Frontiers in Dentistry

Early Prediction of Oral Precancerous Lesions Using Artificial Intelligence

Pradeep kumar Rathinavelu[1]*, Vyshali Pillai[1], Pradeep Kumar Yadalam[2], Sri Shivasankari Thilagar[3], Amir Raee[4,5], Artak Heboyan[6,7,8]

1. Department of Public Health Dentistry, Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India
2. Department of Periodontics, Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India
3. Department of Periodontology and Implantology, Priyadarshini Dental College and Hospital, Thiruvallur Taluk & District, India
4. Dental Implant Research Center, Dentistry Research Institute, Tehran University of Medical Sciences, Tehran, Iran
5. Department of Periodontics, School of Dentistry, Tehran University of Medical Sciences, Tehran, Iran
6. Department of Research Analytics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India
7. Department of Prosthodontics, Faculty of Stomatology, Yerevan State Medical University after Mkhitar Heratsi, Yerevan, Armenia
8. Department of Prosthodontics, School of Dentistry, Tehran University of Medical Sciences, Tehran, Iran

Article Info

Article type:
Original Article

Article History:
Received: 22 Mar 2025
Accepted: 10 Jul 2025
Published: 20 Mar 2026

**** Corresponding author:***
Department of Public Health Dentistry, Saveetha Dental College, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, India

Email: drrpradeepkumar@gmail.com

A B S T R A C T

Objectives: Oral potentially malignant disorders may present as white, red, or mixed red-white lesions and require accurate early recognition. This study evaluated whether texture-analysis features extracted from clinical digital images could distinguish oral precancerous lesions from other oral mucosal lesions and normal mucosa using gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), and wavelet analysis.

Materials and Methods: Sixty-four clinical digital images were selected according to predefined inclusion and exclusion criteria. The dataset included leukoplakia, erythroplakia, oral submucous fibrosis, candidiasis, lichen planus, leukoderma, frictional keratosis, white spongy nevus, and normal mucosa. Regions of interest were extracted from each image, and texture features were derived using GLCM, GLRLM, and wavelet analysis. A support vector machine (SVM) classifier was then used to categorize images as oral precancerous lesions or non-precancerous/normal mucosa.

Results: GLCM yielded the highest classification accuracy (88%), followed by GLRLM (81%) and wavelet analysis (79%). The corresponding sensitivity values were 77%, 64%, and 60%, and the specificity values were 93%, 90%, and 89%, respectively. The positive predictive values were 83% for GLCM, 75% for GLRLM, and 75% for wavelet analysis.

Conclusion: GLCM-based texture features provided the best diagnostic performance in this dataset. These image-analysis methods may be useful as non-invasive adjuncts to conventional clinical examination and histopathological diagnosis; however, larger datasets and external validation are required before clinical implementation.

Keywords: Wavelet Analysis; Leukoplakia; Erythroplasia; Oral Submucous Fibrosis; Candidiasis; Artificial Intelligence

➢ **Cite this article as:** Rathinavelu PK, Pillai V, Yadalam PK, Thilagar SSh, Raee A, Heboyan A. Early Prediction of Oral Precancerous Lesions Using Artificial Intelligence. *Front Dent.* 2026:23:14. http://doi.org/10.18502/fid.v23i14.21568

INTRODUCTION

A premalignant or precancerous lesion is a benign, morphologically altered tissue with a greater-than-normal risk of malignant transformation. Several types of premalignant lesions occur in the mouth [1]. Oral cancers begin as white patches (leukoplakia), red patches (erythroplakia), or mixed red and white patches (erythroleukoplakia or "speckled leukoplakia"). Other common premalignant lesions include oral lichen planus (particularly the erosive type), oral submucous fibrosis, and actinic cheilitis [2,3].

India accounts for 86% of oral cancer cases [4]. Because of the difficulty in detecting oral cancer early, it has one of the worst survival rates of all cancers. Ahmedabad is considered the capital of oral cancers, with 40% of cancers recorded and mostly caused by tobacco and gutkha chewing [5–8]. A high incidence of oral cancer is mainly due to the late diagnosis of potential precancerous lesions and conditions. Population-based screening programs for oral premalignant lesions are more expensive. A simulation model of population screening for oral precancerous lesions indicated that approximately 18000 patients should be screened to save one life. Screening examinations were, therefore, associated with reduced mortality among high-risk patients. The most logical approach to reducing morbidity and mortality is to increase early detection of suspicious oral premalignant lesions through preventive screening [7].

The diagnosis of Oral pre-cancer remains a challenge, particularly in the detection, evaluation, and management of early-phase alterations. Currently, various specialized investigations, including CT (Computed Tomography), CAT (Coaxial Tomography), PET (Positron Emission Tomography), MRI (Magnetic Resonance Imaging), and tissue biopsy, are employed for diagnosis [9]. A non-invasive brush biopsy (BrushTest) can rule out dysplasia (pre-cancer) and cancer in areas of the oral mucosa exhibiting unexplained color variation or lesions [10-12]. Tissue biopsies of the lesion can only confirm the diagnosis of oral cancer or pre-cancer. These diagnostic procedures are unsuitable for community screening due to their time-consuming and expensive nature. Current oral cancer screening and case-finding aids are toluidine blue, brush cytology, tissue reflectance, and autofluorescence. However, the brush biopsy results do not always detect cellular atypia. Toluidine blue is significantly less useful in detecting premalignant lesions due to high false-negative staining rates for carcinoma in situ and false-positive results in ulcerated inflammatory or traumatic lesions [13].

Artificial neural networks can be a simple, effective, inexpensive, and non-invasive diagnostic and prognostic tool for various premalignant lesions and conditions. It varies with the complexity of the lesion, is economical, less time-consuming, and an accurate tool for measuring the progression of premalignant lesions [1,14-16]. Wavelet analysis enhances image recognition. It uses wavelets, mathematical algorithms that capture local and global information, to break down an image into its frequency components. The analysis extracts key visual features and patterns at various scales and resolutions. Representing an image in the wavelet domain makes image denoising, edge detection, texture analysis, and object recognition efficient. Wavelet analysis is useful in image recognition because it captures both fine features and overall structure in a multi-resolution representation [17,18].

This study aims to develop a tool that automates the diagnostic procedure and applies optimal texture feature extraction to better capture underlying physiology, thereby improving cancer-recognition accuracy [19].

Thus, computer-assisted disease diagnosis (CAD) is important and has become a major research subject. Different texture analysis techniques will be automatically applied to images to determine whether a disease is present in the analyzed samples [20]. Also, this study will help determine the different grades or severity of disease, if the disease is present in the sample. This study determines the tool's ability to diagnose precancerous lesions, differentiate it from other oral lesions, such as normal oral mucosa, compare the results with the actual diagnosis, and assess the diagnostic ability in

identifying oral precancerous lesions using various methods of texture analysis like gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM) and wavelet analysis compared to the actual diagnosis.

MATERIALS AND METHODS

The manuscript is approved with the ethical code of IHEC/SDS/FACULTY/23/PERIO/346. The study is designed to predict precancerous lesions using texture analysis of clinical digital images. The study was designed in the Department of Public Health Dentistry with assistance from the Department of Oral Medicine and Radiology, Saveetha Dental College, and the hospital. The sample size was calculated based on an Anuradha study. K et al. 201, N= 64 (calculated using G-power) (at 95% power and 5% α-error)

Inclusion criteria:

Following a retrospective study protocol, digital images of patients diagnosed with precancerous lesions, such as leukoplakia, erythroplakia, and oral submucous fibrosis, other surface lesions of oral mucosa like Candidiasis, white spongy nevus, lichen planus, leukoderma, and images of buccal mucosa of healthy volunteers without any oral habits and systemic diseases.

Exclusion criteria:

Images are not technically of good diagnostic quality (diffuse Image or distortion).

Study procedure:

The data for this study consisted of 64 digital images of various oral mucosal lesions selected after evaluation against the inclusion and exclusion criteria. All images were in digital form. Various oral lesions included in the study are leukoplakia, erythroplakia, oral submucous fibrosis, Candida, lichen planus, leukoderma, frictional keratosis, and white spongy nevus. This data was collected from the Department of Oral Medicine and Radiology, Saveetha Dental Hospital. In this research, we developed a tool to predict and differentiate oral precancerous lesions from similar oral lesions and healthy mucosa. The assessed diagnosis was then correlated with the actual diagnosis.

Proposed method:

I. Data acquisition

A 64 set of digital images was collected from Saveetha Dental Hospital. The Image of the ground truth was considered a reference. All photos were digital. Leukoplakia, erythroplakia, oral submucous fibrosis, Candida, lichen planus, leukoderma, frictional keratosis, and white spongy nevus were studied. The extracted and selected attributes were fed into a classifier to classify the photos into oral precancerous and other oral lesion categories. Intra-oral Images were received from the Saveetha Digital Management Software from the Oral Medicine Department. Dental clinical practitioners have diagnosed these images.

II. Preprocessing and Enhancement

Preprocessing is performed to improve image quality. This stage leads to less complicated and more reliable feature extraction phases. The preprocessing procedure removes the artifacts and noise from the Image. In this work, the Gaussian filter was used to remove noise and smooth the images, thereby improving their contrast.

A median filter using a 3-by-3 square was applied to reduce noise. The median filter was chosen because it helps remove outliers without reducing the image's sharpness and is less sensitive to extreme values. After applying the median filter, abnormalities become more conspicuous in the homogenous background. There is a need to intensify the lesion region. Imadjust is used for image enhancement. Imadjust maps the intensity of a grayscale image to a new value, which is 1% of data saturated at high and low intensities. This enhances the contrast of the Image. Imadjust is a user-friendly image enhancement method with contrast-stretching, adjustable parameters, and speed. It enhances feature visibility, maintains a natural appearance, and is computationally efficient, making it a reliable tool.

III. Selection Region of interest (ROI) of the lesion

In this research, the ROIs were the oral mucosal lesions. All lesions were manually selected from the images by a well-trained operator and

further confirmed by a radiologist. An ROI of size 40×40 pixels was extracted, with the mass centered in the window. A 40×40-pixel ROI enables precise examination of oral mucosal lesions, excluding background noise and facilitating detailed analysis. Effectiveness depends on image resolution and avoiding irrelevant features. Then, the masses were divided into two sets: the learning set and the testing set. Using one of three hold-out cross-validation folds, the learning set comprised 70% training and 30% test data.

IV. Feature extraction using texture and wavelet analysis

a. *Feature extraction using texture analysis:*

The texture of the Image is the local variation in the image intensity. Texture analysis can be statistically done using a co-occurrence matrix. This matrix is generated from the estimation of pairwise statistics of the image pixel intensity. The GLCM is constructed on the theory that identical gray-level configurations are recurrent in a texture. The co-occurrence matrix is given by P (i, j | d, Θ), where i and j are gray-level values at a distance d with an angle Θ. GLCM can be used to extract second-order statistical features. The number of columns and rows is equal in the GLCM. The following notations are used in GLCM.

μ-Mean value of P

μx and μy-Mean value of Px and Py

σx and σy- Standard deviation of Px and Py

G– Size of co-occurrence matrix

The subsequent GLCM features were used in this study:

-Energy: It gives the image uniformity. For a constant image, energy is 1.

-Correlation: measures the level of dependency between the pixels at the relative position to each other, which is specified. For a considerable linear structure image, the correlation will be high.

-Contrast: the number of local differences in an image. This gives the degree of contrast that favors the influence of the diagonal element p (i, j) for i = j. The greater the variations in the Image, the more the diagonal P (i, j) s will be concatenated, resulting in higher contrast.

-Homogeneity: It measures the degree to which the components of the GLCM are near the diagonal of the GLCM matrix. For the diagonal elements, homogeneity is 1. The co-occurrence matrix of a homogenous Image is a mixture of high and low P (i, j), whereas heterogeneous images give an equal spread of P (i, j). The input vector to the neural network and the wavelet transform coefficients were given.

b. *Feature extraction using GLRLM:*

GLRLM is a statistical method used in image processing and recognition, particularly texture analysis. It captures the frequency of runs of consecutive pixels with the same gray level in different directions in an image. The GLRLM helps characterize image texture and patterns, which are commonly used in medical imaging and material science. The algorithm involves Image preprocessing, gray-level quantization, and matrix creation. Various statistical features can be derived from the GLRLM, such as Short Run Emphasis (SRE), Long Run Emphasis (LRE), Run Length Non-uniformity (RLN), Gray Level Non-uniformity (GLN), and Run Percentage (RP), which are crucial for image recognition tasks.

c. *Feature extraction using wavelet analysis:*

Feature selection aids in identifying texture, morphology, color, boundary irregularity, and statistical properties of lesions. It helps distinguish between benign and malignant lesions, measure lesion size, and understand lesion boundaries. Combining these features into machine learning models can enhance diagnostic accuracy and patient management in oral healthcare, reducing dimensionality and ensuring interpretability. The wavelet transform is a mathematical tool for analyzing signals by dividing them into different frequency bands that reveal spatial or temporal patterns. It decomposes a signal into components at different frequency bands, making it suitable for examining spatial or temporal patterns. The choice of wavelet functions and levels of decomposition is crucial for effective application. Decomposition levels in signal analysis balance detail and computational load, considering feature extraction and signal compression purposes. Contextual adaptation also influences level determination.

An image can be approximated by a matrix A

whose elements are the corresponding gray-level pixel intensities P(i, j) [21]. A is a square of dimension 2n X 2n, where n is an integer. The wavelet transform procedure is as follows: filters G and H are applied to the rows of the A matrix. The resulting matrices are HrA and LrA, both of size 2n X 2n-1. Now the column matrix of HrA and LrA is again applied with G and H filters, resulting in four matrices HcLrA, HcHrA, LcLrA, LcHrA of size 2n-1X2n-1. The matrix LcLrA is the average matrix, and the others are the detail matrices. Similarly, many levels of decomposition can be performed.

V. Identifying oral precancerous lesions and other oral lesions using the support vector machine (SVM) method:

After the extraction and selection, the features were input into a classifier to categorize the images into the oral precancerous and other oral lesion images. We used the Support Vector Machine (SVM) method to categorize these lesions. SVM is a popular classifier for its advantages over decision trees and neural networks. It excels in high-dimensional data scenarios, minimizes classification errors, and generalizes effectively. SVMs handle non-linear relationships, simplify training, and offer clearer interpretability. They require less computational power and can scale to large datasets using techniques such as stochastic gradient descent.

VI. Evaluating SVM using the receiver operating characteristic (ROC) curve and computing the AUC:

This research employed the ROC curve because of its comprehensive and fair evaluation ability [7]. A ROC curve plots the true positive fraction (TPF) as a function of the false positive fraction (FPF) [22, 23]. The area under the ROC curve (AUC) can be used as a criterion.

Statistical Analysis:

Data was entered into a Microsoft Excel spreadsheet and analyzed using SPSS software (version 21). Descriptive statistics were used. A p-value <0.05 was considered statistically significant at the significance level.

Measures of Performance Evaluation:

Various measures are used to evaluate the system's performance. Classification Accuracy (AC) and Mathews Correlation Coefficient (MCC) are the measures used. These values are calculated from the Confusion Matrix. A confusion matrix (Kohavi and Provost, 1998) contains information about actual and predicted classifications done by a classification system. The performance of such systems is commonly evaluated using the matrix data. From the confusion matrix, accuracy (AC) can be obtained. The following formula shows the accuracy measurement.

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

TN (True Negative): Correct Prediction as normal

FN (False Negative): Incorrect prediction of normal

FP (False Positive): Incorrect prediction of abnormal

TP (True Positive): Correct prediction of abnormal

The Matthews correlation coefficient (MCC) is used in machine learning to measure the quality of binary (two-class) classifications. The MCC is a correlation coefficient that measures agreement between observed and predicted binary classifications; it ranges from −1 to +1. A coefficient of +1 represents a perfect prediction, 0 is no better than a random prediction, and −1 indicates total disagreement between prediction and observation.

RESULTS

All experiments were conducted in MATLAB R2007 b running on an Intel Pentium Centrino PC with 1GB of RAM. Sixty-four precancerous lesions, other mucosal lesions, and normal mucosa images measuring 40×40 pixels were transformed into wavelet analysis, GLCM, and GLRLM. We obtained up to 79% accuracy using wavelet analysis, 88% GLCM, and 81% GLRLM. Using the sensitivity and specificity values, the ROC curve is plotted to evaluate the performance of GLCM, GLRLM, and wavelet analysis. Sensitivity and specificity values obtained using wavelet analysis, GLCM, and GLRLM are 69%, 77%, 64%, 89%, 93%, and 89%, respectively (Fig. 1), (Table 1, 2, 3, and 4).

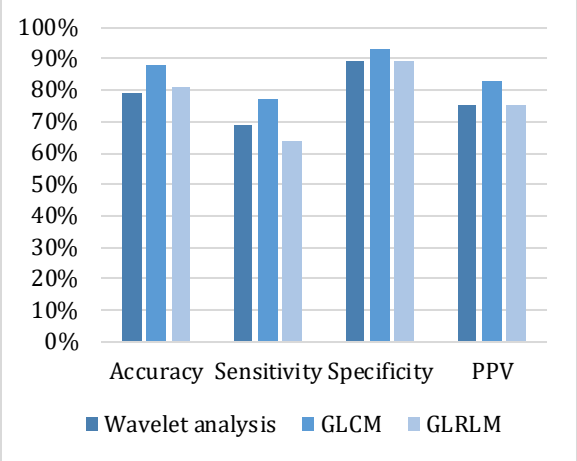

Fig 1. Comparative diagnostic performance of wavelet analysis, gray-level co-occurrence matrix (GLCM), and gray-level run-length matrix (GLRLM) feature extraction for classifying oral precancerous lesions.

Table 1. Confusion matrix for gray-level co-occurrence matrix (GLCM) features

Feature set	Actual class	Predicted negative	Predicted positive
GLCM	Negative	28 (TN)	2 (FP)
GLCM	Positive	3 (FN)	10 (TP)

TN: true negative; FP: false positive; FN: false negative; TP: true positive.

Table 2. Confusion matrix for gray-level run-length matrix (GLRLM) features

Feature set	**Actual class**	**Predicted negative**	**Predicted positive**
GLRLM	Negative	26 (TN)	3 (FP)
GLRLM	Positive	5 (FN)	9 (TP)

TN: true negative; FP: false positive; FN: false negative; TP: true positive.

Table 3. Confusion matrix for wavelet analysis

Feature set	Actual class	Predicted negative	Predicted positive
Wavelet analysis	Negative	25 (TN)	3 (FP)
Wavelet analysis	Positive	6 (FN)	9 (TP)

TN: true negative; FP: false positive; FN: false negative; TP: true positive.

The study evaluated the effectiveness of three feature extraction methods -Wavelet analysis, Gray-Level Co-occurrence Matrix (GLCLM), and Gray-Level Run-Length Matrix (GLRLM)- in classifying precancerous lesions. GLCLM outperformed all methods in predicting outcomes with 88% accuracy. However, it had low Matthews Correlation Coefficient values, indicating poor prediction quality. GLCLM excelled in sensitivity (77%), specificity (93%), and Positive Predictive Value (83%). Further enhancements are needed to improve predictive capabilities, as explained in Table 4.

Table 4. Evaluation results for gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), and wavelet analysis

Measure	**Wavelet analysis**	**GLCLM**	**GLRLM**
AC	79%	88%	81%
MCC	0.004	0.006	0.005
SN	69%	77%	64%
SP	89%	93%	89%
PPV	75%	83%	75%

AC: accuracy; MCC: Matthews correlation coefficient; SN: sensitivity; SP: specificity; PPV: positive predictive value.

This research employed the ROC curve because of its comprehensive and fair evaluation capabilities. A ROC curve (Fig. 2) plots the true positive fraction (TPF) as a function of the false positive fraction (FPF). The area under the ROC curve (accuracy) can be used as a criterion.

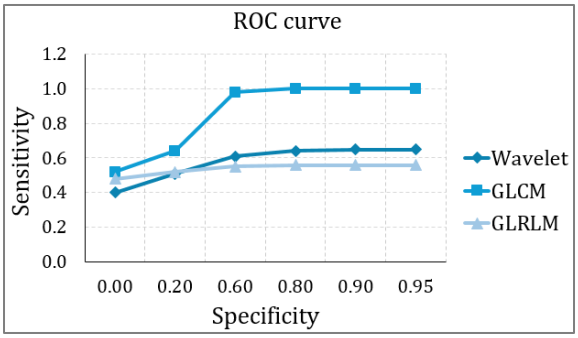

Fig 2. Receiver operating characteristic (ROC)-based performance plot for wavelet analysis, gray-level co-occurrence matrix (GLCM), and gray-level run-length matrix (GLRLM).

DISCUSSION

Early detection includes diagnosing oral potentially malignant disorders (OPMDs) and regular follow-ups.

OPMDs include oral leukoplakia, proliferative verrucous leukoplakia, erythroplakia, oral lichen planus, oral submucous fibrosis, palatal lesions in reverse smokers, oral lupus erythematosus, actinic keratosis, dyskeratosis congenita, oral lichenoid lesion, and oral chronic graft-versus-host disease. Detecting oral potentially malignant disorders (OPMD) is crucial for reducing oral cancer morbidity and mortality. However, visual screening is problematic due to inadequate training, heterogeneity in lesion appearance, and delays in patient referral. Early-stage OSCC lesions and OPMD are often asymptomatic, leading to late presentation and further diagnostic delays. Soft computation techniques such as texture analysis [3,17,18,21-23] and neural networks for feature extraction will lead to an intelligent, low-cost, and interpretable solution compared to traditional techniques. Today, texture-based medical imaging has a vast amount of research focused on the early detection of stroke; Automatic Diagnostic Systems for CT Liver and Computed tomography (CT) images have been widely used for liver disease diagnosis, and a computer-aided classification system for cancer detection from digital mammograms [24,25].

This study approaches the problem by applying texture analysis (GLCM and GLRM) [3,17,18] and wavelet analysis to extract features from the Image. We calibrated the values using images of precancerous lesions, other oral mucosal lesions, and normal mucosa without lesions (Figs. 3, 4A).

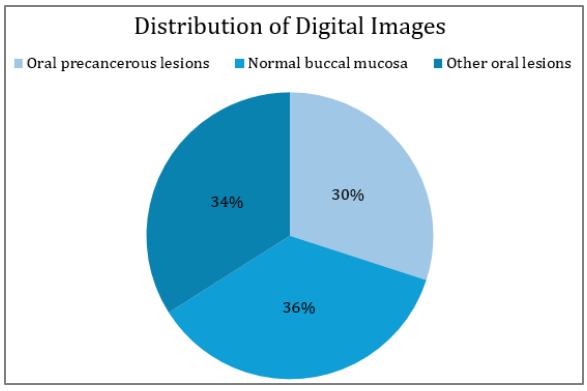

Fig 3. The distribution of digital images based on oral precancerous lesions, other oral lesions, and normal oral mucosa.

After this, the recognition and differentiation of precancerous lesions, other oral lesions, and normal mucosa using the methods mentioned will occur automatically [26]. All methods analyzed the change in expression for each lesion. Later, the classification's sensitivity, specificity, accuracy, Mathews correlation coefficient (MMC), and Positive Predictive Value (PPV) were calculated (Figs. 3, 4, 5). The features' values from both classes overlap, but the minimum and maximum values differ for all three cases. This result indicates that the classification process cannot be easily performed (i.e., it is not linear) due to overlapping values. However, the maximum and minimum feature values for each class are used for classification. For example, the mean values for precancerous lesions range from 13.7650 to 213.2888, while the mean values for other oral lesions range from 85.0038 to 252.3044, and for other mucosal lesions from 114.5560 to 362.2385. This means that distinguishing oral precancerous lesions, other oral lesions, and normal mucosa is now possible using a support vector machine.

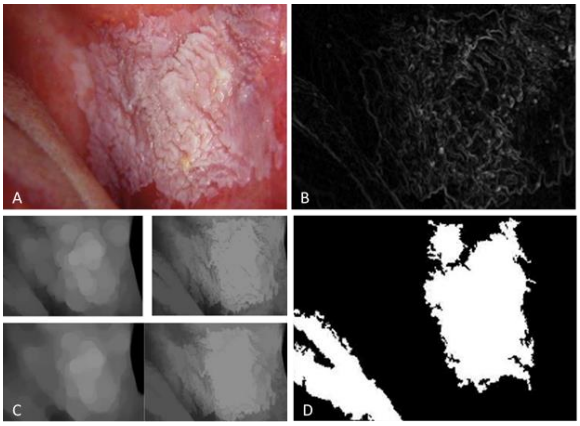

Fig 4. Image-processing workflow for an oral precancerous lesion. A, clinical digital image of leukoplakia showing increased keratinization of the left buccal mucosa extending from the oral commissure toward the mandibular molar region; B, preprocessing and enhancement to reduce artifacts and noise; C, morphological processing for lesion-region extraction; D, regional maxima superimposed on the original image to refine the region of interest before wavelet, GLCM, and GLRLM texture analysis. In this step, the lesion is associated with the nearby anatomical landmark and is separated from the digital Image.

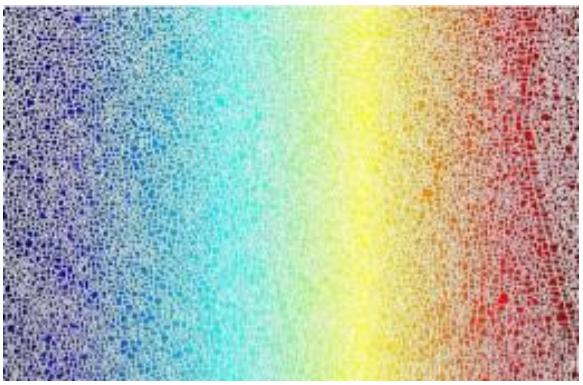

Fig 5. Wavelet-based feature extraction from image coefficients followed by support vector machine (SVM) classification.

A novel directional wavelet-based convolutional neural network algorithm has been proposed to reduce radiation exposure in X-ray CT scans, thereby reducing image artifacts and improving diagnostic reliability, with potential for large datasets and new directions in low-dose CT research [27].

The researchers developed a deep learning algorithm for low-dose X-ray CT, placing second in the 2016 AAPM Grand Challenge. They introduced a novel framelet-based denoising algorithm using a wavelet residual network, which improved performance while preserving the original image textures [28].

Furthermore, using a wavelet subband discriminator enables the removal of artifacts in all subbands by utilizing frequency-specific wavelet subband discriminators. Our method demonstrates competitive performance in noise and scatter removal for CXRs, compared to existing approaches, without requiring additional processing steps during testing. Additionally, we illustrate that our wavelet subband discriminator, combined with a switchable CycleGAN, offers flexibility by generating different levels of artifact removal [27,29].

In this study, we observed that the texture features can be used to predict precancerous lesions. Texture is one of the crucial characteristics used to identify objects or ROI (region of interest) in an image. Five texture features were extracted based on wavelet analysis, 13 on GLCM, and seven on GLRLM [23]. The five features from wavelet analysis are mean, median, standard deviation, kurtosis, and skewness. Autocorrelation, Contrast, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Homogeneity, Maximum Probability, Information Measure, Inverse Difference, Inverse Difference normalized, and Inverse Difference moment normalized were used by GLCM. Short Runs Emphasis, Long Runs Emphasis, Gray Level Non-uniformity, Run Percentage, Run Length Non-uniformity, Low Gray Level, Run Emphasis, and High Gray Level Run Emphasis were used by GLRLM. This study uses deep learning to achieve high-quality 3D images under sparse-sampling CT conditions, similar to previous studies showing that DenseNet201 and Swin Transformer models demonstrated exceptional classification performance on an internal test set, with F1-scores of 0.84 and 0.83, respectively. The DenseNet201 AI model, with an F1 Score of 0.73, has the potential to assist in the early detection and referral of oral potentially malignant disorders using photographic images [29]. The Cancer Imaging Archive dataset was trained using a fully convolutional network and wavelet transform model. The method effectively removes streak artifacts, offering a new perspective on sparse sampling CT reconstruction [7–19,30-35]

The strength of this method is that it is simple, non-invasive, and quick. As the main purpose of this study was to optimize the diagnosis of these oral precancerous lesions, using larger sample sets to improve diagnostic training accuracy can be beneficial [27,29,36,37,38]. Although we achieved high sensitivity and specificity for diagnostic classification, we believe that with larger arrays and more samples, these models can be improved to increase diagnostic sensitivity in clinical practice. Future applications of these methods will include studies to classify cancers by stage and biological behavior, to predict prognosis, and to guide therapy. The results show that the system's performance improves with more image training, as fewer FP detections are produced.

CONCLUSION

GLCM provided the highest classification accuracy (88%) among the evaluated feature-extraction methods. GLRLM and wavelet-analysis features also showed diagnostic potential but require further optimization. These methods may serve as non-invasive adjuncts to clinical assessment and histopathological diagnosis, but they should not replace biopsy-based confirmation. Larger prospective studies are required to validate their clinical utility and improve diagnostic performance.

CONFLICT OF INTEREST STATEMENT

None declared.

GENERATIVE AI IN SCIENTIFIC WRITING

The authors declare that no generative artificial intelligence tool was used for study design, data collection, image analysis, statistical analysis, interpretation of results, or generation of scientific conclusions. AI-assisted language editing, if used, was limited only to improving grammar and readability, and all content was critically reviewed and approved by the authors.

REFERENCES

1. Nurtanio I, Purnama IK, Hariadi M, Purnomo MH. Cyst and tumor lesion segmentation on dental panoramic images using active contour models. IPTEK The Journal for Technology and Science. 2011 Aug 3;22(3).
2. Li R, Xiao L, Gong T, Liu J, Li Y, Zhou X, et al Role of oral microbiome in oral oncogenesis, tumor progression, and metastasis. Mol Oral Microbiol. 2023 Feb;38(1):9-22.
3. Vijayakumari B, Ulaganathan G, Banumathi A, Banu AF, Kayalvizhi M. Dental cyst diagnosis using texture analysis. In2012 International Conference on Machine Vision and Image Processing (MVIP) 2012 Dec 14 (pp. 117-120). IEEE. from: http://dx.doi.org/10.1109/mvip.2012.6428774
4. Bankfalvi A, Piffko J. Prognostic and predictive factors in oral cancer: the role of the invasive tumour front. Journal of Oral Pathology & Medicine: Review article. 2000 Aug;29(7):291-8.
5. Ezhov M, Gusarev M, Golitsyna M, Yates JM, Kushnerev E, Tamimi D, et al. Clinically applicable artificial intelligence system for dental diagnosis with CBCT. Sci Rep. 2021 Jul 22;11(1):15006.
6. Braakhuis BJ, Leemans CR, Brakenhoff RH. A genetic progression model of oral cancer: current evidence and clinical implications. J Oral Pathol Med. 2004 Jul;33(6):317-22.
7. Feller LL, Khammissa RR, Kramer BB, Lemmer JJ. Oral squamous cell carcinoma in relation to field precancerisation: pathobiology. Cancer Cell Int. 2013 Apr 3;13(1):31.
8. Whitmore SE, Lamont RJ. Oral bacteria and cancer. PLoS Pathog. 2014 Mar 27;10(3):e1003933.
9. Patil S, Yadalam PK, Hosmani J, Khan ZA, Shankar VG, Shaukat L, et al. Modulation of oral cancer and periodontitis using chemotherapeutic agents - A narrative review. Dis Mon. 2023 Jan;69(1):101348.
10. Joseph B, Yadalam PK, Anegundi RV. Management of oral lesions following COVID-19 vaccination. Oral Dis. 2022 Nov;28 Suppl 2:2634-2635.
11. Patil S, Khan SS, Hosmani J, Khan ZA, Muruganandhan J, Mushtaq S, et al. Identification of oral immune disorders- A review and a diagnostic algorithm. Dis Mon. 2023 Jan;69(1):101350.
12. Patil S, Yadalam PK, Hosmani J, Khan ZA, Ahmed ZH, Shankar VG, et al. Oral immune-mediated disorders with malignant potential/association: An overview. Dis Mon. 2023 Jan;69(1):101349..
13. Patil S, Mustaq S, Hosmani J, Khan ZA, Yadalam PK, Ahmed ZH, et al. Advancement in therapeutic strategies for immune-mediated oral diseases. Dis Mon. 2023 Jan;69(1):101352.
14. Yadalam PK, Trivedi SS, Krishnamurthi I, Anegundi RV, Mathew A, Al Shayeb M, Narayanan JK, Jaber MA, Rajkumar R. Machine learning predicts patient tangible outcomes after dental implant surgery. IEEE Access. 2022 Dec 12;10:131481-8.
15. Kumar VS, Kumar PR, Yadalam PK, Anegundi RV, Shrivastava D, Alfurhud AA, et al. Machine learning in the detection of dental cyst, tumor, and abscess lesions. BMC Oral Health. 2023 Nov 6;23(1):833.
16. Kekre HB, Gharge SM, Sarode TK. Image segmentation of MRI images using vector quantization techniques. InThinkquest~ 2010: Proceedings of the First International Conference on Contours of Computing Technology 2011 (pp. 171-176). New Delhi: Springer India.
17. Banu AF, Kayalvizhi M, Arumugam B, Gurunathan U. Texture based classification of dental cysts. In2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT) 2014 Jul 10 (pp. 1248-1253). IEEE.
18. Mohanty AK, Senapati MR, Beberta S, Lenka SK. Texture-based features for classification of mammograms using decision tree. Neural Computing

and Applications. 2013 Sep;23(3):1011-7.
19. Haddon JF, Boyce JF, Strens M. Autonomous segmentation and neural network texture classification of IR image sequences. InIEE Colloquium on Image Processing for Remote Sensing 1996 Feb 13 (pp. 6-1). IET.
20. Mohammadi E, Makkiabadi B, Shamsollahi MB, Reisi P, Kermani S. Wavelet-Based Biphase Analysis of Brain Rhythms in Automated Wake-Sleep Classification. Int J Neural Syst. 2022 Feb;32(2):2250004.
21. Theodorou SJ, Theodorou DJ, Sartoris DJ. Imaging characteristics of neoplasms and other lesions of the jawbones: part 1. Odontogenic tumors and tumorlike lesions. Clinical imaging. 2007 Mar 1;31(2):114-9.
22. Lin JH, Cheng TY. Dynamic clustering using support vector learning with particle swarm optimization. In18th International Conference on Systems Engineering (ICSEng'05) 2005 Aug 16 (pp. 218-223). IEEE.
23. ChandraPrabha K, Bharathi C. Texture analysis using GLCM & GLRLM feature extraction methods. Int J Res Appl Sci Eng Technol (IJRASET). 2019 May;7(5):2059-64.
24. Gadodia A, Gamanagatti S, Neyaz Z. Segmental cystic renal disease: Sonographic and CT findings. Journal of Clinical Ultrasound [Internet]. 2009 Sep;37(9):525–526. Available from: http://dx.doi.org/10.1002/jcu.20636
25. Chen EL, Chung PC, Chen CL, Tsai HM, Chang CI. An automatic diagnostic system for CT liver image classification. IEEE Trans Biomed Eng [Internet]. 1998 Jun;45(6):783–794. Available from: http://dx.doi.org/10.1109/10.678613
26. Reddy TK, Kumaravel N. Wavelet based texture analysis and classification of bone lesions from dental CT. International Journal of Medical Engineering and Informatics. 2010 Jan 1;2(3):319-27.
27. Ji D, Xue X, Xu C. Truncated total variation in fractional B-spline wavelet transform for micro-CT image denoising. J Xray Sci Technol. 2023;31(3):555-572.
28. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med Phys. 2017 Oct;44(10):e360-e375.
29. Li M, Wang Y, Sun H. Single-Frame Infrared Image Non-Uniformity Correction Based on Wavelet Domain Noise Separation. Sensors (Basel). 2023 Oct 12;23(20):8424.
30. Diamanti N, Duxbury AJ, Ariyaratnam S, Macfarlane TV. Attitudes to biopsy procedures in general dental practice. British dental journal. 2002 May;192(10):588-92.
31. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006 Jun 1;27(8):861-74.
32. Li HS, Fan P, Peng H, Song S, Long GL. Multilevel 2-D Quantum Wavelet Transforms. IEEE Trans Cybern. 2022 Aug;52(8):8467-8480.
33. Zhang Z, Huang M, Jiang Z, Chang Y, Torok J, Yin FF, et al. 4D radiomics: impact of 4D-CBCT image quality on radiomic analysis. Phys Med Biol. 2021 Feb 11;66(4):045023.
34. Yu Y, She K, Liu J, Cai X, Shi K, Kwon OM. A super-resolution network for medical imaging via transformation analysis of wavelet multi-resolution. Neural Netw. 2023 Sep;166:162-173.
35. Tong C, Pang Y, Wang Y. HIWDNet: A hybrid image-wavelet domain network for fast magnetic resonance image reconstruction. Comput Biol Med. 2022 Dec;151(Pt A):105947.
36. Bhardwaj J, Nayak A. Haar wavelet transform-based optimal Bayesian method for medical image fusion. Med Biol Eng Comput. 2020 Oct;58(10):2397-2411.
37. Zhang Y, Ding W, Pan Z, Qin J. Improved Wavelet Threshold for Image De-noising. Front Neurosci. 2019 Feb 8;13:39.
38 Sivakani R, Ansari GA. Imputation using machine learning techniques. In2020 4th International conference on computer, communication and signal processing (ICCCSP) 2020 Sep 28 (pp. 1-6). IEEE.