Content available at: https://www.ipinnovative.com/open-access-journals

Journal of Dental Specialities

Journal homepage: https://www.jdsits.in/

Original Research Article

Automated detection of oral potential malignant disorders using exfoliative cytology

Heena Zainab[1], Ameena Sultana[1]*, Rajmohan Pardeshi[2]

[1]Dept. of Oral Pathology and Microbiology, Al-Badar Dental College and Hospital, Kalaburagi. Rajiv Gandhi University of Health Sciences, Bengaluru, Karnataka, India

[2]Dept. of Computer Science, KRE Society's Karnatak Arts, Science and Commerce College, Bidar, Karnataka, India

Abstract

Background: Oral exfoliative cytology serves as a non-invasive diagnostic tool for detecting cellular abnormalities. However, manual analysis can be time-consuming and prone to subjective interpretation. With advancements in Artificial Intelligence (AI), automated systems offer promising solutions for improving diagnostic accuracy and efficiency.

Aim: This study aims to evaluate the effectiveness of AI-based techniques, including machine learning and deep learning models, for classifying normal and abnormal oral exfoliative cells through cytomorphological analysis.

Materials and Methods: The study employed two AI approaches. The first involved use of cellular and nuclear dimensions, such as cell and nuclear diameters, which were analyzed using a Decision Tree classifier. The second method utilized a deep learning model based on the AlexNet Convolutional Neural Network (CNN) architecture for image-based classification. Grad-CAM visualizations were used to identify biologically significant regions contributing to the classification.

Results: The Decision Tree classifier, based on cytomorphometric features, achieved an accuracy of 100% in distinguishing normal from abnormal cells. The AlexNet-based CNN model achieved a classification accuracy of 93%. Grad-CAM results provided interpretability by highlighting relevant morphological areas in the cytological images.

Conclusion: The study demonstrates that integrating cytological morphometry with AI techniques significantly enhances the accuracy and interpretability of oral cell abnormality detection. These findings support the potential of AI-assisted cytology as a reliable tool for early screening of oral pathologies.

Keywords: Decision tree classifier, Deep learning, Morphometric analysis, Exfoliative cytology, Oral malignant disorders

Received: 12-05-2025; **Accepted:** 13-08-2025; **Available Online**: 29-09-2025

1. Introduction

Cancer remains a major public health challenge globally and is one of the leading causes of mortality. According to GLOBOCAN 2020, oral cancer accounts for an estimated 377,713 new cases annually. In India, oral cancer represents a significant cause of morbidity and mortality in the oral and maxillofacial regions. Oral squamous cell carcinoma (OSCC), comprising over 90% of all malignant neoplasms of the oral cavity, is the sixth most common cancer worldwide and the third most common in India.[1] Tobacco consumption, alcohol use either independently or in combination with chewing tobacco and betel quid are well established carcinogens contributing to the high incidence of OSCC. Additionally, a major yet often overlooked contributor to the global cancer burden is outdoor air pollution. Emissions from industrial activities, power generation, transportation, and domestic burning significantly degrade air quality. Due to their fine particulate size, these pollutants can be deeply inhaled into the lungs and bronchi, posing severe health risks.[2,3] Exposure to outdoor air pollution has been associated with increased hospital admissions for cardiovascular and respiratory conditions and heightened mortality rates. Furthermore, air pollution is recognized as both carcinogenic and mutagenic, largely due to the presence of polycyclic aromatic hydrocarbons (PAHs), dioxins, 3-nitrobenzanthrone, and sulfur-containing compounds. PAHs are known to form DNA adducts, thus increasing the risk of

*Corresponding author: Ameena Sultana
Email: mrsabdulmajid@gmail.com

https://doi.org/ 10.18231/j.jds.76024.1759127383

various human cancers.[4] Several studies have also linked air pollution to epigenetic modifications and altered gene expression patterns, further implicating it in cancer development. However, limited research has focused specifically on the association between air pollution and oral cancer.

Research published in[5] challenges traditional views on the etiological factors of oral cancer. The findings suggest that elevated levels of air pollution may contribute to malignancies of the lips, tongue, cheeks, and palate. This underscores that ambient air may pose more serious health risks than previously assumed, extending beyond respiratory ailments to oncological threats. This suggests that early detection and diagnosis of oral diseases is crucial, and therefore, artificial intelligence (AI) presents a promising approach for developing tools that enable rapid identification of abnormalities through pathological tissue imaging.

The integration of artificial intelligence (AI) into diagnostic pathology has shown promising potential in the early detection and classification of oral potentially malignant disorders (OPMDs) and oral squamous cell carcinoma (OSCC).[6] In this study, we utilize exfoliative cytology images to explore the potential of morphometric parameters such as nuclear and cytoplasmic dimensions as reliable biomarkers for the early detection of oral cancer. Given its non-invasive nature and cost-effectiveness, exfoliative cytology remains a valuable diagnostic tool for the early identification of premalignant and malignant oral lesions.[7] To evaluate the significance of these morphometric features, we adopted two approaches: first, a decision tree classifier was used to assess their diagnostic potential through traditional machine learning, and finally, AlexNet, a deep learning architecture, was applied to leverage convolutional feature extraction for automated classification.

2. Materials and Methods

2.1. Details of the dataset

The present study was conducted on exfoliative cytology samples collected from healthy individuals. A total of 50 subjects were included in this pilot study based on sample size estimation, comprising 30 individuals in the case group and 20 in the control group diagnosed with oral squamous cell carcinoma (OSCC). The case group consisted of healthy individuals aged between 25 and 40 years, with no history of tobacco consumption and no clinically evident oral mucosal pathology. Participants underwent a comprehensive oral clinical examination, following which exfoliative cytology samples were obtained. Cellular and nuclear diameters were measured from cells captured at 40x magnification across 10 fields per slide using LYNX AUTO BIOLUX image analysis software. Cell diameter and nuclear size, these parameters were selected to evaluate potential morphological alterations in oral exfoliative cells 30 case samples and 20 control samples. The captured cytological images were then analyzed using a deep learning pipeline, employing Convolutional Neural Networks (CNN) based on the AlexNet architecture, with Grad-CAM visualization techniques used to interpret and highlight discriminative regions in the image classification process. Inclusion criteria for the case group required the absence of oral pathology and tobacco habits, while subjects with any oral lesions or a history of tobacco use (chewing or smoking) were excluded.

2.2. Decision tree classifier

Decision Tree classifier was implemented using two key morphometric parameters cellular diameter and nuclear diameter extracted from exfoliative cytology images. These features were selected due to their well established diagnostic relevance in distinguishing normal epithelial cells from those undergoing dysplastic or malignant transformation. The Decision Tree algorithm,[8] constructs a flowchart-like model that recursively splits the dataset based on feature thresholds that maximize information gain or reduce impurity, allowing for interpretable rule-based classification. In this study, the classifier effectively learned decision boundaries based on variations in cell and nuclear size, enabling the discrimination between case group and control group samples. This approach offered a transparent and explainable model for understanding the morphological criteria associated with oral malignancy risk.

2.3. AlexNet

Given the limited size of the exfoliative cytology image dataset, transfer learning allowed effective feature extraction by leveraging the robust low-level and mid-level features learned from a large, diverse image corpus.[9] For the classification task, a transfer learning approach was employed using the AlexNet architecture, a deep convolutional neural network originally developed and pre-trained on the ImageNet dataset.[10] The final fully connected layers of AlexNet were fine-tuned to accommodate a binary classification task-differentiating between cytological images of street vendors (case group) and those diagnosed with oral squamous cell carcinoma (control group). Input images were resized to 227x227 pixels to match the input requirements of the network. To enhance generalization and prevent overfitting, data augmentation techniques such as rotation, flipping, and contrast normalization were applied.[11] Furthermore, Grad-CAM (Gradient-weighted Class Activation Mapping) was utilized to visualize the discriminative regions within the images that contributed most significantly to the network's predictions, thus improving model interpretability and clinical relevance.[12]

3. Results

The Decision Tree classifier given in[13] demonstrated 100% accuracy in distinguishing between normal 2 of 7 and abnormal classes using cellular morphometric parameters, specifically cell diameter and nuclear size. The high classification accuracy is further supported by the Confusion

Matrix given in **Figure 1**(a), which indicates perfect classification performance, with no misclassifications in either class. Additionally, the ROC Curve given in **Figure 1**(b), confirms the classifier's excellent performance, with an area under the curve (AUC) close to 1, signifying a near perfect ability to discriminate between the normal and abnormal categories based on the selected morphometric features.

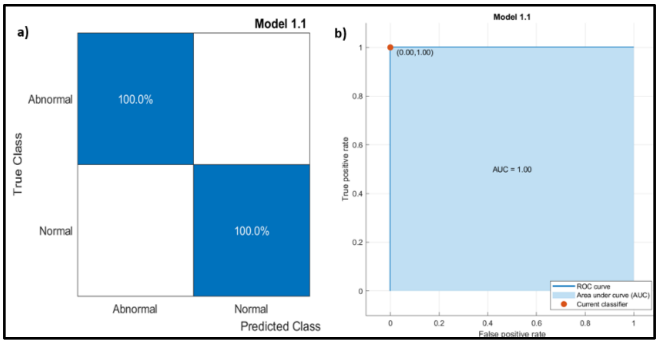

Figure 1: a: Confusion matrix of Decision Tree classifier and **b:** ROC curve illustrating the classifier's performance.

The classification performance of the AlexNet-based Convolutional Neural Network (CNN) was evaluated using key metrics including precision, recall, F1-score, and overall accuracy, as shown in **Table 1**. The model achieved an impressive overall accuracy of 93%, indicating high reliability in distinguishing between normal and abnormal oral exfoliative cytology images.

Table 1: Precision, Recall, F1-score, accuracy, and macro average for the for the alexnet based cytology image classification

Class	Precision	Recall	F1-Measure
Abnormal	1.00	0.86	0.92
Normal	0.89	1.00	0.94
Macro-average	0.94	0.93	0.93
Average Accuracy	93		

The precision for the abnormal class was 1.00, meaning that all predictions labeled as abnormal were correctly identified, with no false positives. However, the recall for this class was 0.86, suggesting that a few abnormal samples were misclassified as normal. In contrast, the normal class showed a precision of 0.89 and a recall of 1.00, indicating that all actual normal samples were correctly retrieved, though a small number of predictions labeled as normal may have been incorrect. The balanced performance across both classes is reflected in the macro average F1-score of 0.93, confirming the model's generalizability across the dataset.

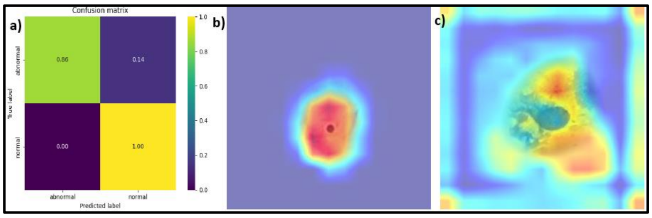

Figure 2: Visualization of Model Interpretability and Performance: **a:** Confusion matrix summarizing classification results of the AlexNet based CNN model; **b:** Grad-CAM high- lighting discriminative regions in a normal epithelial cell, and; **c:** Grad-CAM focusing on key features in an abnormal (OSCC) cell.

The confusion matrix, shown in **Figure 2**(a), visually reinforces the numerical results by displaying the correct and incorrect predictions for each class. The majority of samples were correctly classified, with only a few misclassifications observed in the abnormal category. This aligns with the slightly lower recall for the abnormal class.

To further validate the interpretability of the model, Grad-CAM visualizations were generated, as illustrated in **Figure 2**(b) and **Figure 2**(c). These heatmaps highlight the regions of the cell images that contributed most to the model's decision. In the abnormal cell image, the Grad-CAM focused on nuclear irregularities and atypical morphological patterns, whereas in the normal cell image, the attention was distributed around regular, uniformly shaped nuclei. This supports the model's biologically meaningful learning process.

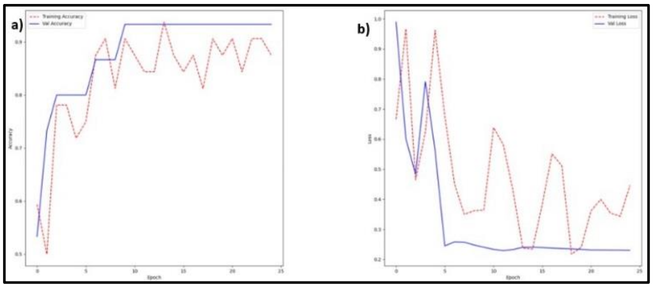

Figure 3: Performance of the CNN model using AlexNet architecture; **a:** Accuracy curve showing model performance on training and validation datasets; **b:** Loss curve depicting the error reduction over epochs.

The training and validation accuracy/loss curves given in **Figure 3** indicate smooth convergence without overfitting, as the validation performance closely follows the training performance. The final validation accuracy stabilized near the training accuracy as given in **Figure 3**(a), and the loss curves given in **Figure 3**(b) show a consistent decrease, confirming that the model was well-trained and capable of generalizing to unseen data.

4. Discussion

India experiences a high incidence of oral cancer, leading to an annual expenditure of $386 million due to productivity loss surpassing economically growing countries. Healthcare services in rural areas have limited access to pathology services and expertise, causing delayed diagnosis, increased treatment costs, and reduced survival rates.[14] The use of AI offers an opportunity to improve accuracy for timely intervention and appropriate management strategies. The deleterious role of air pollutants on respiratory tract is well known. This study highlights the role of Air Pollutants as the carcinogen and creating societal awareness 4 of 7 against the possible causative for oral cancer besides the role of tobacco and other known carcinogens. In our study statistically significant differences in nuclear diameters between oral cancer control and the case group through cyto-morphometric evaluation were noted indicating the role of air pollution in causing cellular changes. To the best of our knowledge this is the pioneering study to evaluate mucosal changes in street vendors through AI. Early efforts have focused on the use of machine learning models for cytological image analysis.

For instance, in,[15] employed support vector machines (SVMs) to classify exfoliative cytology images of oral lesions, achieving notable accuracy in distinguishing malignant from benign cells. Similarly in[16] demonstrated the use of cytomorphometric analysis with satisfactory performance in detecting OSCC. Deep learning has further advanced the field by enabling automatic feature extraction from histopathological images.[17] Utilized convolutional neural networks (CNNs) to detect cancerous regions in oral tissue images, reporting robust performance with minimal preprocessing. Building upon this,[18] applied transfer learning using pre-trained CNN architectures for classifying OSCC, significantly reducing the need for large annotated datasets. Recent work has also explored the combination of clinical data with image-based AI models.[19] Integrated patient metadata with cytology images in a hybrid AI framework to improve diagnostic reliability in OPMD screening. Moreover,[20] proposed an ensemble learning approach combining deep features with traditional machine learning classifiers, yielding superior classification performance for oral lesion grading.[21]

These studies collectively suggest that AI, particularly deep learning, holds significant promise in augmenting the early detection and diagnosis of oral malignant disorders. In this study, we utilize exfoliative cytology images to explore the potential of morphometric parameters such as nuclear and cytoplasmic dimensions-as reliable biomarkers for the early detection of oral cancer. Exfoliative cytology, being a non-invasive diagnostic technique, is widely recognized for its utility in identifying premalignant and malignant lesions at an early stage.[17] Previous studies have established that changes in cellular morphology are indicative of pathological transformations.[21-24]

We incorporate machine learning algorithms and deep learning models to classify cellular images, thereby enhancing diagnostic precision through automated, image-based approaches. From a machine learning standpoint, the application of a Decision Tree classifier demonstrated perfect accuracy in distinguishing normal from abnormal cytology samples using just two morphometric features: nuclear and cellular diameter. This supports the hypothesis that even simple machine learning models can yield clinically useful results when robust and relevant biomarkers are selected. Furthermore, the deep learning model using the AlexNet architecture achieved an overall classification accuracy of 93%. The precision and recall metrics across both classes were high, with the Grad-CAM visualizations reinforcing the biological relevance of the model's decisions by focusing on nuclear regions exhibiting morphological anomalies.

These results emphasize the potential of combining traditional cytomorphometric techniques with AI driven analysis for non-invasive, early-stage screening of individuals at risk due to environmental exposures. The integration of explainable AI techniques, such as Grad-CAM,[12] not only enhances model trust but also aligns clinical observations with machine-learned features-a step toward building reliable diagnostic tools in resource limited settings. Nonetheless, the pilot nature of this study and the relatively small sample size limit the generalizability of findings. Future work should involve a larger and more diverse cohort to validate these observations and explore additional features, such as nuclear contour irregularities, chromatin texture, and mitotic figures, to enhance classification robustness. Integration with patient metadata and real-time pollutant exposure data could further improve predictive accuracy and clinical utility.

5. Conclusions

This study demonstrates that morphometric analysis, particularly nuclear enlargement, can serve as an early indicator of cytological atypia. By combining these measurable cellular features with AI-based image classification and machine learning algorithms such as Decision Trees and Convolutional Neural Networks (CNNs), it is possible to accurately distinguish between normal and ab- normal exfoliative cell samples. The integration of morphometric measurements with image- based deep learning models presents a powerful, non-invasive approach for early detection of oral cellular abnormalities. Our findings further suggest that environmental pollutants may contribute to early cytological changes associated with oral carcinogenesis, extending their risk beyond respiratory impact. This highlights the importance of implementing AI-assisted cytological screening programs, especially targeting vulnerable populations like street vendors and outdoor workers, etc. Such early intervention strategies can play a crucial role in timely diagnosis, prevention, and the enhancement of public health outcomes in high risk environments.

6. Ethical Approval

Consent for treatment and open access publication was obtained or waived by all participants in this study. Institutional Ethics Committee (IEC) Al-Badar Rural Dental College and Hospital, Kalaburagi issued approval No./IEC/2021-22/38.

7. Source of Funding

None.

8. Conflict of Interest

None.

9. Acknowledgements

Authors would like to thank Madiha Zufa and Bibi Ayesha, Students of BDS for Supporting Data Acquisition.

References

1. Jain AK. Oral Cancer Screening: Insights into Epidemiology, Risk Factors, and Screening Programs for Improved Early Detection. *Cancer Screen Prev*. 2024;3(2):97-105.
2. Lin WJ, Jiang RS, Wu SH, Chen FJ, Liu SA. Smoking, alcohol, and betel quid and oral cancer: a prospective cohort study. *J Oncol*. 2011;2011:525976.
3. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers*. 2020;6(1):92.
4. Turner MC, Andersen ZJ, Baccarelli A, Diver WR, Gapstur SM, Pope-3rd CA, et al. Outdoor air pollution and cancer: An overview of the current evidence and public health recommendations. *CA Cancer J Clin*. 2020:21632
5. Chu YH, Kao SW, Tantoh DM, Ko PC, Lan SJ, Liaw YP: Association between fine particulate matter and oral cancer among Taiwanese men. *J Investig Med*. 2019;67(1):34–8.
6. Chowdhury T, Kasralikar P, Aleem Syed A, Tumati R, Patel S, Kommineni D. Role of Artificial Intelligence in the Diagnosis of Oral Squamous Cell Carcinoma: A Systematic Review. *Cureus.* 2025;17(4):81800.
7. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol*. 2009;45(4-5):309–16.
8. Quinlan JR. Induction of decision trees. *Machine Learning*. 1986;1:81–106.
9. Kaur G, Sharma N. Automated Detection of Oral Squamous Cell Carcinoma using Transfer Learning Models from Histopathological Images. 3rd International Conference for Advancement in Technology (ICONAT) GOA, India. 2024;1:6. 10.1109/ICONAT61936.2024.10775135
10. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans Med Imaging.* 2016;35(5):1285–98.
11. Perez L, Wang J. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *Computer Sci viXra*. 2017.
12. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM Visual Explanations from Deep Networks via Gradient-Based Localization. IEEE International Conference on Computer Vision (ICCV), Venice, Italy. 2017;618:626.
13. Pyo JS, Sohn JH, Kang G. Diagnostic assessment of intraoperative cytology for papillary thyroid carcinoma: using a decision tree analysis. *J Endocrinol Invest*. 2017;40(3):305–11.
14. Bakshi SS, Kalidoss VK. A Need to Estimate the Economic Burden of Oral Cancer in India. *Ann Maxillofac Surg*. 2021;11(1):214.
15. Mhaske S, Ramalingam K, Nair P, Patel S, Menon P A, Malik N, Mhaske S. Automated Analysis of Nuclear Parameters in Oral Exfoliative Cytology Using Machine Learning. *Cureus*. 2024;16(4):58744.
16. Verma R, Singh A, Badni M, Chandra A, Gupta S, Verma R. Evaluation of exfoliative cytology in the diagnosis of oral premalignant and malignant lesions: A cytomorphometric analysis. *Dent Res J*. 2015;12(1):83–8.
17. Aubreville M, Knipfer C, Oetter N, Jaremenko C, Rodner E, Denzler J, et al. Automatic Classification of Cancerous Tissue in Laser endomicroscopy Images of the Oral Cavity using Deep Learning. *Sci Rep*. 2017;7(1):11979.
18. Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Netw*. 2020;128:47–60.
19. Wang W, Liu Y, Wu J. Early diagnosis of oral cancer using a hybrid arrangement of deep belief network and combined group teaching algorithm. *Sci Rep*. 2023;13;13:22073.
20. Sukegawa S, Tanaka F, Nakano K, Hara T, Ochiai T, Shimada K, et al. Training high-performance deep learning classifier for diagnosis in oral cytology using diverse annotations. *Sci Rep*. 2024;14(1):17591.
21. Alsarraf A, Kujan O, Farah CS. Liquid-based oral brush cytology in the diagnosis of oral leukoplakia using a modified Bethesda Cytology system. J Oral Pathol Med. 2018;47(9):887–94.
22. Koss LG. The Papanicolaou test for cervical cancer detection. A triumph and a tragedy. *JAMA*. 1989;261(5):737–43.
23. Pina PSS, Mendes AT, Correa L, Coracin FL, Sousa SCOM. The role of exfoliative cytology in diagnosis of oral lesions. *Clin Oral Investig*. 2024;29(1):2.
24. Wang S, Zheng Z. Advances in Oral Exfoliative Cytology: From Cancer Diagnosis to Systemic Disease Detection. *Diagn Cytopathol*. 2024;52(11):697–706.

Cite this article: Zainab H, Sultana A, Pardeshi R. Automated detection of oral potential malignant disorders using exfoliative cytology. *J Dent Spec*. 2025;13(2):210-214.