, and their pixel intensity profiles are compared. The intensity of the pixels is compared across the original input image, noisy image, and pre-processed image, and shown in Figure 7. In Figure 7, the ‘x-axis’ represents the pixel position in the selected region and the ‘y-axis’ represents the corresponding intensity value. This approach is used to observe the changes in the patterns, texture, and contrast of the image. It is inferred that the noisy image has an uneven distribution of pixels when compared to the original image. However, the de-noised image has a pixel variation similar to that of the original image. Hence, it is evident that the proposed pre-processing method is suitable for thermal image processing.

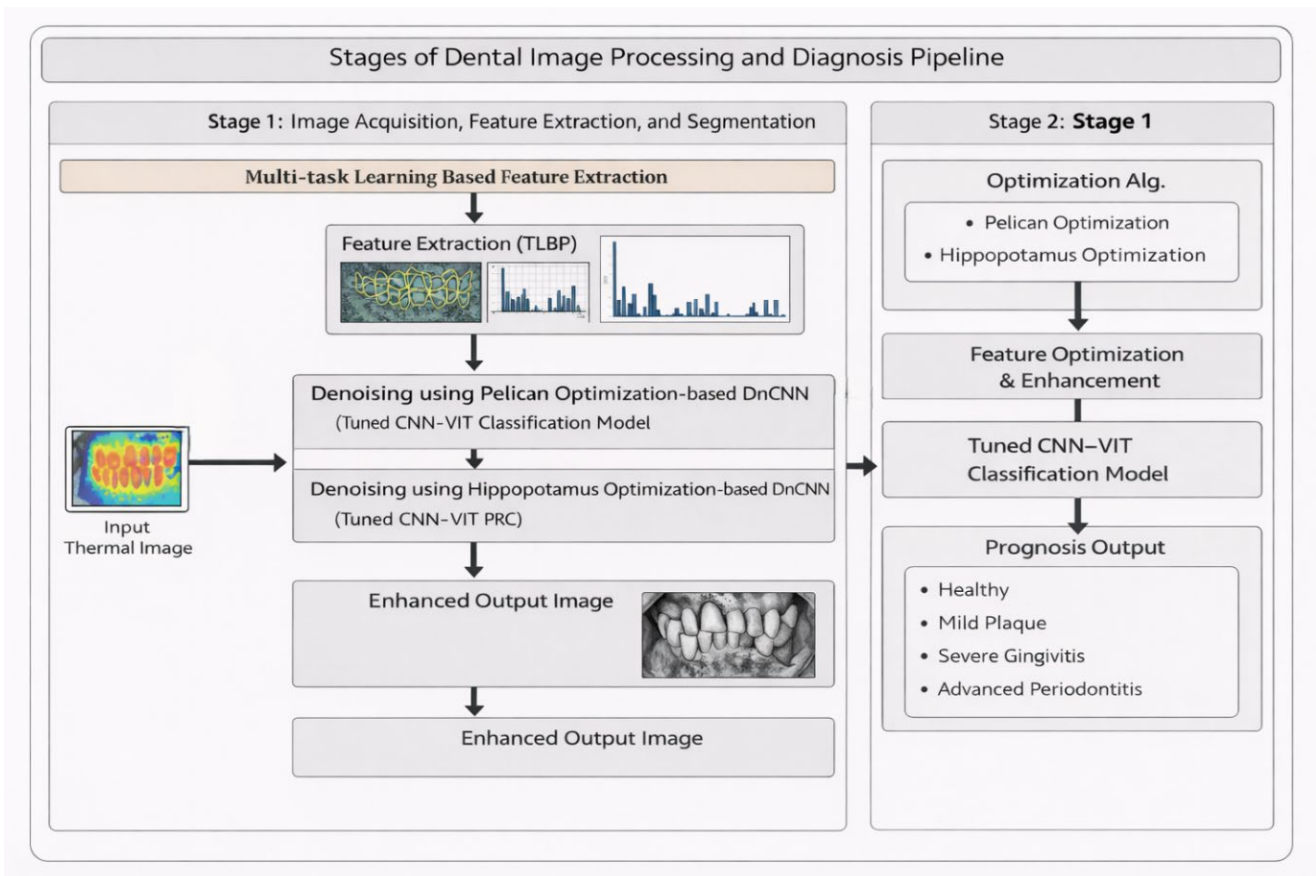


Figure 4. Architecture of the proposed model for classifying the coronal plaque, gingivitis, and Periodontitis using thermal image

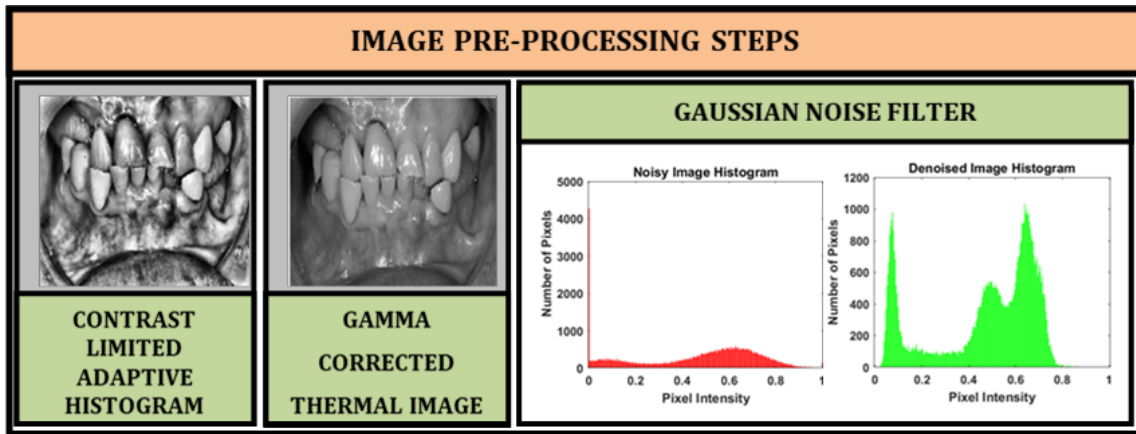


Figure 5. Proposed pre-processing methods using Contrast Limited Adaptive Histogram Equalization (CLAHE), gamma correction, and a Gaussian filter

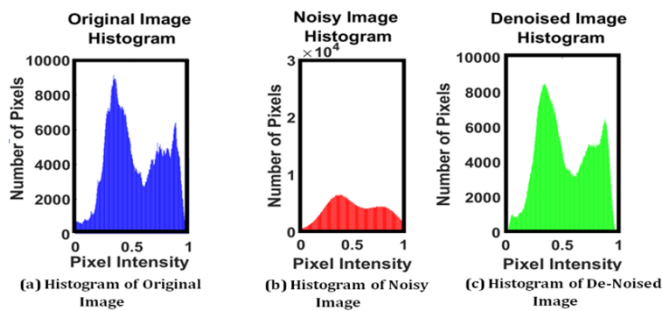


Figure 6. Analyzing the variation of image pixels using the histogram

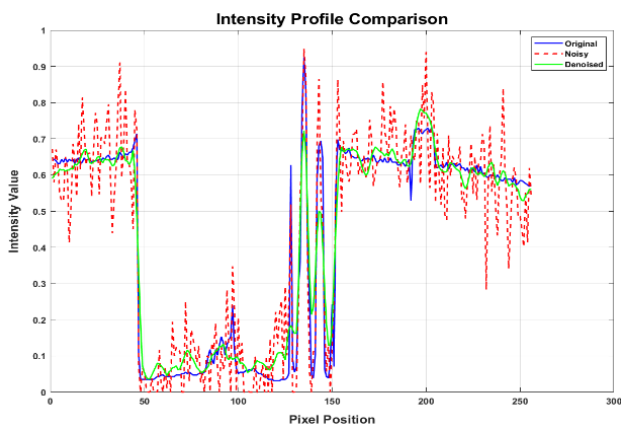


Figure 7. Intensity profile comparison of the original, noisy, and de-noised image

3.3 Image feature extraction and segmentation

The pre-processed thermal images have enhanced distinct features that distinguish the object from the background. The features, such as edges and contours of the teeth and gums, are extracted efficiently to classify the diseases. In this proposed work, the features from the pre-processed thermal image are extracted using the TLBP Feature Extraction algorithm. The TLBP algorithm performs effectively for teeth thermal images because it is rotation invariant, highlights, and distinguishes the subtle patterns. The proposed TLBP algorithm compares the neighboring pixel intensity and analyzes its pattern in binary format. If the intensity of the neighboring pixel (I_p) is greater than or equal to the central pixel's intensity $I(x, y)$ then that region is considered as a brighter region (1). In case it is dark and less than the central pixel, the value is assigned as '0'.

The mathematical representation for this TLBP is given in Eq. (1).

$$LBP_{P,R}(x, y) = \sum_{p=0}^{P-1} s(I_p - I(x, y)) \cdot 2^p \quad (1)$$

The intensity of the center pixel $I(x, y)$ is compared with neighboring pixel I_p and further undergoes step function $s(x)$ in which the brighter region is assigned as '1' and the darker region is assigned as '0'. Finally, the regions of interest across thermal images are combined in the form of a histogram as shown in Figure 8.

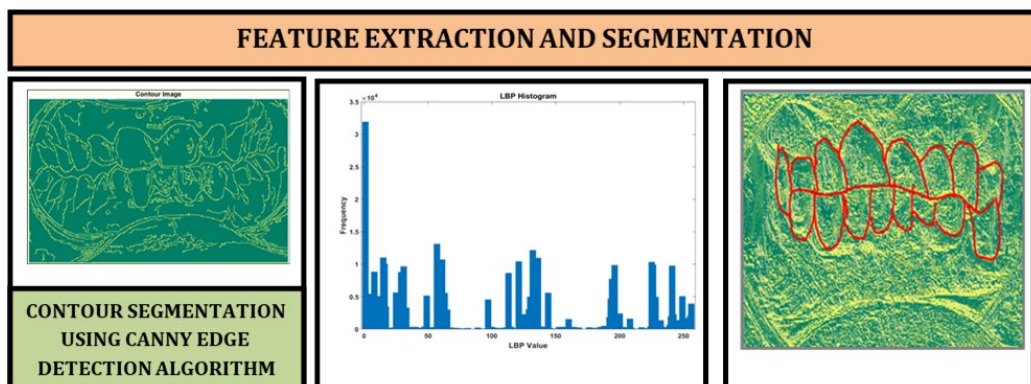


Figure 8. Tuned Linear Binary Pattern (TLBP) based feature extraction and Canny edge-based contour segmentation

Table 2. Tuning and analyzing the parameters of the linear binary pattern

Radius (R)	Neighbors (P)	Accuracy Rate (%)	Effect on Texture Detail	Computational Complexity	Best Fit
1	8	96.78	Localize fine details	Low	Good for fine textures
2	8	94.5	Slightly focus on coarser details, comparing radius = 1	Moderate	Suitable for large features
2	16	92.45	The fine details are balanced with a high spatial domain	Higher	Perfect for small to medium patterns
3	8	88.5	Focus broadly, compromising the details of features	Moderate	Effective for complex patterns in a higher spatial domain
3	24	85.5	High-resolution features	High	Best for complex and large-scale features

In this proposed TLBP method, the LBP parameters such as radius (R) and number of circularly symmetric neighbors (P) are tuned, and the features are extracted precisely. These parameters are tuned, and their observed values are tabulated in Table 2. It is inferred that the alignment of the pixel with its corresponding grid location is not possible with the higher R value. Therefore, it requires the bilinear interpolation process, which is computationally expensive. The larger P value needs more neighbor comparisons; therefore, it results in a longer binary encoding process. Considering all the above factors, radius R is chosen as ‘1’ and the number of neighbors P is chosen as ‘8’ in this proposed TLBP method. The Tuned LBP converts the resultant binary patterns into feature vectors that denote the texture of the teeth image. Though it extracts the meaningful information from thermal images, particular features are used for the classification of gum disease. The remaining non-important features may introduce noise during the classification of gum disease. To avoid this, the extracted feature vectors are further optimized using the Pelican Optimization Algorithm (POA) and hippopotamus optimization algorithm (HOA). The above algorithm searches the feature space and focuses on the promising areas in the feature vector. These methods narrow down the image features by filtering out the redundant and less important details. The obtained LBP feature vector is given as input to both the POA and HOA. The essential texture information is selected accurately through several iterative optimizations. The performance of both POA and HOA is shown in Figure 9 as follows. The pixels are distributed as in Figure 9(a), and the important features are extracted using the POA algorithm. The HOA algorithm has uneven pixel distribution as shown in Figure 9(b).

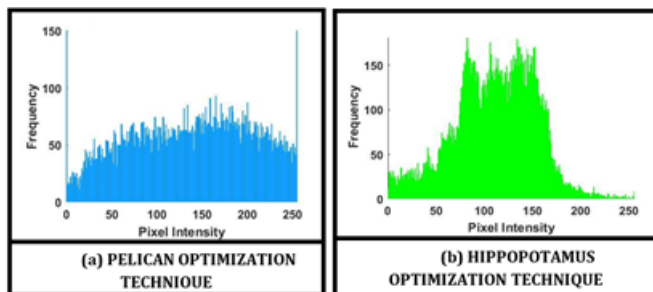


Figure 9. Comparison of the performance of the Pelican Optimization Algorithm and the hippopotamus optimization algorithm (HOA) using a histogram

The POA algorithm selects 56 features from the input thermal image and predicts gingival infection in 4.21 seconds. The HOA algorithm selects 73 features from the thermal

image and predicts the infection in 6.87 s. The prediction accuracy rate for POA is 93.5%, whereas 90.2% for the HOA method. It is evident that the prominent features are selected better using POA comparing the HOA in this proposed work. When comparing the HOA and POA, the POA balances the precision and speed, focusing on the critical features. Hence, POA is preferred for this proposed study.

3.4 Pelicon optimization algorithm

The POA algorithm selects the most discriminative features from the teeth image and reduces the dimensions. LBP algorithm extracts a high-dimensional feature vector, which makes the classification harder. The LBP algorithm acquires the textures and temperature variations in the obtained teeth image. However, it generates high-dimensional feature vectors that contain redundant or irrelevant thermal features, which increases the computational complexity. In order to improve the classification model’s performance, the POA algorithm is used sequentially after the LBP feature extraction process, as in Eq. (2).

$$F (LBP \text{ features}) = \{f_1, f_2, \dots, f_n\} \quad (2)$$

In Eq. (2), let f_1, f_2, \dots, f_n be the LBP feature vector extracted from the input thermal teeth image. Similarly, the POA features are extracted from the input thermal image using Eq. (3).

$$(POA \text{ features}) = \{p_1, p_2, \dots, p_m\} \subseteq F \quad (3)$$

where, $m < n$.

In this proposed work, each pelican in the POA algorithm denotes a binary feature mask X_i as in Eq. (4), where:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\} \quad (4)$$

If $x_{ij} = 1$ then it denotes that j^{th} LBP feature is selected from thermal image. Else if $x_{ij} = 0$ then that particular irrelevant feature is discarded. This process is repeated until the fitness of each pelican is evaluated and its positions are updated through the optimization process. Once the optimal features are obtained, they are masked using the sigmoid function, and it is given in Eq. (5).

$$S(x) = \frac{1}{1 + e^{-x}}, \text{ then } x_{ij} = \begin{cases} 1, & \text{if } S(x) > rand() \\ 0, & \text{Otherwise} \end{cases} \quad (5)$$

In this proposed method, the POA algorithm efficiently handles high-dimensional LBP feature vectors by reducing the

redundant and noisy features. The extracted POA features are passed into the Canny edge detection model, which detects the contour of the teeth from the thermal image. The edge detector primarily focuses on the infected gingival region and segments the gum region for classification. Canny detector enhances the strong gradients and de-noises the irrelevant pixels. Canny edge detection is superior to other algorithms because it predicts even the prominent edges using the gradient intensity and orientation. As a result, the boundaries and structures of human teeth and gums are visualized clearly and avoid false

edge identification. It focuses on both the dark and bright regions of edges and defines the boundary effectively without any compromise, as shown in Figure 10. The Gaussian filters are used as the pre-processing method prior to the Canny edge segmentation process. It helps to smooth the thermal image features and reduces noise. The Canny edge detector employs both upper threshold and lower threshold values for focusing on the prominent features of teeth. These gradients are essential in edge-linking steps for determining the edges and contours of the gingiva.

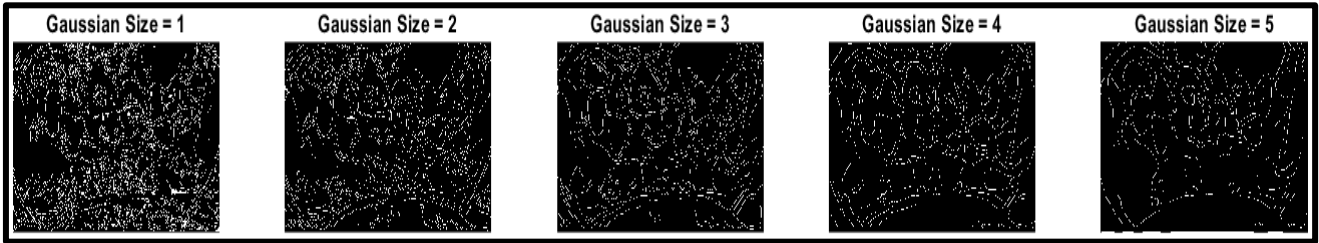


Figure 10. Effect of tuning the Gaussian filter size on edge detection using the Canny method

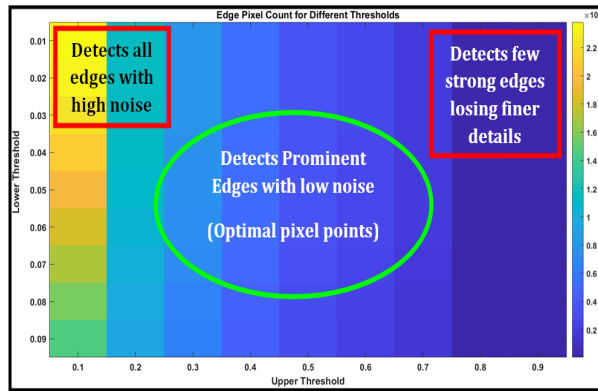


Figure 11. Heat map visualization of the Canny edge detector for extracting features from the gingivitis thermal image

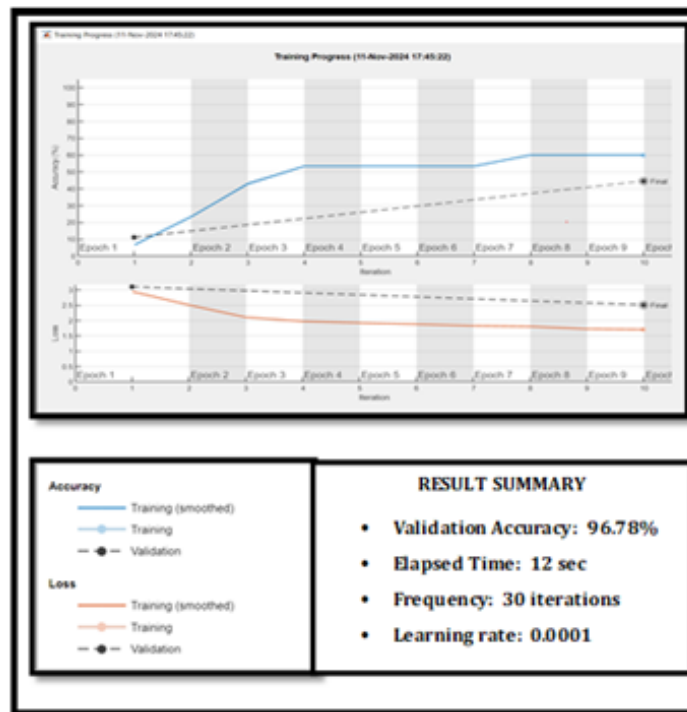


Figure 12. Training performance of the proposed Convolutional Neural Network with Vision Information Transformers (CNN-ViT) algorithm

The Gaussian filter size is tuned up to ‘value = 5’ and enhances the edges and boundaries (Figure 10). Further, the edge pixel count of dental thermal images is computed subject to different threshold values ranging from ‘0.1’ to ‘0.9’. This pixel count plays an important role in defining the contours of a pre-processed image. Figure 11 provides an insight into the allocation of pixel count with its respective threshold value. This heat map elaborates the pixel allocation and helps to determine the position of the meaningful features with less noise. The yellow regions (left corner) have more texture details and noise, whereas the blue or purple regions have low noise with few edge points (right corner). Hence, the Canny edge detection algorithms and Gaussian filters identify the important edges and contours for the classification of the coronal plaque, gingivitis, and Periodontitis.

Flow of the proposed CNN-ViT classification algorithm.

Pseudocode for the Proposed CNN-ViT Algorithm
BEGIN
STEP 1: Load the dataset
STEP 2: Split the dataset into Testing (80%) and Training (20%) datasets
STEP 3: Normalize the image size to ‘[224,224,3]’
STEP 4: Data Pre-processing and Augmentation (CLAHE, Rotation, Translation)
STEP 5: Extract features using Local Binary Pattern (LBP)
STEP 6: Use the Pelican Optimization Algorithm (POA) for selecting the optimized feature set
STEP 5: Load the ResNet-50 model for CNN classification
STEP 6: Train the CNN model with the optimized feature dataset
STEP 7: Create Vision Transformer block
7.1 Apply Patch embedding ($16 \times 16 \times 3$){Height, Width, Channel}
7.2 Add Positional encoding
7.3 Apply Transformer Blocks (Multi-Head Self Attention (MHSA))
7.4 Layer Normalization
7.5 Feed Forward Network (FNN)
7.6 Flatten and pass through MLP head
STEP 8: Train both CNN and ViT models on training data
STEP 9: Test the image using both the CNN and the ViT classification models
STEP 10: Display the predicted class of the thermal image using the softmax classifier
STEP 11: Compute the statistical metrics accuracy, precision, recall, and F1-Score
END

3.5 Image optimization and classification

In this proposed work, the extracted contours are analyzed using the hybrid Convolutional Neural Network with Vision Information Transformers (CNN-ViT) for gum disease classification. The CNN model extracts the low-level features such as corners, contours, edges, and textures from the segmented thermal image. The convolution and pooling operations detect small abnormalities in the gum regions of teeth. In contrast, ViT models are capable of capturing the global spatial relationship by understanding the structure of the mouth and teeth. Once the local features are extracted using the CNN, it is passed to the ViT layers, and the images are segregated into patches. From continuous analysis, the patterns across the mouth regions are learned, and this process is repeated until it reads all the regions of the mouth. The

combination leverages the strength of the classification model and enhances the medial image processing. The overall working of the proposed CNN-ViT model is briefed using algorithmic steps in the section below.

This proposed hybrid model focuses on both the local and global features without any compromise. Initially, the pre-processed image is divided into 196 patches. Patches are flattened linearly and reduces its dimension and convert into 1D vector. This 1D vector retains the essential information without any loss, and it is embedded as a discrete token. Each patch of the image is considered as a discrete token and fed as input to the transformer layers. The Multi-Head Self Attention (MHSA) mechanism observes the dependencies between the tokens to order the discrete tokens. The heads of the MHSA are tuned up to 12, and inferred that it requires more training and inference time. So, the number of heads is assigned as ‘8’ and its respective attention score is calculated using the dot product. After this process, a residual connection is created to extract the information from previous layers, and the outputs are standardized using the layer normalization parameters (mean & variance). The vanishing gradient problem is addressed using residual connections, whereas the internal covariate is stabilized with the normalization process. Each token is subjected to the non-linear transformations through a Feed-Forward Neural Network (FFN) and enhances the performance of the model. The Rectified Linear Unit (ReLU) activation function is used in this network because it is nonlinear, computationally efficient, and mitigates the vanishing gradient problem. The performance of the proposed CNN-ViT algorithm during the training process is shown in Figure 12. The convolution layers study the features effectively with a learning rate of 0.001% and achieve an accuracy rate of 96.78% approximately. The learning process continues for 30 iterations, and finally, the classification layer classifies the types of oral disease. Its validation is done using statistical metrics.

4. RESULTS AND DISCUSSION

The thermal images are acquired at a standardized clinical imaging setup using a Fluke VT02 thermal camera. Data from patients with varying levels of oral disease severity are collected with the help of Dr. R.S.Kaviya, Dental Surgeon (Reg.No: 22049). To handle the class imbalance problem, the dataset was augmented using different processes such as rotation, flipping, and translation. Data have been balanced to prevent model bias between majority classes. The training dataset is increased by using the data augmentation techniques, through flipping, i.e., Horizontal or vertical rotation (0° - 90°), translation, and scaling (zoom in and out). Initially, there were 500 images for training; later, it was expanded to 7000 images. Moreover, 300 visible and thermal images have been collected from Mendeley [20] and Kaggle [27] online dataset is used for training the classification model. The proposed model’s performance is verified and validated using the statistical metrics such as Accuracy, Precision, Recall, and F1-Score, and the ROC curve. In this paper, multi-task analysis is done to ensure the correctness of the predicted results in all aspects. True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are identified from the confusion matrix of the proposed classification method. The accuracy determines the overall correctness of the method and is calculated as in Eq. (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

The precision identifies the true positives among all the positive predictions as in Eq. (7).

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall predicts the true positive values among the correctly predicted values as in Eq. (8).

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

The precision and recall value are correctly balanced using the F1-score, and it is mathematically calculated as in the following Eq. (9).

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

Relying on the above equations' precision, recall, and F1-Score of the proposed Tuned CNN-ViT classification model is calculated and shown in Figure 13. Among them, Periodontitis is predicted well at a rate of 98.5% approximately.

The visualization of gum degradation and infection is visible using thermal images, and increases the classification accuracy. Secondly, the healthy teeth prediction rate is 98% approximately. Classifying plaque and gingivitis has a minor confusion since they both have similar symptoms. To correct this issue, the model's learning rate is improved through tuning the weights of the proposed ViT model during the backpropagation process. Later, its performance is observed by computing the Mean Squared Error (MSE) value and error histogram as shown in Figures 14 and 15.

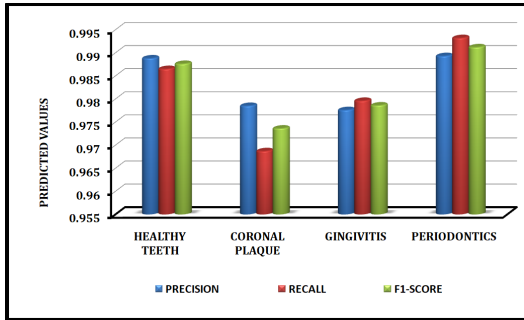


Figure 13. Variations in statistical metrics for classifying the oral diseases using proposed model

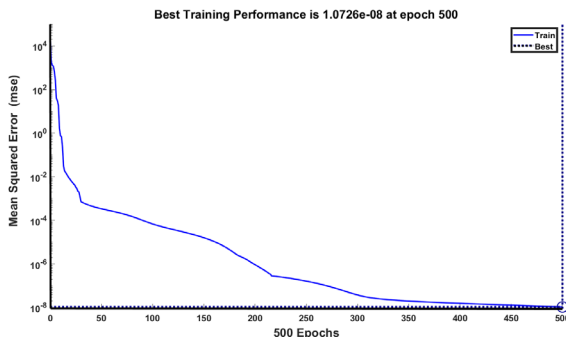


Figure 14. Mean Squared Error plot for the training process

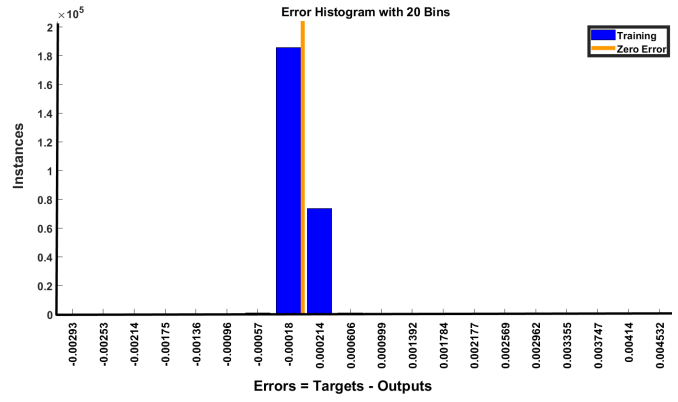


Figure 15. Error histogram of proposed Convolutional Neural Network with Vision Information Transformers (CNN-ViT) classification model

The training loss of the proposed CNN-ViT model is visually represented using the MSE Value. In Figure 14, the blue line decreases with each epoch during the learning process, and therefore, it leads to the low MSE value. A low MSE value implies that the proposed classification model's learning rate is improved. The error rate for the proposed classification model is depicted as an error histogram in Figure 15. It is observed that the X-axis represents the error difference between the dental thermal image and the predicted image. The Y-axis represents how many predictions fall under that particular histogram bin. The orange line is the threshold point where there is no difference between the original and predicted image. The better model has the distribution closer to the orange line. Henceforth, it is proven that the proposed CNN-ViT classification model has a high positive prediction rate and a low error rate.

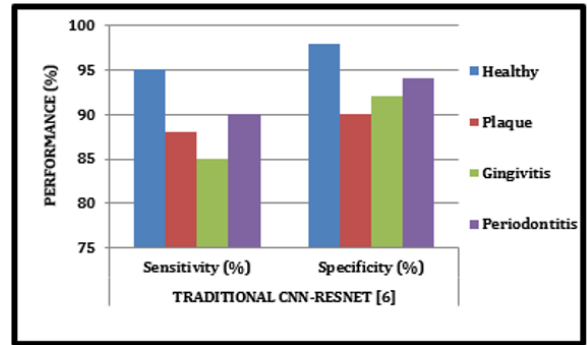


Figure 16. Statistical analysis of traditional CNN-RESNET methodology

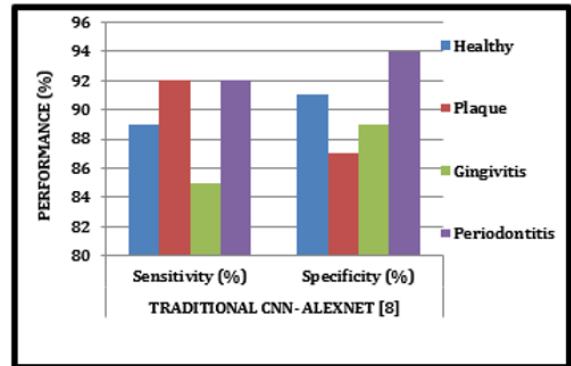


Figure 17. Statistical analysis of traditional CNN-ALEXNET methodology

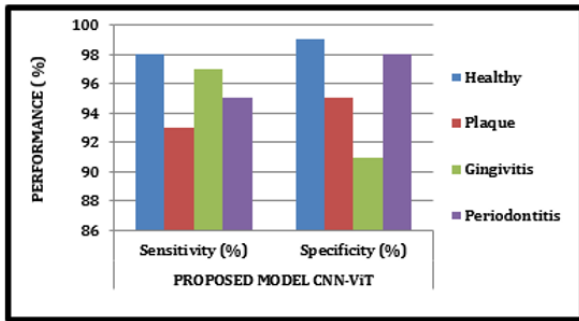


Figure 18. Statistical analysis of proposed tuned CNN-ViT methodology

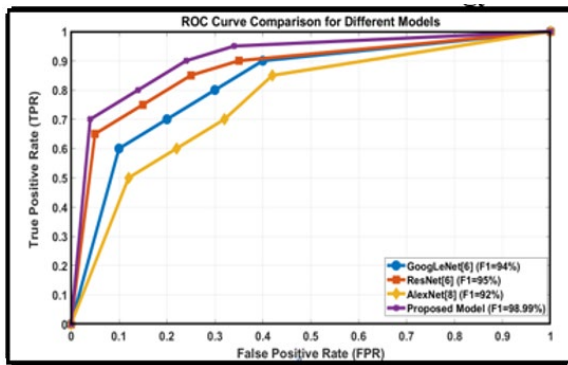


Figure 19. Receiver Operating Characteristics (ROC) curve comparison between existing methodologies and proposed system

The statistical analysis of the traditional CNN-RESNET, CNN ALEXNET, and the proposed CNN-ViT model is shown in Figure 16, Figure 17, and Figure 18, respectively. From the figures, it is proven that the True Positive Rate (TPR) of the proposed model is high, i.e., 98.5%, when comparing the traditional models such as CNN-RESNET (accuracy = 95.5%) and CNN-ALEXNET (accuracy = 94.5%). To enhance the performance of the proposed classification model, the Receiver Operating Characteristics (ROC) curve is estimated and shown in Figure 19. ROC curve is plotted across TPR and False Positive Rate (FPR), which provides insight about the model’s capability, at different stages of oral disease. It also reduces the false diagnosis through balancing the sensitivity and specificity of the proposed model. The ROC curve is plotted by calculating the sensitivity and specificity using the following Eqs. (10) and (11).

$$\text{Sensitivity (TPR)} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (10)$$

$$\text{Specificity (FPR)} = \frac{\text{False Positives (FP)}}{\text{False Positives (FP)} + \text{True Negatives (TN)}} \quad (11)$$

The ROC curve of the proposed Tuned CNN-ViT classification model is compared with the existing systems. It is reported that the proposed model outperforms the other predefined models and classifies oral diseases with a performance rate of 98.99% approximately using thermal images. The other pre-trained models, such as GoogLeNet, ResNet, and AlexNet, have the performance rate of 94%, 95%,

and 92%, respectively. They have lower rates compared to the proposed CNN-ViT model, mainly due to the poor image quality and inefficient pre-processing model. Henceforth, the proposed classification model functions perform better due to the proper balanced dataset, effective pre-processing, and optimization techniques. This proposed model not only classifies the oral disease but also provides the prognosis for identifying its severity level. Based on the classification results, the stage of the tooth infections is derived as given in Table 3.

Table 3. Prognosis derivation for oral diseases based on the output of the proposed classification model

Teeth Condition	Status	Prognosis
HEALTHY TEETH	GOOD	<ul style="list-style-type: none"> Excellent prognosis. Maintain regular oral hygiene and routine check-ups every 6 months.
CORONAL PLAQUE	MILD	<ul style="list-style-type: none"> Fair prognosis Requires intensive cleaning, scaling, and antibacterial treatments at home.
GINGIVITIS	SEVERE	<ul style="list-style-type: none"> Fair to guarded prognosis Scaling, root planning, surgical interventions, and lifestyle change are compulsory.
PERIODONTICS	ADVANCED	<ul style="list-style-type: none"> Guarded to poor prognosis Aggressive treatment and regular follow-ups required, including complex surgery and tooth loss

The reliability of the Proposed Tuned CNN-ViT model is validated through standard datasets obtained from sources such as Mendelely [20] and Kaggle [27], and its values are plotted in Table 4. To validate the stability of the proposed CNN-ViT model statistically, K-fold validation is applied in the CNN layers, and the value of ‘k’ is assigned as ‘5’. The mean training and validation accuracy rates are around 98.34% and 97.62%, respectively, as shown in Table 5. The model learns the data well and improves the training process in the proposed CNN-ViT method. Among the 5 folds, Fold 3 (98.5%) has the highest validation accuracy, and Fold 4 (99.7%) has the highest training accuracy. Fold 1 has the lowest training and validation accuracy because it fails to focus on the deeper layers. As the process continues, the accuracy rate is gradually improved in the upcoming folds. Moreover, the proposed CNN-ViT model’s performance is validated across the clinical diagnosis by the dentist and shown in Table 6.

Table 4. Statistical comparison of the proposed tuned CNN-ViT model across existing methodologies

Dataset	Accuracy (%)	Precision	Recall	F1-score
Thermal Dataset [Ours]	98.78	0.9823	0.9698	0.9760
Mendelely [20]	96.67	0.9674	0.9578	0.9626
Kaggle [27]	95.94	0.9543	0.9365	0.9453

Table 5. Training validation of the proposed CNN-ViT model across 5 folds

Fold	Training Accuracy (%)	Validation Accuracy (%)
Fold 1	96.7	95.6
Fold 2	98.5	97.8
Fold 3	97.9	98.5
Fold 4	99.7	97.9
Fold 5	98.9	98.3
Mean	98.34	97.62

Table 6. Statistical comparison between the proposed model and clinical diagnosis

Metric	Proposed Model	Clinical Diagnosis @ Kaviya Dental Clinic, Chennai, India
Accuracy	98.95%	97.43%
Precision	96.53%	98.43%
Recall	96.67%	97.65%
F1-Score	96.14%	98.04%

The accuracy of the proposed CNN-ViT model is 1.52% is higher than the clinical diagnosis. The precision rate is high in both the clinical and proposed CNN-ViT method, which implies that gum diseases are predicted and classified accurately. However, the recall rate of the proposed model is 0.98% lower compared to the clinical diagnosis because the model gets confused between plaque and gingivitis. This is improved by training the model with more real-time datasets in the future. Moreover, the F1-score of both clinical diagnosis and the proposed model is about 98% approximately. Despite of this, the clinical method requires some second opinion while classifying the depth of the infected teeth pockets. Clinical diagnosis to perform using the probing method sometimes fails to observe the deep infected teeth's gingiva because the dentist cannot visualize through the naked eye. The proposed thermal image classification is considered as the second opinion for the precise prediction.


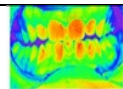

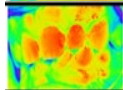

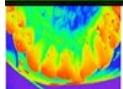

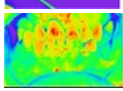
From Table 7, it is inferred that the presence of pre-processing methods has improved the classification accuracy of oral diseases. Moreover, a comparison of execution time between the standard methods, such as Visual examination, Probing, and X-rays, and the proposed methodology is

conducted. The standard method requires more time to classify the oral diseases compared to the proposed tuned CNN-ViT classification method. Therefore, the robustness of the proposed classified model is justified with both the accuracy rate and the execution time. The classification of oral diseases using the traditional method involves a series of processes such as visual examinations, probing, palpation, and Scans (X-ray & CT-Scan). A sample of patients' photos is collected manually, and the accuracy of the prediction rate of the traditional method is compared with the proposed CNN-ViT classification model, as shown in Table 8. The accuracy rate for the traditional method is as similar as the proposed model, yet there is a mild deviation for predicting the gingivitis and coronal plaque. Both plaque and gingivitis have a similar yellowish buildup at the initial stage that leads to confusion when diagnosed manually. Gingival forms at the gum line of teeth, whereas plaque accumulates on the tooth crown. Hence, the proposed model has a high accuracy rate (97.89%) for differentiating plaque and gingivitis compared to the traditional method (95%). The proposed model is used by the dentist as a second opinion to classify the plaque and gingivitis.

The dentists used the dental mirror and plaque disclosing dye solution (Erythrosine) for predicting the severity of the teeth disease. Healthy teeth have no stains, and teeth with plaque have a bright red or pink color near the gum margins. The gingivitis teeth have soft tissues with an intense red color. Deep red stains occur around the gum pockets between the teeth due to Periodontitis infection. Thus, the patients were tested clinically as per the doctor's advice.

Apart from clinical diagnosis, the patient's teeth were acquired using a HIKIMICRO thermal camera, subject to different criteria as shown in Table 9. The MSX feature of thermal camera blends visible light with the thermal image; therefore, it analyzes the minute variation irrespective of environment and lighting conditions. Mainly, it was useful while predicting the depth of the teeth pockets in periodontic disease. The HIKIMICRO thermal camera is mounted on the Real Me Narzo 70 smartphone, and the patients were pictured by subjecting to the above criteria. The sample real-time implementation set up is shown in Figure 20. All the patients were informed about this research study. Hence, the findings were consistent with the clinical thermal image data, and the proposed CNN-ViT classification model classifies the plaque, gingivitis, and periodontitis more accurately.

Table 7. Impact evaluation of pre-processing techniques in the proposed oral disease classification model

Patients Teeth Image Condition	Visible Image	Thermal Image	Without Pre Processing (Accuracy Rate)	With Pre-Processing (Accuracy Rate)	Standard Prediction Method (Time in Seconds)	Proposed Classification Model (Time in Seconds)
Healthy			0.9045	0.9987	60	5
Plaque			0.8997	0.9632	75	6
Supra Gingival			0.8587	0.9785	80	6
Sub Gingival			0.9356	0.9812	90	7


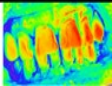
Periodontics			0.9432	0.9854	95	8
--------------	---	---	--------	--------	----	---

Table 8. Comparison of accuracy rates between traditional methods and the proposed CNN-ViT Model for oral disease classification

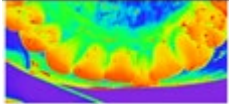
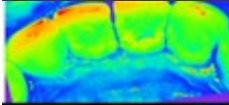
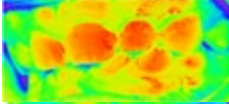
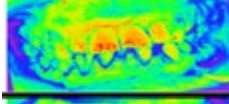
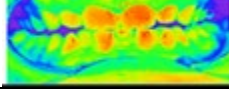
Patient Teeth Photos	Oral Disease Type	Traditional Method Accuracy (%) (Probing)	Proposed Method Accuracy (%) (Non Invasive Method)
	Gingivitis (Severe)	97	98.43
	Gingivitis (Severe)	95	97.89
	Coronal Plaque (Mild)	98	97.68
	Periodontics (Advanced)	99	97.98
	Healthy Teeth (Good)	99	98.99

Table 9. Experimental criteria for image acquisition using fluke VT02 thermal camera

Inclusion Criteria	Exclusion Criteria
<ol style="list-style-type: none"> 1. Camera positioning: Fluke thermal camera is fixed at a distance of 50 cm \pm 20°–30° inclination, focusing on the infected regions in the frontal view (perpendicular to mouth). 2. Patient’s mouth position: Maximum comfortable opening with less air flow. The excessive air flow causes temperature fluctuations in the thermal camera. 3. Age limitation: 18–60 years. 4. Room temperature: Patients were seated in an indoor air-conditioned environment (20°–28°C). 5. Appropriate consent form is collected from the patient for imaging and data use. 	<ol style="list-style-type: none"> 1. Patients with recent dental surgeries within 6 months are excluded. 2. Diabetic patients are excluded from the experiment. 3. Heavy smokers and frequent alcohol users are excluded, as it highly impacts the thermal imaging results. 4. Patients undergoing antibiotics are also excluded due to differences in their inflammatory response. 5. Motion-blurred images are removed if they affect the classification rate.

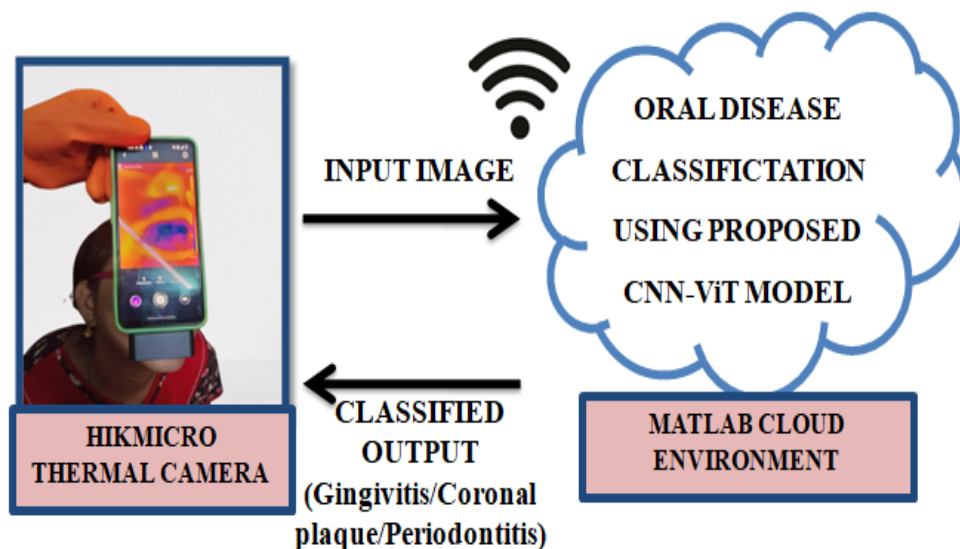


Figure 20. Real time integration of the proposed model into the clinical diagnosis

Table 10. Sample experimental results for 10 patients using the clinical diagnosis and proposed CNN-ViT model using thermal images

Patient ID	Patient's Condition	Clinical Diagnosis / Classification (Probing Method)	Proposed CNN-ViT Model	Clinical Diagnosis Accuracy Rate (%)	Proposed CNN-ViT Model (%)
Patient 1	Gingivitis (Severe)	✓	✓	96.78	97.65
Patient 2	Coronal Plaque (Mild)	✓	✗	97.34	89.54
Patient 3	Periodontitis (Advanced)	✓	✓	97.84	98.13
Patient 4	Gingivitis (Severe)	✓	✓	98.34	98.04
Patient 5	Healthy	✓	✓	98.65	98.76
Patient 6	Healthy	✓	✓	98.99	99.02
Patient 7	Gingivitis (Severe)	✗	✓	97.78	97.89
Patient 8	Coronal Plaque (Mild)	✓	✓	96.38	97.29
Patient 9	Periodontitis (Advanced)	✗	✓	88.65	98.74
Patient 10	Healthy	✓	✓	98.95	98.31

Table 11. Comparison of the proposed and existing methods

Methods	Accuracy (%)	Precision	Recall	F1-Score
ViT (Base) [33]	96.5 ± 1.0	0.950 ± 0.010	0.940 ± 0.012	0.945 ± 0.011
DeiT [34]	97.2 ± 0.8	0.960 ± 0.008	0.960 ± 0.009	0.960 ± 0.008
Swin Transformer [35]	98.2 ± 0.6	0.970 ± 0.006	0.980 ± 0.007	0.975 ± 0.006
PVT (Pyramid Vision Transformer) [36]/[40]	97.1 ± 0.9	0.960 ± 0.009	0.970 ± 0.010	0.965 ± 0.009
CvT [37]/[41]	97.4 ± 0.8	0.970 ± 0.007	0.960 ± 0.008	0.965 ± 0.007
T2T-ViT [38]/[42]	97.7 ± 0.7	0.970 ± 0.007	0.970 ± 0.007	0.970 ± 0.007
CrossViT [39]	97.9 ± 0.7	0.970 ± 0.006	0.970 ± 0.006	0.970 ± 0.006
Proposed CNN-ViT	98.99 ± 0.4	0.9765 ± 0.005	0.9689 ± 0.006	0.9726 ± 0.005

Table 12. Ablation study of the proposed and existing methods

Configuration	Preprocessing (CLAHE+γ+Gauss)	TLBP (R = 1, P = 8)	POA Selection	Canny Contours	ViT Head	Accuracy (%)	Precision (%)	Recall (%)
Baseline: ResNet-50 raw thermal	✗	✗	✗	✗	✗	93.2	91.0	90.4
+ Preprocessing only	✓	✗	✗	✗	✗	95.8	94.1	93.7
+ TLBP texture	✓	✓	✗	✗	✗	96.8	95.4	95.0
+ POA (vs. no opt.)	✓	✓	✓	✗	✗	97.6	96.3	96.1
+ HOA baseline	✓	✓	HOA	✗	✗	96.1	94.8	94.2
+ Canny contours	✓	✓	✓	✓	✗	98.1	97.1	96.8
ViT-only (preproc only)	✓	✗	✗	✗	✓	96.9	95.8	95.4
CNN-only full pipeline	✓	✓	✓	✓	✗	98.2	97.2	96.9
Full Tuned CNN-ViT	✓	✓	✓	✓	✓	98.99	97.65	96.89

From the new samples, experimental results of 10 patients are shown in Table 10. In clinical diagnosis, the periodontic patient was misclassified as having gingivitis because the deep pockets in the infected teeth are not visualized through the probing method. This drawback is addressed by the proposed CNN-ViT method using thermal images. Similarly, the proposed model misclassifies coronal plaque as gingivitis due to the coexisting symptoms between plaque and gingivitis. The training dataset contains highly correlated cases of plaque and leads to gingivitis. Hence, it misclassifies the coronal plaque as gingivitis. The proposed CNN-ViT model can be used as a second opinion to classify the plaque and gingivitis at an early

stage. The proposed CNN-ViT model assists dentists and analyzes the oral images efficiently, supporting earlier diagnosis of dental plaque.

The proposed TUNED CNN-ViT method is used by the dentist to identify the phases of the infection at an earlier stage using a mobile app. The teeth images are acquired using a smartphone camera attached with a USB thermal camera in a real-time environment, and based on the severity of the infection and classification of the stages, treatment and drugs can be prescribed. However, existing devices/methods detect the presence of plaque and never classify. Table 11 shows the comparison of the proposed and existing methods. Table 12

shows the Ablation study of the proposed and existing methods. The clinical evaluation was conducted by dental surgeon. Standard periodontal diagnostic guidelines, such as visual inspection, probing depth assessment, and identification of gingival inflammation and plaque accumulation, are followed. The classification of oral conditions, such as healthy, plaque, gingivitis, and periodontitis, was diagnosed based on the clinical indicators, such as bleeding on probing, pocket depth, and tissue condition. The dentist's diagnosis was considered the ground truth for validating the proposed CNN-ViT model. Ethical approval was obtained, and all participants provided informed consent prior to data collection.

5. CONCLUSION

The presence of Coronal plaque, Gingivitis, and Periodontitis affects both physical and mental health of the person. It leads to social embarrassment and isolation due to halitosis (bad breath). Considering the above problem, the proposed study differentiates the healthy teeth and infectious teeth due to coronal plaque, gingivitis, and Periodontitis using the Tuned CNN-ViT classification method. Thermal images are acquired using a smartphone attached with a thermal USB camera and then pre-processed using the proposed hybrid model (CLAHE, gamma correction & Gaussian Noise filter) to eradicate the noise signal and improve the image quality. The pre-processing techniques equalize the image pixels and tune the contrast of the image using gamma correction. Then the features are extracted using the Tuned LBP algorithm, and further, it is optimized using POA to acquire vital features of the thermal image. Based on the severity of the tooth infection, the prognosis is analyzed and determined. This avoids unnecessary tooth loss, bone loss, and malignant diseases. The early prediction of Coronal plaque and Gingivitis avoids the occurrence of Periodontitis, tooth loss, and other malignant diseases. The proposed CNN-ViT method identifies the healthy teeth and infectious teeth, i.e., coronal plaque, gingivitis, and periodontitis, using the Tuned CNN-ViT classification method with an accuracy of 98.95% approximately. The Proposed CNN-ViT method assists the dentists and analyzes the oral thermal images efficiently, and supports immediate diagnosis, as shown in Figure 20. For the experimental set up HIKIMICRO thermal camera is attached to the RealMe Narzo 70 smartphone to diagnose the plaque, gingivitis, and periodontitis. As future work, different machine learning algorithms such as tuned Support Vector Machine (SVM), Random Forest, and XGBoost can be used for extraction or alternate feature selection, and their performance can be investigated for optimal results. Different thermal image palettes can be utilized, and their performance can be analyzed through experimentation using the proposed real-time setup. As future work, different machine learning and deep learning algorithms will be investigated to find the optimal model. Also, different thermal images from the Seek compact Pro and FLIR thermal camera can be utilized, and their performance can be analyzed through experimentation

5.1 Clinical relevance

5.1.1 Scientific rationale for the study

- Effective classification of coronal plaque, gingivitis, and periodontitis helps in understanding the disease progression and its risk.

- The advanced technologies (CNN & ViT) improve the accuracy of clinical diagnostics.

5.1.2 Principal findings

- Thermal image-based feature extraction and segmentation identify the plaque density well by focusing on the prominent feature.
- Predicting the disease at an early stage improves the treatment process and the patient's prognosis.

5.1.3 Practical implications

- The AI-based classification streamlines the diagnostic processes that save time for clinicians, and also prevents severe diseases like diabetes and cardiovascular issues.

ACKNOWLEDGMENT

We acknowledge "Chase Technologies, Chennai, Tamil Nadu, India" for their support.

REFERENCES

- [1] Chau, C.W., Li, G., Tew, I.M., Thu, K., McGrath, C., Lo, W.L., Ling, W.K., Hsung, T.C., Lam, W. (2023). Accuracy of artificial intelligence-based photographic detection of gingivitis. *International Dental Journal*, 73(5): 724-730. <https://doi.org/10.1016/j.identj.2023.03.007>
- [2] Alalharith, D.M., Alharthi, H.M., Alghamdi, W.M., Alsenbel, Y.M., Aslam, N., Khan, I.U., Shahin, S.Y., Dianišková, S., Alhareky, M.S., Barouch, K.K. (2020). A deep learning-based approach for the detection of early signs of gingivitis in orthodontic patients using faster region-based convolutional neural networks. *International Journal of Environmental Research and Public Health*, 17(22): 8447. <https://doi.org/10.3390/ijerph17228447>
- [3] Aroonratana, P., Lertpimonchai, A., Samaranayake, L., Vathesatogkit, P., Thienpramuk, L., Tavedhikul, K. (2024). The association between interdental cleaning and periodontitis in an urban Thai adult cohort: A cross-sectional study. *BMC Oral Health*, 24: 1185. <https://doi.org/10.1186/s12903-024-04980-6>
- [4] Li, W., Liang, Y., Zhang, X., Liu, C., He, L., Miao, L., Sun, W. (2021). A deep learning approach to automatic gingivitis screening based on classification and localization in RGB photos. *Scientific Reports*, 11: 16831. <https://doi.org/10.1038/s41598-021-96091-3>
- [5] Etta, I., Kambham, S., Girigosavi, K.B., Panjiyar, B.K. (2023). Mouth-heart connection: A systematic review on the impact of periodontal disease on cardiovascular health. *Cureus*, 15(10): e46585. <https://doi.org/10.7759/cureus.46585>
- [6] Li, W., Guo, E., Zhao, H., Li, Y., Miao, L., Liu, C., Sun, W. (2024). Evaluation of transfer ensemble learning-based convolutional neural network models for the identification of chronic gingivitis from oral photographs. *BMC Oral Health*, 24(1): 814. <https://doi.org/10.1186/s12903-024-04460-x>
- [7] Kurt-Bayrakdar, S., Bayrakdar, İ.Ş., Yavuz, M.B., Sali, N., Çelik, Ö., Köse, O., Uzun Saylan, B.C., Kuleli, B.,

- Jagtap, R., Orhan, K. (2024). Detection of periodontal bone loss patterns and furcation defects from panoramic radiographs using deep learning algorithm: A retrospective study. *BMC Oral Health*, 24(1): 155. <https://doi.org/10.1186/s12903-024-03896-5>
- [8] Dai, F., Liu, Q., Guo, Y., Xie, R., Wu, J., Deng, T., Zhu, H., Deng, L., Song, L. (2024). Convolutional neural networks combined with classification algorithms for the diagnosis of periodontitis. *Oral Radiology*, 40(3): 357-366. <https://doi.org/10.1007/s11282-024-00739-5>
- [9] Revilla-León, M., Gómez-Polo, M., Barmak, A.B., Inam, W., Kan, J.Y.K., Kois, J.C., Akal, O. (2023). Artificial intelligence models for diagnosing gingivitis and periodontal disease: A systematic review. *The Journal of Prosthetic Dentistry*, 130(6): 816-824. <https://doi.org/10.1016/j.prosdent.2022.01.026>
- [10] Iniesta, M., Chamorro, C., Ambrosio, N., Marín, M.J., Sanz, M., Herrera, D. (2023). Subgingival microbiome in periodontal health, gingivitis and different stages of periodontitis. *Journal of Clinical Periodontology*, 50(7): 905-920. <https://doi.org/10.1111/jcpe.13793>
- [11] Di Gianfilippo, R., Pini Prato, G., Franceschi, D., Castelluzzo, W., Barbato, L., Bandel, A., Di Martino, M., Pannuti, C.M., Chambrone, L., Cairo, F. (2025). Diagnostic reproducibility of the 2018 classification of gingival recessions: Comparing photographic and in-person diagnoses. *Journal of Periodontology*, 96(5): 467-477. <https://doi.org/10.1002/JPER.24-0173>
- [12] Ndjidda Bakari, W., Thiam, D., Mbow, N.L., Samb, A., Guirassy, M.L., Diallo, A.M., Diouf, A., Diallo, A.S., Benoist, H.M. (2021). New classification of periodontal diseases (NCPD): An application in a sub-Saharan country. *BDJ Open*, 7(1): 16. <https://doi.org/10.1038/s41405-021-00071-8>
- [13] Ertaş, K., Pence, I., Cesmeli, M.S., Ay, Z.Y. (2023). Determination of the stage and grade of periodontitis according to the current classification using machine learning algorithms. *Journal of Periodontal & Implant Science*, 53(1): 38-53. <https://doi.org/10.5051/jpis.2201060053>
- [14] Chang, J., Chang, M.F., Angelov, N., Hsu, C.Y., Meng, H.W., Sheng, S., Glick, A., Chang, K., He, Y.R., Lin, Y.B., Wang, B.Y., Ayilavarapu, S. (2022). Application of deep machine learning for the radiographic diagnosis of periodontitis. *Clinical Oral Investigations*, 26(11): 6629-6637. <https://doi.org/10.1007/s00784-022-04617-4>
- [15] Chang, H.J., Lee, S.J., Yong, T.H., Shin, N.Y., Jang, B.G., Kim, J.E., Huh, K.H., Lee, S.S., Heo, M.S., Choi, S.C., Kim, T.I., Yi, W.J. (2020). Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis. *Scientific Reports*, 10(1): 7531. <https://doi.org/10.1038/s41598-020-64509-z>
- [16] Sabri, H., Nava, P., Hazrati, P., Alrmali, A., Galindo-Fernandez, P., Saleh, M.H.A., Calatrava, J., Barootchi, S., Tavelli, L., Wang, H.L. (2025). Comparison of ultrasonography, CBCT, transgingival probing, colour-coded and periodontal probe transparency with histological gingival thickness: A diagnostic accuracy study revisiting thick versus thin gingiva. *Journal of Clinical Periodontology*, 52(4): 547-560. <https://doi.org/10.1111/jcpe.14139>
- [17] PPatil, S., Joda, T., Soffe, B., Awan, K.H., Fageeh, H.N., Tovani-Palone, M.R., Licari, F.W. (2023). Efficacy of artificial intelligence in the detection of periodontal bone loss and classification of periodontal diseases: A systematic review. *Journal of the American Dental Association*, 154(9): 795-804. <https://doi.org/10.1016/j.adaj.2023.05.010>
- [18] Nancy, V., Balakrishnan, G. (2019). Thermal image-based object classification for guiding the visually impaired. *The Computer Journal*, 64(11): 1747-1759. <https://doi.org/10.1093/comjnl/bxaa097>
- [19] 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. *Journal of Clinical Periodontology*, 51(12): 1547-1560. <https://doi.org/10.1111/jcpe.13856>
- [20] Chandrashekar, H.S., Geetha Kiran, A., Murali, S., Dinesh, M.S., Nanditha, B.R. (2021). Oral images dataset (Version 2). Mendeley Data. <https://doi.org/10.17632/mhjyrm35p4.2>
- [21] Liu, Q., Dai, F., Zhu, H., Yang, H., Huang, Y., Jiang, L., Tang, X., Deng, L., Song, L. (2023). Deep learning for the early identification of periodontitis: A retrospective, multicentre study. *Clinical Radiology*, 78(12): e985-e992. <https://doi.org/10.1016/j.crad.2023.08.017>
- [22] Jundaeng, J., Chamchong, R., Nithikathkul, C. (2025). Periodontitis diagnosis: A review of current and future trends in artificial intelligence. *Technology and Health Care*, 33(1): 473-484. <https://doi.org/10.3233/THC-241169>
- [23] Zhao, D., Homayounfar, M., Zhen, Z., Wu, M.Z., Yu, S.Y., Yiu, K.H., Vardhanabhuti, V., Pelekos, G., Jin, L., Koochi-Moghadam, M. (2022). A multimodal deep learning approach to predicting systemic diseases from oral conditions. *Diagnostics*, 12(12): 3192. <https://doi.org/10.3390/diagnostics12123192>
- [24] Chen, Y., Chen, X. (2020). Gingivitis identification via GLCM and artificial neural network. In *Medical Imaging and Computer-Aided Diagnosis: Proceedings of MICAD 2020*, vol 633, Springer, Singapore, pp. 95-106. https://doi.org/10.1007/978-981-15-5199-4_10
- [25] Aykol-Sahin, G., Yucel, O., Eraydin, N., Keles, G.C., Unlu, U., Baser, U. (2025). Efficiency of oral keratinized gingiva detection and measurement based on convolutional neural network. *Journal of Periodontology*, 96(6): 652-662. <https://doi.org/10.1002/JPER.24-0151>
- [26] Lakshmi, T.K., Dheeba, J. (2022). Digital decision making in dentistry: Analysis and prediction of periodontitis using machine learning approach. *International Journal of Next-Generation Computing*, 13(3): 305-322.
- [27] Sajid, S. (n.d.). Oral diseases dataset. Kaggle. <https://www.kaggle.com/datasets/salmansajid05/oral-diseases>.
- [28] Scannapieco, F.A., Dongari-Bagtzoglou, A. (2021). Dysbiosis revisited: Understanding the role of the oral microbiome in the pathogenesis of gingivitis and periodontitis: A critical assessment. *Journal of Periodontology*, 92(8): 1071-1078. <https://doi.org/10.1002/JPER.21-0120>
- [29] Yan, Y.J., Wang, B.W., Yang, C.M., Wu, C.Y., Ou-Yang, M. (2021). Autofluorescence detection method for dental plaque bacteria detection and classification. *Dentistry Journal*, 9(7): 74. <https://doi.org/10.3390/dj9070074>

- [30] Hong, I., Pae, H.C., Song, Y.W., Cha, J.K., Lee, J.S., Paik, J.W., Choi, S.H. (2020). Oral fluid biomarkers for diagnosing gingivitis in humans: A cross-sectional study. *Journal of Clinical Medicine*, 9(6): 1720. <https://doi.org/10.3390/jcm9061720>
- [31] Ma, Y., Wang, H., Shen, H., Duan, S., Wen, S. (2025). Analog spiking U-Net integrating CBAM ViT for medical image segmentation. *Neural Networks*, 181: 106765. <https://doi.org/10.1016/j.neunet.2024.106765>
- [32] Zhang, Z., Wu, H., Zhao, H., Shi, Y., Wang, J., Bai, H., Sun, B. (2023). A novel deep learning model for medical image segmentation with convolutional neural network and transformer. *Interdisciplinary Sciences: Computational Life Sciences*, 15(4): 663-677. <https://doi.org/10.1007/s12539-023-00585-9>
- [33] Hörst, F., Rempe, M., Heine, L., Seibold, C., Keyl, J., Baldini, G., Ugurel, S., Siveke, J., Grünwald, B., Egger, J., Kleesiek, J. (2024). CellViT: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94: 103143. <https://doi.org/10.1016/j.media.2024.103143>
- [34] Yu, S., Wang, X., Ma, L., Li, Y., Yin, W., Zheng, H., Chen, H. (2026). Investigating the thermal failure mechanisms of teeth: insights from an improved thermal-mechanical-damage coupling meshless numerical method. *Arabian Journal for Science and Engineering*, pp. 1-22. <https://doi.org/10.1007/s13369-026-11166-5>
- [35] Çankaya, Z.T., Koyuncu, A., Gürbüz, S. (2026). Artificial intelligence assisted thermal imaging for gingival inflammation assessment: A novel approach. *Journal of Esthetic and Restorative Dentistry*, 38(2): 362-370. <https://doi.org/10.1111/jerd.70045>
- [36] Özcan, C., Yiğitarıslan, K. (2025). Gingival temperature variations in dogs: Assessing healthy and inflamed gingiva using thermal imaging before and during anaesthesia. *Veterinary Medicine and Science*, 11(4): e70475. <https://doi.org/10.1002/vms3.70475>
- [37] Kabakci, A., Yilmaz, A., Helvacıoglu-Yigit, D., Nawar, N.N., Kim, H.C. (2025). Thermal behaviour of teeth with internal root resorption during obturation and enhancing thermal simulations: A finite-element analysis. *International Dental Journal*, 75(6): 103903. <https://doi.org/10.1016/j.identj.2025.103903>
- [38] Sundar, K., Ravikumar, S., Berna, E., Vijay, K., Vimalan, M. (2025). A deep learning framework for glaucoma diagnosis and visual field progression prediction using ResNet-50 and LSTM. In *2025 International Conference on Intelligent Computing, Information and Control Systems (ICOIICS)*, Lalitpur, Nepal, pp. 1606-1611. <https://doi.org/10.1109/ICOIICS67115.2025.11390341>
- [39] Ferrara, E., Rapone, B., D'Albenzio, A. (2025). Applications of deep learning in periodontal disease diagnosis and management: A systematic review and critical appraisal. *Journal of Medical Artificial Intelligence*, 8: 23. <https://doi.org/10.21037/jmai-24-241>
- [40] Tao, L.R., Li, Y., Wu, X.Y., Gu, Y., Xie, Y., Yu, X.Y., Lai, H.C., Tonetti, M.S. (2026). Deep learning photo processing for periodontitis screening. *Journal of Dental Research*, 105(2): 226-235. <https://doi.org/10.1177/00220345251347508>
- [41] Sabri, R.K., Abdulkadir, L.Y., Khidhir, A.M., Saleh, H.A. (2025). Diagnosing gingiva disease using artificial intelligence techniques. *Diyala Journal of Engineering Sciences*, 179-190. <https://doi.org/10.24237/djes.2024.18211>
- [42] Israr, S., Ilyas, M., Atif, M., Irfan, M., Ahmad, H., Ashraf, M., Abbas, T. (2025). Accurate periodontal disease classification from dental radiographs using deep learning models. *Statistics, Computing and Interdisciplinary Research*, 7(2): 411-434. <https://doi.org/10.52700/scir.v7i2.222>