International Journal of Basic and Applied Sciences, 14 (1) (2025) 377-387

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/8c1tmt79
Research paper

AI-powered classification of oral pre-cancer: a histopathological image approach

Dr. Sharmila Sengupta [1, 2] *, Dr. Priya R. L. [1, 2], Anuj Bagad [1, 2], Aayush Talreja [1, 2], Mansi Bellani [1, 2], Dr Harsha Karwa [3], Dr Shrijha G [3]

[1] *Department of Computer Engineering*
[2] *Vivekanand Education Society's Institute of Technology, Chembur, Mumbai, India*
[3] *Department of Oral Pathology and Microbiology, Government Dental College & Hospital, Mumbai, India*
**Corresponding author E-mail: etechday@gmail.com*

Received: April 23, 2025, Accepted: May 18, 2025, Published: May 24, 2025

Abstract

A high fatality rate characterises the prognosis outlook for oral cancer due to delayed diagnosis and significantly hinders the advancements in early detection techniques. Traditional histological investigation, a cornerstone of precancerous lesion diagnosis, faces substantial challenges in terms of labour intensity in manually processing tissue samples, is time-consuming and the inferences are mostly guided by microscopic investigation. Therefore, a system is proposed for oral dysplasia grading from histopathological tissues using machine learning. Distinguished from other methodologies, this approach incorporates a comprehensive array of features, not limited to cell nuclei properties but also morphometric analysis of epithelial stratification and enhancements in digital pathology. Moreover, early pre-cancer grading enhances the survival chances of patients by identifying abnormalities that might be missed through conventional examination. The study presented here has an overall accuracy of 94%, considering the limited training cases available under each feature. This approach would benefit both doctors and patients by streamlining the diagnostic process and improving its outcomes. The fundamentals of oral cancer surgical management, when it comes to risk factors, regional predispositions, treatment response, and outcome, oral cavity squamous cell carcinoma (OSCC) is a heterogeneous illness. Surgery is the main treatment option for oral malignancies, even though non-surgical treatment is used in other head and neck subsites. Adjuvant treatment, such as radiation or chemoradiation, is then administered based on the risk factors of the final histopathology. Before beginning treatment, a multidisciplinary tumor board must discuss each patient and develop a treatment strategy. This chapter aims to explain the fundamental surgical concepts used in the reconstruction of defects and the management of various oral cavity subsites. The surgical management the technical process of brain arteriovenous malformation (bAVM) excision, patient selection for surgery, and perioperative care that optimizes the likelihood of the best result are all included in surgical treatment.

Keywords: *Atypical Mitotic Figures; Irregular Epithelial Stratification; Hyperchromasia; Apoptotic Mitoses; Keratinization; Loss of Polarity of Basal Cells; Keratin Pearls; Rete Ridges; Histopathological Images; H&E Stains; Template Matching; Digital Pathology.*

1. Introduction

An important worldwide health concern, oral cancer, primarily presented as Oral Squamous Cell Carcinoma (OSCC), develops from the epithelial cells. Oral epithelial dysplasia, the precursor to this cancer, indicates an early precancerous stage characterised by cellular and architectural abnormalities within the epithelium. The term "dysplasia," derived from the ancient Greek words "dys" for "bad" and "plasis" for "formation," accurately characterises the aberrant cellular processes that predispose to OSCC. A complex interaction between genetic and environmental factors drives these processes, forming many atypical epithelial layers.

Compared to their non-dysplastic counterparts, oral lesions exhibiting dysplastic alterations are substantially contributing to the advancements into OSCC, highlighting the critical need for an early and precise diagnosis. Traditionally, a thorough evaluation of cytological characteristics, including nuclear and cellular pleomorphism, as well as architectural abnormalities such as uneven stratification and keratinisation patterns, has been used to grade oral epithelial dysplasia. Although the grading procedure is divided into discrete classes (hyperplasia, mild, moderate, severe, and cancer in situ), it is very subjective. It frequently leads to significant inter- and intra-observer variability. Such variation emphasises the urgent need for more impartial and trustworthy diagnostic techniques.Machine Learning and digital pathology have ushered in a new era in histopathological examination. High-resolution annotated digitised images of whole slide images (WSIs) from experienced pathologists provide a comprehensive view of tissue samples and allow ML algorithms to be applied for automated grade prediction. This technological advancement is expected to reduce the subjective nature of manual grading, enabling more reliable and effective diagnostic procedures. These algorithms are used in conjunction with a morphometric examination of epithelial stratification, which examines the thickness and arrangement of the keratin, basal, and epithelial layers. This method aims to automate the prediction of

dysplasia grades with unprecedented levels of objectivity and accuracy.Grading is based on feature extraction of the histopathological images. The ones observed in the lower third of the image are classified as mild, the middle as moderate, and features in the top third are deemed severe. In this research, after individual grading, a final judgement for mild, moderate, or severe grades is made based on majority voting. Individual grading is done based on extracted features from the training dataset using a Support Vector Machine(SVM) classifier, a promising ML method for efficient classification. The WSIs are carefully monitored and segregated by the Government Dental College & Hospital, Mumbai, Maharashtra, India, doctors. The datasets collected from the hospital include samples of normal epithelium and a range of dysplasia grades. The ML-based diagnostic approach improves the accuracy of diagnosis and substantially contributes to optimising patient care and therapeutic strategies.

2. Literature review

Several research papers have been consulted to develop the tool described in this paper. Gupta, R. K., Manhas, J., & Kour M. [1] developed a Hybrid Feature Extraction Based Ensemble Classification Model for diagnosing Oral Carcinoma using Histopathological Images, employing only selective 5 textural (mean, standard deviation, third moment, uniformity, smoothness, and entropy) feature extraction with algorithms like CNN, GLCM, and LBP alongside classifiers such as SVM, KNN, and Naïve Bayes with a Boosted Tree. Despite achieving an accuracy of 98.71%, this approach may lack generalizability across diverse datasets due to potential overfitting, and the integration of various algorithms could introduce complexity in clinical applications. Meanwhile, Panigrahi, S., Nanda et al. [2] utilised deep transfer learning to classify histological pictures of oral squamous cell carcinoma, proposing a baseline CNN model and using pre-trained DCNNs such as VGG16 and ResNet50. Although beneficial for accuracy, the reliance on pre-trained networks might not fully capture the unique characteristics of oral cancer histopathology, limiting the model's sensitivity to subtle yet critical features. Rahman AU et al. [3] focused on a histopathological image dataset, employing the AlexNet CNN model, though their study was limited to a binary classification of OSCC presence. Advancing the field, R. A. Welikala et al. [4] developed a CNN-based system for oral lesion detection and classification without grading severity or accounting for imaging condition variations. Nanditha B R et al. further explored the potential of machine learning and deep learning models [5] using a lesion image dataset, yet did not address the grading of oral cancer severity. Adding to the body of knowledge, Saad Bashir, R.M., Mahmood, H., Shaban, M., et al. presented a study on automated grade classification of oral epithelial dysplasia using morphometric analysis of histology images [6]. Despite showcasing a novel approach in digital pathology, the focus on automated grading through morphometric analysis alone may not fully encapsulate the complex interplay of histological features indicative of dysplasia progression. Khanagar et al. [7], in their systematic review, assessed the efficiency of AI in the realm of oral cancer through an analysis of histopathological images. Here, AI is used for diagnosis, classification, and prediction of oral cancer (OC) using histopathological images, unlike the model presented here for oropharyngeal squamous cell carcinoma (OPC). In the review [8], the emphasis on deep learning (DL) presents a contrast to machine learning (ML) approaches in oral cancer diagnosis and prognosis. Dixit et al. (2023), in paper[9] delve into the advancements and challenges in employing AI, particularly ML and DL, for oral cancer diagnosis. This comprehensive analysis underscores the potential AI holds in revolutionising early detection and treatment methodologies. However, it also brings to light the hurdles, such as data scarcity, model generalisation, and the integration of AI into clinical workflows that need to be overcome. Most of the literature papers referred to have used deep learning techniques, which are time and processing complex.In this context, the work presented here outlines a future direction for oral cancer diagnosis by employing computer vision to overcome the limitations of conventional histological techniques. The development of the HistoGrade Predictor Tool aims to improve therapeutic outcomes through prompt and accurate diagnosis. Given the challenges oral pathologists face with the complexity and overlapping features of head and neck tumors, AI tools can indeed serve as valuable support for objective interpretation, enhancing diagnostic accuracy and aiding prognosis[10] Recent studies underscore that AI models, once trained on large and diverse datasets, can distinguish subtle morphological features in digital slides, assisting pathologists in identifying oral squamous cell carcinoma (OSCC) with higher precision. However, further research is necessary for clinical validation to ensure these AI. An AI medical imaging model may miss a pertinent relationship between data inputs (features) and intended outputs (predictions) during the modeling phase due to systematic errors brought on by incorrect assumptions about the data.

The ORCHID (ORal Cancer Histology Image Database) discussed in [11] is a notable advancement aimed at supporting AI-based research for the histopathological analysis of oral cancer, particularly for challenging cases like oral squamous cell carcinoma (OSCC) and oral submucous fibrosis (OSMF). Given the global burden of oral cancer and the diagnostic hurdles posed by histopathological analysis, such as inter-observer variability and the limited availability of experienced pathologists, ORCHID provides a valuable resource. Such a rich, annotated dataset could be transformative for training deep learning models that can accurately classify and grade oral cancer types, aiding in faster and potentially more accurate diagnoses.

An AI-driven approach is presented in [12] for assessing the risk of malignant transformation in oral epithelial dysplasia (OED), addressing a significant limitation in current OED grading systems. Traditional grading suffers from variability and has limited prognostic reliability, which can result in inconsistent treatment plans. By developing an AI algorithm that produces an Oral Malignant Transformation (OMT) risk score based on Haematoxylin and Eosin (H&E) stained whole slide images, your team offers a novel, data-driven method for more objective assessment. The AI pipeline’s segmentation model, which detects and segments nuclei and epithelium, serves as a foundation for the OMT score.

The critical issue of grading OED is time-intensity and often yields inconsistent results among pathologists. By developing the E-MOD-plus system[13], an objective and efficient AI-based grading tool for OED, your team provides a promising solution to improve diagnostic consistency and speed. The methodology involved comparing four convolutional neural networks (CNNs) to identify the most effective model, EfficientNet-B0, which was selected based on its performance on labeled image patches from 56 whole-slide images of oral leukoplakia. Using this network, feature detection models were trained, validated, and tested on 1,000 image patches, culminating in the creation of E-MOD-plus. This comprehensive system combines the feature detection models with a multiclass logistic model, allowing it to perform nuanced grading on whole-slide images.

Head and neck squamous cell carcinoma (HNSCC), presented in paper [15], especially oral cavity squamous cell carcinoma (OCSCC), presents substantial challenges due to its frequent lethality and limited improvements in survival despite therapeutic advances. With OCSCC often arising from premalignant lesions, early detection through screening is particularly valuable for intervention. Currently, the conventional visual and tactile exam (CVTE) combined with biopsy is the gold standard for diagnosis. However, CVTE alone lacks the sensitivity to distinguish between benign reactive lesions and premalignant dysplastic lesions, and histologic grading is subjective with limited prognostic reliability. This subjectivity creates a challenging clinical dilemma, balancing the risks of unnecessary aggressive treatment against potential delays in intervention. Given these issues, the development of innovative diagnostic approaches, such as AI-based

platforms and molecular biomarkers, holds significant promise. AI algorithms could assist by analyzing histological features objectively, potentially increasing diagnostic consistency and predicting malignancy risk with higher accuracy than conventional methods. Concurrently, molecular biomarkers can provide additional layers of prognostic information, offering insights into which lesions are more likely to progress. These novel platforms could reduce the psychological and economic burdens of the waiting and watching approach while improving patient outcomes through earlier and more precise interventions.

With further research and clinical validation, AI and biomarker-based diagnostic tools could transform the current standard of care, supporting more personalized management strategies for OCSCC and reducing uncertainty in treatment planning.

The study described in paper [16] addresses an important clinical need by developing time-to-event models using machine learning to predict the risk of malignancy arising from oral white lesions. Since these lesions often precede the development of mouth cancer, predicting their malignancy risk over time can significantly enhance treatment and monitoring strategies. The models leverage machine learning to estimate the time-factored malignancy risk based on pathological diagnoses, providing a dynamic and personalized risk assessment. After extensive testing, the models demonstrated strong discrimination and calibration, indicating their reliability and potential utility in a clinical setting. This platform could support decision-making in managing patients with oral white lesions, enabling earlier and more accurate intervention strategies.The oral pathology in clinical dentistry is an important area for several reasons. Prevention and early detection: Oral disease progression and consequences can be avoided with routine screenings and early identification. For example, early intervention and improved patient outcomes can result from the detection of precancerous lesions.

3. Oral pre-cancer classification

The accuracy of a system, such as a machine learning model, in properly categorizing oral lesions as either benign or cancerous, or as precancerous phases, is known as classification error in oral pre-cancer. It measures the frequency of inaccurate classifications made by the model by comparing its predictions with the actual diagnoses. Classification Error Calculation: Confusion Matrix. The classification model is performing. It contrasts the actual classes with the expected classes. True Positives (TP): Accurately diagnosed with cancer. True Negatives (TN): Identified as non-cancerous with accuracy. False Positives (FP): A Type I error in which a patient is mistakenly diagnosed with cancer.False Negatives (FN): A Type II error where a test result is incorrectly categorized as non-cancerous. Misclassification Rate or Error Rate. Oral cancer falls under mouth cancer and comes in three grades. When the cancer cells are grade 1 (low grade), they resemble healthy oral cells. The cancer cells in grade 2 (intermediate grade) differ slightly from healthy oral cells in appearance. Grade 3 (high grade) denotes that the cancer cells are very aberrant and do not resemble typical oral cells.

4. Proposed work

AI-powered image analysis, AI-based automated OC identification, and classification emphasize the use of AI techniques, customized algorithms, and predictive models to improve the accuracy of oral cancer diagnosis. intends to solve the OC diagnosis delay and assess the viability of AI-powered methods for automated OC detection and categorization. Machine learning point-of-care screening for oral cancer and other potentially harmful conditions shows encouraging false-positive and sensitivity rates, with a strong AUC of 0.935. demonstrates ML's potential to transform early identification and treatment of illnesses of the oral cavity despite obstacles. Deep learning for finding and categorizing oral lesions shows promise for improving early detection rates, enabling remote patient monitoring, and screening for oral cancer. provides a revolutionary change in direction of an affordable, noninvasive, and easy-to-use oral cancer screening method.

5. Comparative analysis of oral cancer detection and categorization

The oral Cancer Detection and Categorization through machine learning methods, health problems, and it is most common in middle- and low-income nations. The early diagnosis of this condition is crucial for starting the oral cancer treatment process. The main method used to help diagnose cancer is the Convolutional Neural Network (CNN) model, which uses datasets like Visual Object Classification and Common Object in Context. CNN thus contributes to the development of a framework for object detection. It is also used for picture classification. Since the symptoms are moderate, it can be difficult to distinguish this disease from other minor conditions. Early treatment can prevent this disease detection process and save many lives. Cancer can be cured, and its incidence reduced with improved screening and sophisticated detection. The entire method of classifying and identifying oral cancer was clarified by this specific study. Two kinds of approaches, such as automated classification and automated detection process of oral lesions, are based on the deep learning-based detection framework and deep learning-based classification framework. The goal of this study is to identify the many forms of oral cancer, with a focus on buccal mucosa cancer, or inner cheek cancer.

Both oral and buccal mucosa malignancies can be accurately detected with the aid of machine learning algorithms and deep learning [10] procedures. For the machine learning process to be successful, several components must be correctly integrated. The identification and detection technique in this work was carried out using the CNN model and VGG-16. The process's data frequently contains a lot of undesirable elements that should be eliminated before processing, like noise, hazy images, or click angle. The screening procedure uses both normal and histopathological camera images, and data preparation should be completed before the screening procedure. Conversely, feature extraction, data transformation, data cleaning, data standardization, and selection.

This processed data improves the screening process and is used as the final data set. This research focuses on accurate data, thus, the “Convolutional Neural Network” (CNN) model, the method of machine learning [1], and the deep learning approach are employed in this study. To improve the model's performance, biopsy, histopathology, and camera images of malignant and non-cancerous tumors are used independently. With the use of a transfer learning strategy and hyperparameter adjustment, this model produced the most accurate results. For the model, 164 histopathological and photographic images were combined. This study has demonstrated that the model yields the most accurate and superior results from datasets. The rate of detection process accuracy was raised by feature extraction and hyperparameter adjustment. Algorithm for cancer risk the likelihood that a patient with specific symptoms has cancer is determined using statistical formulas known as cancer risk algorithms. Their goal is to assist primary care providers in determining if a patient needs an urgent oncologist's consultation.

6. Materials and methods

An important worldwide health concern, oral cancer, primarily presented as Oral Squamous Cell Carcinoma (OSCC), develops from the epithelial cells. Oral epithelial dysplasia, the precursor to this cancer, indicates an early precancerous stage characterised by cellular and architectural abnormalities within the epithelium. The term "dysplasia," derived from the ancient Greek words "dys" for "bad" and "plasis" for "formation," accurately characterises the aberrant cellular processes that predispose to OSCC. A complex interaction between genetic and environmental factors drives these processes, forming many atypical epithelial layers. Compared to their non-dysplastic counterparts, oral lesions exhibiting dysplastic alterations are substantially proposes on the advancements of OSCC, highlighting the critical need for an early and precise diagnosis. Traditionally, a thorough evaluation of cytological characteristics, including nuclear and cellular pleomorphism, as well as architectural abnormalities such as uneven stratification and keratinisation patterns, has been used to grade oral epithelial dysplasia. Although the grading procedure is divided into discrete classes (hyperplasia, mild, moderate, severe, and cancer in situ), it is very subjective. It frequently leads to significant inter- and intra-observer variability. Such variation emphasises the urgent need for more impartial and trustworthy diagnostic techniques.

Machine Learning(ML) and digital pathology have ushered in a new era in histopathological examination. High-resolution annotated digitised images of whole slide images (WSIs) from experienced pathologists provide a comprehensive view of tissue samples and allow ML algorithms to be applied for automated grade prediction. This technological advancement is expected to reduce the subjective nature of manual grading, enabling more reliable and effective diagnostic procedures. These algorithms are used in conjunction with a morphometric examination of epithelial stratification, which examines the thickness and arrangement of the keratin, basal, and epithelial layers. This method aims to automate the prediction of dysplasia grades with unprecedented levels of objectivity and accuracy.

Grading is based on feature extraction of the histopathological images. The ones observed in the lower third of the image are classified as mild, the middle as moderate, and features in the top third are deemed severe. In this research, after individual grading, a final judgement for mild, moderate, or severe grades is made based on majority voting. Individual grading is done based on extracted features from the training dataset using a Support Vector Machine(SVM) classifier, a promising ML method for efficient classification. The WSIs are carefully monitored and segregated by the Government Dental College & Hospital, Mumbai, Maharashtra, India, doctors. The datasets collected from the hospital include samples of normal epithelium and a range of dysplasia grades. The ML-based diagnostic approach improves the accuracy of diagnosis and substantially contributes to optimising patient care and therapeutic strategies.

In this context, the work outlines a future direction for oral cancer diagnosis by employing computer vision to overcome the limitations of conventional histological techniques. The development of the HistoGrade Predictor Tool represents a significant advancement in the grading of potentially malignant conditions of the mouth, ultimately aiming to improve therapeutic outcomes through prompt and accurate diagnosis.

The materials used in this study include Visual Studio Code or any IDE. The computer code and data associated with this research are publicly available on GitHub at the following repository link for the frontend of the web app: https://github.com/aayu5hgit/histograde-react and for the backend of the web app: https://github.com/aayu5hgit/histograde-flask

7. Methodology

Detecting cancer in its early stages significantly lowers the trauma and stress experienced by patients. The quest for early detection of oral cancer based on several features is pivotal to enhancing patient prognosis and developing various screening methodologies. Traditional systems have demonstrated capability in grading OPC based predominantly on structured data, leaving a gap in the application of unstructured data forms, such as images. The Architectural Features (Irregular Epithelial Stratification, Premature Keratinization in single cell, Keratin Pearls) and Cytological Features (Abnormal Variation in nuclear and cell size and shape, Increased N: C ratio, Atypical Mitotic Figures and Apoptotic mitoses, Increased number and size of nucleoli, Hyperchromasia, Single Cell keratinisation and Premature keratinisation in single cells) The system extracts features individually from the dataset and different methodologies are applied to feed these features to the machine learning model.

The proposed system extracts features from each identified element in the histopathological image, incorporating the cell, nucleus, and cytoplasm. The cell, as can be seen in Fig.1, encapsulates the nucleus, the repository of genetic material, and the cytoplasm, a gel-like substance housing various organelle. The images in the dataset are stained with Hematoxylin and Eosin (H&E) to accentuate cellular details and significantly enhance the visibility of nuclei, enabling the subsequent identification of nuclei for detailed examination. The RGB value of its colour is captured to accurately characterise the nucleus, leveraging the unique colouration of cellular components exhibited when stained with H&E, and particularly, the nucleus presents a distinct colour signature. A predefined range of RGB values is employed to represent this specific nucleus colour. This range acts as a critical filter, allowing the algorithm to efficiently locate and pinpoint areas exhibiting the predefined colouration, facilitating precise nucleus identification.

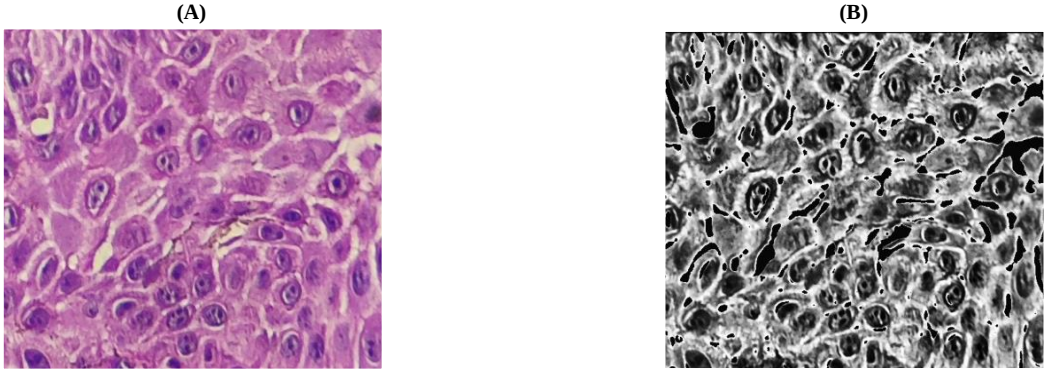

Fig. 1: A) H&E Stained. B) CLAHE Image.

7.1. Preprocessing of dataset

The processing of grading of the OPC starts with converting the input images to Reyhang normalised images, as illustrated in Fig 1 with respect to the H&E stained image, followed by the application of Contrast-Limited Adaptive Histogram Equalization (CLAHE) to distribute the bright areas evenly in the image. This foundational work sets the stage for dynamic contouring, a technique that more precisely delineates the varied morphologies of nuclei and cells.
By adopting dynamic contouring, we surpass the accuracy of traditional rectangular border detection techniques, specifically targeting and refining the issue of light region dominance. Dynamic contouring for cell and nucleus detection markedly improves the detection of the cell, nucleus and cytoplasm, as can be observed in Fig.2, by leveraging contouring techniques, as opposed to conventional methods, offering a significant boost in the accuracy of feature extraction.

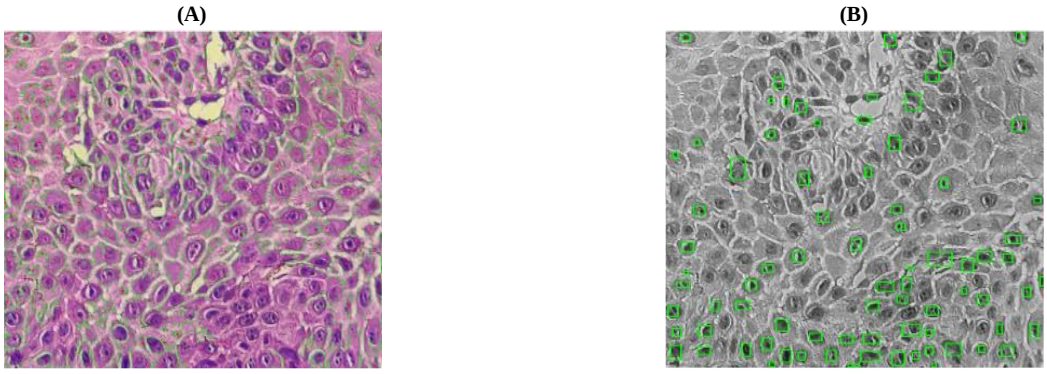

Fig. 2: a) Static Contouring. b) Dynamic Contouring.

7.2. Feature extraction and analysis from the processed dataset

A comprehensive analysis of each feature, including those that have been modified and improved upon from previous studies, underscores the innovative approach taken to boost diagnostic capabilities in the healthcare sector. The system extracts features individually are compiles them into a CSV file to train the ML algorithm used by HistogGrade model for grading oral epithelial dysplasia, as shown in Fig. 3. Methodologies used for determining each of the features for OPC detection and their prediction by the ML model are explained in subsequent sections.

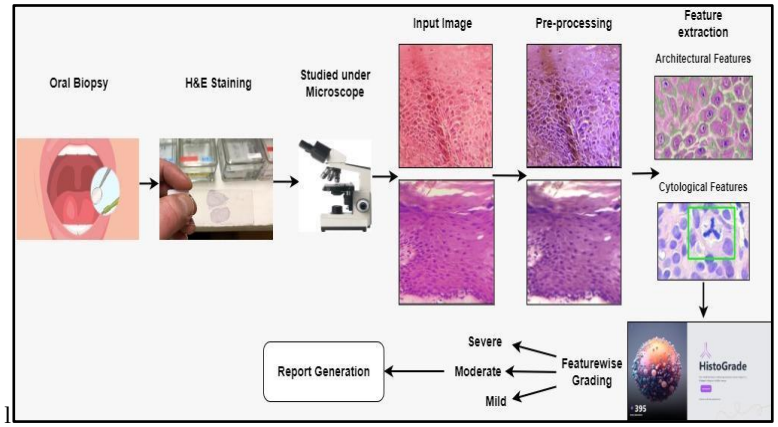

Fig. 3: Block Diagram of Histograde.

7.3. Increased number of nucleoli

An unusual rise in the number of nucleoli, crucial substructures found in the cell nucleus, suggests underlying cellular abnormalities or specific cellular activity. The steps involved in detecting an increased number of nucleoli within cells follow a cohesive and structured flow, beginning with the application of a morphological technique known as dilation, which enlarges the boundaries of cellular structures, facilitating the separation of individual cells within an image. Further sophistication in the analysis is achieved through cell contour analysis, designed to ascertain the presence of multiple nucleoli within a single cell. This analysis is vital for identifying cells with increased nucleoli, indicated by the detection of several nuclei within one cell.

7.4. Table 2 Analysis of the Increased number of nucleoli with Threshold = 80 sq. pixels

213	21	313.645	Severe

This sequential approach not only streamlines the process of detecting an increased number of nucleoli but also enriches the analysis with a depth of understanding and precision, crucial for comprehensive cellular analysis in histopathological examinations. The results can be observed in Fig. 4 and Table 2.

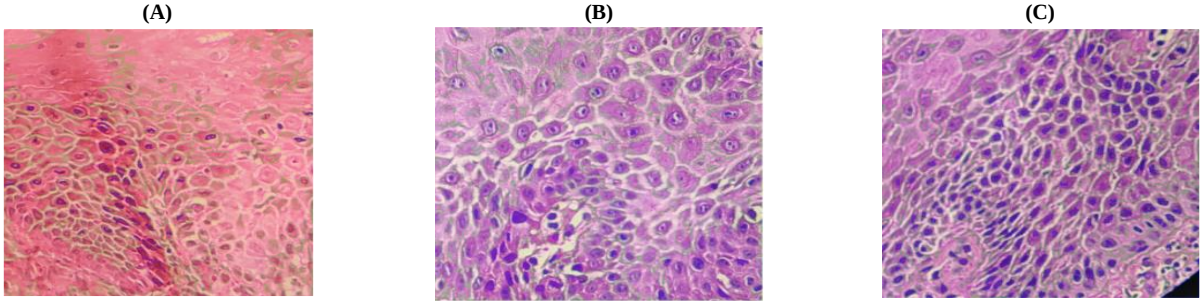

Fig. 4: a) Mild Category. b) Moderate Category. c) Severe Category.

7.5. Increased cell size

The above feature involves a detailed examination of cell nuclei within epithelial tissue, serving as a crucial parameter for cell classification. Reyhang normalisation is applied to the dataset to standardise colour variances and guarantee consistent analysis. After calculating the colour range standard deviation across photos, we normalise the images within these ranges. Average cell sizes for the image's lower, middle, and upper thirds are determined for the training set, yielding nine values per category (mild, moderate, severe). To test for dysplasia severity, picture portions are compared to these averages; disparities between the two indicate normal cell sizes. The grading can be visualised in Fig.5.

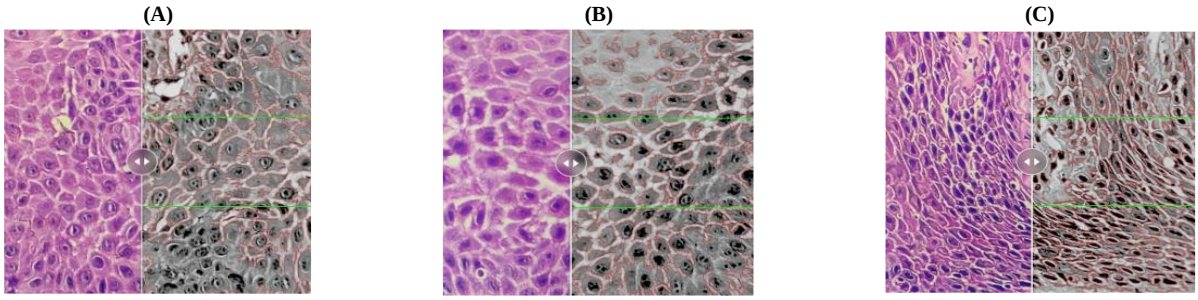

Fig. 5: A) Mild. B) Moderate. C) Severe.

7.6. Abnormal variation in cell shape

This feature highlights irregular cell shapes, often associated with precancerous conditions. In assessing dysplasia severity as in Fig. 6, sections of an image are evaluated against these average metrics; deviations from these standards indicate normal cell dimensions. This methodology, initially applied to cell size, is similarly utilised for assessing cell shape, ensuring a systematic approach to both aspects in the paper. This consistent methodological application allows for a nuanced analysis of dysplasia severity based on both cell size and shape, contributing to the robustness of the diagnostic process.

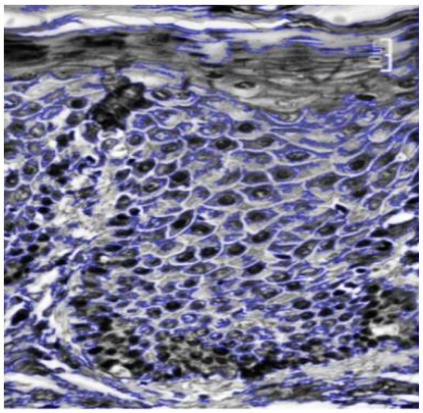

Fig. 6: Cell Shape.

7.7. Increased nucleus size

Here, the enlargement of the nucleus is considered a crucial characteristic for classifying dysplasia from Reyhang normalised images. This part retains its own language and viewpoint while reflecting the technical rigour used in cell size analysis. The segmentation of each image into thirds makes it easier to generate average nucleus sizes during training for categorisation purposes. Table 3 shows the dependency of grading on size and count of nuclei.

7.8. Table 3. Analysis of the increase in nucleus size

Severe	625.8018018018018 px	1602

7.9. Abnormal variation in nucleus shape

This highlights the significance of recognising atypical forms of cell nuclei as indicators of potential cellular dysfunction. This observation is critical because nuclei shape often reflects a cell's physiological state; deviations from the norm may suggest pathological changes or disease processes at play, as can be observed in the grading shown in Fig.7. Such variations in nuclear morphology can be pivotal in the early detection and diagnosis of various conditions, underscoring the intricate link between cellular structure and overall tissue health.

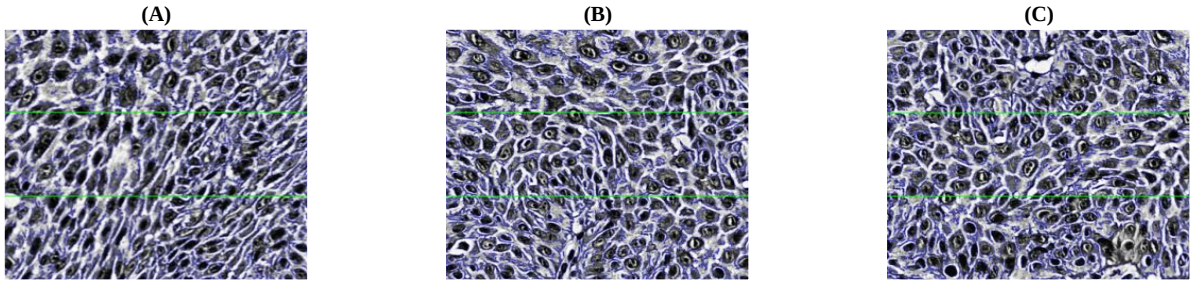

Fig. 7: A) Mild. B) Moderate. C) Severe.

7.10. Hyperchromasia

As seen in Fig. 8a, the cell nuclei's increased hematoxylin staining affinity results in a strong staining intensity. This phenomenon is particularly noteworthy in precancerous disorders like oral dysplasia, and it is symptomatic of various stages of malignancy as well. Hyperchromasia, which is typically evaluated by pathologists using a microscope to examine tissue samples, is a critical indicator of dysplastic alterations.The method automates the measurement of nuclear staining intensity, an essential indicator of hyperchromasia, through the utilisation of the HistomicSTK Python toolkit. This library facilitates an objective estimation of this vital characteristic by analysing both the size and intensity of nuclei, thereby enhancing the accuracy of dysplasia grading.

Severe	910.3422	2334.2113	0.39

7.13. Irregular epithelial stratification

This is a very important feature for identifying OPC and is understood by the disorganisation in the layers of epithelial cells, often seen in precancerous lesions.

The methodology to predict this feature is initiated by segmenting the images into equal parts to detect irregular epithelial stratification. Subsequently, the cells within these segments are contoured, and the centroid of each cell is identified. The centroids of cells that align

horizontally are then connected, creating multiple line segments across the image. The sparsity of the connections is assessed by analysing these line segments.
The segment exhibiting the highest number of sparse line segments is indicative of the degree of irregularity in epithelial stratification. The degree of sparsity within a particular segment, as shown in Fig. 9, is used to assign a grading of mild, moderate, or severe to epithelial dysplasia based on the extent to which this sparsity is pronounced.
The following features lack enough patient databases, because of which template matching is performed, as against ML techniques used in earlier features.

Fig.9: A) Mild. B) Severe.

The adjective mild means "not severe or strong," "gentle and kind," or "slight." Rough, aggressive, serious, stern, and unkind are adjectives that describe harshness.The basic idea behind epithelial dysplasia-based sparsity is to use the dysplasia's aberrant cell growth features to pinpoint regions with low-density cell populations, which may be a sign of an increased risk of developing or reoccurring cancer

7.14. Atypical mitotic figures

Anomalous cell divisions are identified here from images comprising samples of abnormal mitotic figures shown in Fig.10 and are detected with template matching. The diversity of the sample set ensures the approach's effectiveness in identifying various types of uncommon cell divisions, rendering the method valuable for pinpointing potential issues in cancer diagnosis.

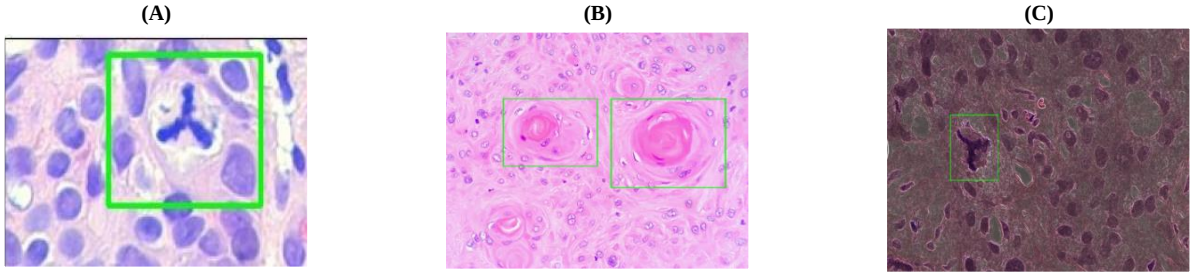

Fig. 10: A) Mitotic Figures. B) Apoptotic Mitoses. C) Premature Keratinization.

7.15. Apoptotic mitoses

In this research, aberrant apoptotic mitoses are detected through a technique known as template matching. The sample images from the web, as shown in Fig. 10 b, are used to train the computer to identify these unusual forms of cell death. The computer scrutinises the primary image against this dataset. Upon identifying a match, the area surrounding the recognised abnormal apoptotic mitosis in the main image is highlighted with rectangular contours.

7.16. Premature keratinization

This feature, as seen in Fig. 10c, is characterised by the early development of keratin layers in cells or cell groups before reaching the skin surface or the upper layers of the mucosa. It's a pathological finding, typically seen in various precancerous lesions and squamous cell carcinoma stages, highlighting disrupted epithelial maturation and differentiation processes. Both conditions are important markers for diagnosing and grading epithelial dysplasia and squamous cell carcinomas.

7.17. Single-cell keratinisation

Single-cell keratinisation refers to the process where individual cells within the epithelium undergo keratinisation, a differentiation that leads to the formation of keratin, a protective protein. This phenomenon is often observed in dysplastic tissue or early-stage squamous cell carcinomas, indicating abnormal cell behaviour and differentiation. Template matching was utilised to identify single-cell keratinisation, as seen in Fig. 11a, with templates specifically designed to recognise individual keratinised cells. The images undergo preprocessing to enhance cell outlines, followed by the application of template matching to detect cells that show keratin formation outside the typical stratification pattern.

(A) **(B)**

Fig. 11: A) Single Cell Keratinization. B) Keratin Pearls Within Rete Ridges.

7.18. Keratin pearls within rete ridges

Keratin pearls within rete ridges, as shown in Fig. 11b, are indicative of squamous cell carcinoma. They appear as concentric accumulations of keratin, a protein, within the epithelial cells, located specifically in the rete ridges—projections of the epidermis into the dermis. These structures are key diagnostic features, suggesting high cell differentiation and malignancy in the tissue examined. To detect keratin pearls within rete ridges, we employed a template matching technique, a method wherein predefined templates of keratin pearls are matched against the histopathological images. This process begins by segmenting the images to focus on regions containing rete ridges. The algorithm then searches for patterns in these areas that closely resemble the keratin pearl templates.

8. Results

For each image, the proposed grading system HistoGrade classifies normal from abnormal cells using individual features, as can be observed from Fig. 12a, and then grading is done on the abnormal cells. A user-friendly GUI is created for this purpose, as shown in Fig. 12 b. and the methodology applied in this research. The system achieves an average accuracy of 94% using a majority voting system on the results obtained using each feature, and individual feature selection can be observed in Fig. 12c. In this system, the grading output that appears most frequently across different features is determined to be the result within a single image. Ultimately, a comprehensive report as seen in Fig.12d. is generated, summarising the grades of all individual features for convenient review and analysis for the doctors as well as the patients.

image_name	Total Nuclei	Lower Size	Middle Size	Upper Size	Mitotic figures	Keratin pearls	lower_third_co	middle_third_cc	upper_third_co	Intensities	class
03.jpg	518	79.83459	499.7551	5701.586	0	0	1	6	3	178.523345	0
11.jpg	303	83.55729	458.4231	5949.604	0	0	7	7	6	171.4694805	0
12.jpg	257	75.45808	518.0588	5894.288	0	0	5	0	0	123.5093985	0
07.jpg	230	111.1731	464.3816	7133.581	0	0	4	7	8	155.3415337	0
05.jpg	339	70.76087	494.1304	4416.633	0	0	2	5	6	110.6996889	0
01.jpg	324	100.8531	501.5965	3526.299	0	0	8	6	7	200.5449901	0
09.jpg	260	115.8045	493.2097	2969.462	1	0	5	8	3	187.6390467	0
02.jpg	473	52.9322	552.75	4068.072	0	0	7	4	8	141.5630703	0
06.jpg	422	89.79817	516.7768	3126.26	0	0	6	0	0	125.4024658	0
04.jpg	284	90.25146	501.9556	5540.985	0	0	9	4	9	160.599575	0
08.jpg	204	112.1792	487.85	5995.147	0	0	7	0	0	193.0254459	0
10.jpg	240	85.55426	513.2121	5722.077	0	0	6	0	0	177.446434	0
26.jpg	270	75.71111	538.76	4097.518	0	0	0	8	1	197.6349506	1
27.jpg	208	101.4813	481.275	5862.068	0	0	8	5	0	258.5070181	1
1.jpg	344	76.34332	487.6719	4586.016	0	0	5	0	7	298.6959352	1
11.jpg	229	55.59146	469.6875	11336.52	0	0	4	8	1	320.6071978	1
14.jpg	361	97.27005	524.7174	3825.598	0	0	4	2	2	309.3146963	1
3.jpg	427	70.58022	511.375	3541.419	0	0	5	8	0	196.9851608	1
21.jpg	234	75.65203	500.2632	11840.51	0	0	5	4	0	261.0182114	1
19.jpg	538	59.09345	465.4464	4080.092	0	0	6	9	2	419.1728001	1
24.jpg	318	80.33742	524.0244	3471.478	0	0	2	9	6	379.506829	1
5.jpg	466	79.2	497.37	3198.678	0	0	9	7	5	228.412629	1
12.jpg	279	85.78662	498.0667	5416.886	0	0	2	2	3	244.187599	1
13.jpg	279	88.52685	506.6458	4833.509	0	0	0	0	0	279.506829	1
8.jpg	393	69.46828	494.1731	4407.27	0	0	5	8	5	411.5630703	1
16.jpg	356	81.95477	500.5577	3653.179	0	1	9	4	0	225.4024658	1

Fig. 12: A) Features Extracted from the Images.

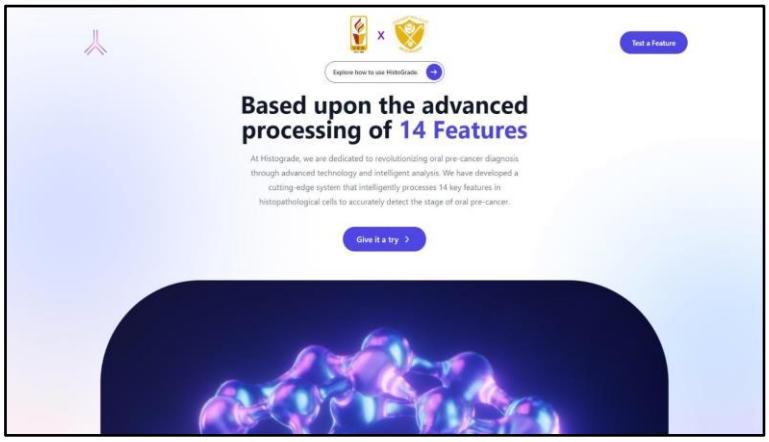

Fig. 12: B) GUI Depicting Various Features & Their Grading Results.

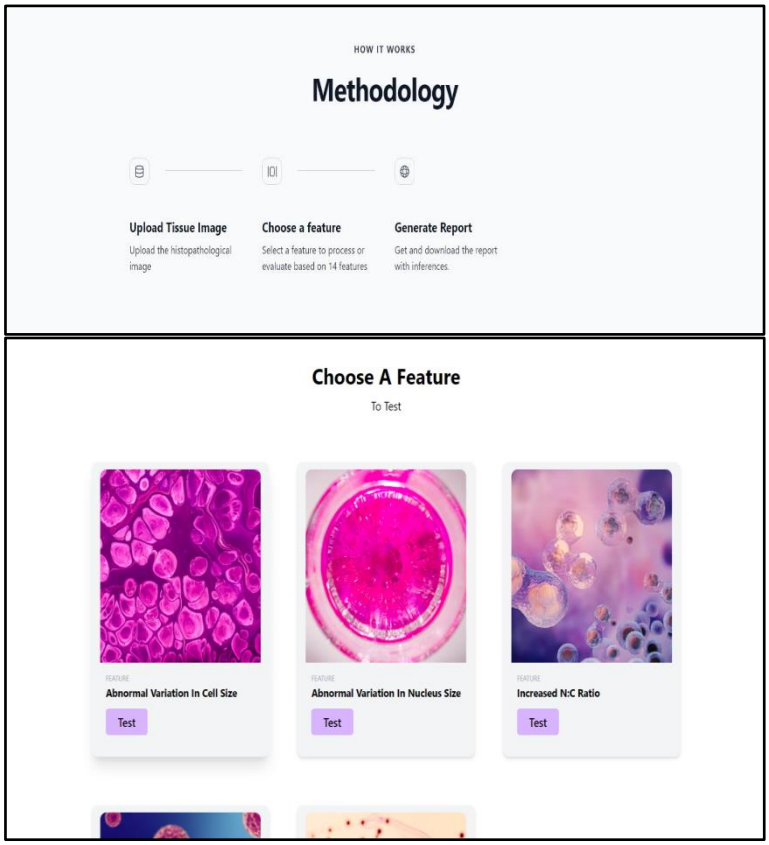

Fig. 12: C) GUI Depicting Various Features & Their Grading Results & User Interface of Histograde.

8.1. User interface of histograde -feature selection page

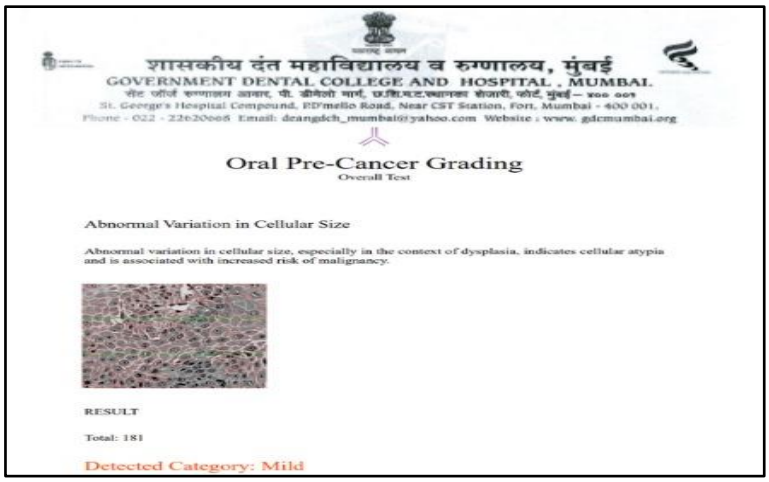

Fig. 12: D) Report of Grading Generated Through Histograde.

The regressive H&E staining technique is associated with the procedural staining phase known as differentiation. The most often used ingredients in differentiating solutions for hematoxylin staining are hydrochloric acid (for quick differentiation) and acetic acid (for slower, more controlled differentiation). The majority of the more than clinical AI applications that the FDA has approved are related to radiology. The Food and Drug Administration has approved dozens of new items that can be controlled by artificial intelligence

9. OSCC

The OSCC mouth Squamous Cell Carcinoma, the most prevalent kind of mouth cancer, is referred to as "OSCC" throughout. The squamous epithelial cells that line the oral cavity are the site of this malignant tumor, which can spread to the floor of the mouth, gums, lips, and tongue. With a high death rate and rising frequency in some areas, OSCC is a serious worldwide health concern. The identification of oral squamous cell carcinoma from histopathologic images using EfficientNetB3. more than 90% of instances of oral cancer worldwide are of the extremely aggressive and common kind known as oral squamous cell carcinoma (OSCC). Early OSCC discovery is essential since it increases patient survival rates dramatically and offers a five-year survival probability of almost 80%. On the other hand, conventional diagnostic techniques based on histological examination are labor-intensive, prone to human error, and require highly skilled specialists. By investigating the use of deep learning to automatically detect OSCC from histopathologic images, this work tackles these problems. These results imply that the suggested approach provides a workable alternative to early, automated OSCC identification, enhancing clinical diagnostic precision and patient outcomes. AI-powered diagnostics have remarkable speed and accuracy when

analyzing thousands of imaging scans, including mris, mammograms, and X-rays. For example, long before symptoms appear, AI technologies can identify tiny calcium deposits in breast tissue, which frequently indicate early breast cancer.

10. Conclusion

This study highlights the possibility for better outcomes through earlier intervention, underscoring the vital necessity of early identification in the management of oral precancers. With the help of this study, pathologists may now more accurately identify important characteristics linked to oral dysplasia. It provides a trustworthy way to confirm the grading of examinations that are typically completed by hand inspection. Because this tool is accessible through a web platform, pathologists worldwide can easily use it to validate their diagnostic results. The system promises better outcomes with more training examples under each feature and also better performance measures if a balanced dataset is available, consisting of enough test cases in each category of grades for analysis of dysplasia.It could also serve as a foundation for future AI-driven diagnostic tools that might enable early detection, improved prognostication, and support for pathologists in low-resource settings.

Acknowledgements

We would like to thank Dean Dr Wasundhara Bhad, Dr Tabita Joy, and Dr. Manish Ahire Sardar for their invaluable contribution and support in making this joint collaborative scientific research on oral pre-cancer detection and analysis.
This research paper is a live testimony to the joint efforts of practising doctors and budding engineers of the Government Dental College & Hospital, Mumbai, and the Vivekanand Education Society's Institute of Technology, Chembur, Maharashtra.
Author Contributions: Conceptualization, Dr Sharmila Sengupta, Anuj Bagad; methodology, Aayush Talreja; software, Aayush Talreja and Anuj Bagad; validation, Dr Sharmila Sengupta, Dr Harsha Karwa, and Dr Shrijha G; formal analysis, Aayush Talreja; investigation, Mansi Bellani; resources, Dr Harsha Karwa, and Dr Shrijha G; data curation, Aayush Talreja and Anuj Bagad; writing—original draft preparation, Mansi Bellani; writing—review and editing, Mansi Bellani; visualisation, supervision, Dr Sharmila Sengupta; All authors have read and agreed to the published version of the manuscript.

References

[1] Gupta, R. K., Manhas, J., & Kour, M. (2022). Hybrid Feature Extraction Based Ensemble Classification Model to Diagnose Oral Carcinoma Using Histopathological Images. Journal of Scientific Research, 66(3), 219-226. Banaras Hindu University. https://doi.org/10.37398/JSR.2022.660327.
[2] Panigrahi, S., Nanda, B.S., Bhuyan et al. (2023). Classifying histopathological images of oral squamous cell carcinoma using deep transfer learning. Heliyon, 9(2023), e13444. https://doi.org/10.1016/j.heliyon.2023.e13444.
[3] Rahman AU, Alqahtani A, Aldhafferi N, Nasir MU, Khan MF, Khan MA, Mosavi A. "Histopathologic Oral Cancer Prediction Using Oral Squamous Cell Carcinoma Biopsy Empowered with Transfer Learning." Sensors (Basel). 2022 May 18;22(10):3833. doi: 10.3390/s22103833. PMID: 35632242; PMCID: PMC9146317. https://doi.org/10.3390/s22103833.
[4] R. A. Welikala et al., "Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer," in IEEE Access, vol. 8, pp. 132677-132693, 2020, https://doi.org/10.1109/ACCESS.2020.3010180.
[5] B R, Nanditha & Annegowda, Geetha. (2022). "Oral Cancer Detection using Machine Learning and Deep Learning Techniques." International Journal of Current Research and Review. 14. 64-70. https://doi.org/10.31782/IJCRR.2021.14104.
[6] Bashir, S., Mahmood, H., Shaban, M., et al. (2020). "Automated grade classification of oral epithelial dysplasia using morphometric analysis of histology images." in H.K. Hahn & M.A. Mazurowski (Eds.), Medical Imaging 2020: Digital Pathology (Vol. 11320, pp. page range). Society of Photo-Optical Instrumentation Engineers (SPIE). https://doi.org/10.1117/12.2549705.
[7] Aldhebaib, A., Khanagar, S. B., Alkadi, L., Alghilan, M. A., Kalagi, S., Awawdeh, M., Bijai, L. K., Vishwanathaiah, S., & Singh, O. G. (2023). "Application and Performance of Artificial Intelligence (AI) in Oral Cancer Diagnosis and Prediction Using Histopathological Images: A Systematic Review." Biomedicines, 11(6), 1612. MDPI. https://doi.org/10.3390/biomedicines11061612.
[8] Kritsasith, W., & Suebnukarn, S. (2024). "Deep Learning in Oral Cancer: A Systematic Review. "BMC Oral Health, 24, Article 39. BioMed Central. https://doi.org/10.1186/s12903-024-03993-5.
[9] Dixit, S., Kumar, A., & Srinivasan, K. (2023). "A Current Review of Machine Learning and Deep Learning Models in Oral Cancer Diagnosis: Recent Technologies, Open Challenges, and Future Research Directions. Diagnostics", 13(7), 1353. MDPI. https://doi.org/10.3390/diagnostics13071353..
[10] Vanesa Pereira-Prado et.al. "Artificial Intelligence for Image Analysis in Oral Squamous Cell Carcinoma: A Review", Diagnostics (Basel) 2023 Jul 20;13(14):2416. https://doi.org/10.3390/diagnostics13142416.
[11] Chaudhary, N., Rai, A., Rao, A.M. et al. "High-resolution AI image dataset for diagnosing oral submucous fibrosis and squamous cell carcinoma." Sci Data 11, 1050 (2024). https://doi.org/10.1038/s41597-024-03836-6.
[12] Jiakuan Peng et.al., "Oral epithelial dysplasia detection and grading in oral leukoplakia using deep learning", BMC Oral Health, 2024 Apr 9;24:434. https://doi.org/10.1186/s12903-024-04191-z.
[13] Simran Uppal et.al., "Machine learning methods in predicting the risk of malignant transformation of oral potentially malignant disorders: A systematic review", International Journal of Medical Informatics, Volume 186, June 2024, 105421 https://doi.org/10.1016/j.ijmedinf.2024.105421.
[14] Rachelle Wolk, et.al. "Proceedings of the North American Society of Head and Neck Pathology Companion Meeting, New Orleans, LA, March 12, 2023: Oral Cavity Dysplasia: Why Does Histologic Grading Continue to be Contentious?", Head Neck Pathol ,2023 May 15;17(2):292–298. https://doi.org/10.1007/s12105-023-01544-y.
[15] John Adeoye et.al., "Deep Learning Predicts the Malignant-Transformation-Free Survival of Oral Potentially Malignant Disorders" , Cancers (Basel) 2021 Dec 1;13(23):6054. https://doi.org/10.3390/cancers13236054.
[16] Amann, J., Blasimme, A., Vayena, E. et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 20, 310 (2020). https://doi.org/10.1186/s12911-020-01332-6.