Early-Stage Oral Cancer Identification Using White Light Imaging and LightGBM Algorithm

AkilarasanB[1], DineshL S[2], Sakthivel S[3], Revathi P[4]
[1, 2, 3] Dept of Computer Science & Engineering
[4]Assistant Professor, Dept of Computer Science & Engineering
[1, 2, 3, 4] Chettinad College of Engineering & Technology, Karur

Abstract- *Oral cancer is a significant public health concern globally, with early detection playing a crucial role in improving patient outcomes. In this project, we propose a novel approach for oral cancer detection leveraging deep learning techniques. The system is developed using Python as the primary coding language, with Flask serving as the web framework and HTML, CSS, and JavaScript for the frontend interface. The dataset utilized in this project comprises 1013 oral cancer images and 294 noncancer oral images, meticulously labelled for easy classification. This dataset serves as a comprehensive resource for researchers and developers in the field of oral cancer detection using machine learning algorithms. With a balanced representation of cancerous and non-cancerous samples, this dataset facilitates the exploration of innovative approaches to enhance diagnostic accuracy. The proposed model employs the LightGBM algorithm to efficiently classify oral cancer stages based on white light images, ensuring a robust and accurate detection system.*

Keywords- Oral cancer detection, Python-based system, medical image classification, Cancerous vs. non-cancerous samples, LightGBM algorithm, Feature extraction and classification.

I. INTRODUCTION

Oral cancer is one of the most prevalent malignancies worldwide, posing a significant challenge to global healthcare systems. Early detection plays a critical role in improving survival rates, as timely intervention can greatly enhance treatment effectiveness. However, conventional diagnostic methods often rely on subjective clinical evaluations, biopsies, and histopathological analyses, which can be time-consuming and resource intensive. To address these challenges, artificial intelligence (AI) and machine learning (ML) have emerged as promising tools for automating and improving the accuracy of cancer detection. In this project, we propose a deep learning-based approach for oral cancer detection using machine learning algorithms. The system is built using Python as the primary programming language, with Flask as the web framework, and HTML, CSS, and JavaScript for the frontend interface.A well-structured dataset comprising 1,013 oral cancer images and 294 non-cancerous oral images is utilized, ensuring a balanced representation for training and evaluation. By integrating machine learning techniques, particularly the LightGBM algorithm, the model effectively classifies oral cancer stages based on white light images.

II. LITERATURE SURVEY

A. A. Risk Factors:

Tobacco use, including cigarettes, smokeless tobacco, and betel quid chewing, is the primary cause of oral squamous cell carcinoma (OSCC). Heavy alcohol consumption also significantly increases the risk, as it damages the oral mucosa and makes it more susceptible to carcinogens.

B. B. Viral Infections:

Human papillomavirus (HPV), particularly type 16, has been identified as a leading cause of oropharyngeal cancer, with increasing prevalence among younger, non-smoking populations.

C. C. Genetic and Environmental Factors:

Genetic mutations and prolonged exposure to harmful environmental agents, such as ultraviolet (UV) radiation and industrial pollutants, also contribute to the disease's development.

D. Diagnosis and Screening:

Early diagnosis significantly improves survival rates, yet most cases are detected at an advanced stage due to a lack of symptoms in early phases. Common diagnostic methods include:

E. Visual and Physical Examination:

The initial step in diagnosing oral cancer involves a thorough visual and physical examination by dentists, oncologists, or oral healthcare providers. This examination helps in identifying suspicious lesions, ulcers, or abnormal tissue growth in the oral cavity.

F. Biopsy and Histopathology:

A biopsy is the gold standard for diagnosing oral cancer. It involves the removal of tissue from a suspected lesion for microscopic examination to confirm malignancy.

G. Imaging Techniques:

Once a biopsy confirms oral cancer, imaging techniques are used to determine the extent of the tumor, assess metastasis, and aid in treatment planning.

H. Emerging Technologies:

Machine learning models, such as LightGBM, have been developed for automated classification of oral cancer stages based on white light images.

III. EXISTING SYSTEM

Oral cancer diagnosis traditionally relies on clinical examinations, biopsies, and histopathological analysis. While these methods are highly accurate, they come with several limitations:

A. Manual Diagnosis:

Doctors visually inspect lesions, ulcers, and abnormalities in the oral cavity. Diagnosis depends on the experience and judgment of medical professionals, which can be subjective.This can lead to inconsistencies, especially in early-stage cancer detection.

B. Biopsy Requirement:

A tissue sample is taken from the suspected area and sent for pathological analysis. This process is invasive, causing discomfort to the patient. The procedure is costly and requires specialized medical expertise.

C. Delayed Results:

Histopathological tests take time (several days to weeks) before confirming cancer. This delay can impact treatment initiation, leading to the progression of cancer.

D. Limited Accessibility:

Rural and underdeveloped regions lack access to expert oncologists and diagnostic centers. Patients may need to travel long distances, making early diagnosis difficult.

E. Lack of Automated Detection:

Conventional methods do not use AI-driven tools for automated screening.

Large-scale screening programs become challenging due to a lack of automation.

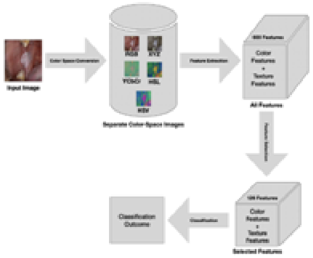
IV. PROPOSED SYSTEM

A. Machine Learning-Based Detection:

The proposed system leverages machine learning techniques to enhance the accuracy and efficiency of oral cancer detection. Traditional diagnostic approaches rely on subjective visual examinations and biopsies, which can lead to delays and inconsistencies in diagnosis. By utilizing the Light Gradient Boosting Machine (LightGBM), the system analyses white light images and classifies oral lesions into benign, pre-cancerous, and malignant stages. This method eliminates human subjectivity and ensures a standardized, data-driven approach to diagnosis. The ML model is trained on a dataset of oral lesion images, allowing it to identify patterns and abnormalities with high precision. Unlike manual assessment, which depends on the experience of the examiner, the machine learning model can process large volumes of data efficiently, leading to faster and more reliable detection.

B. Image Processing for Feature Extraction:

To enhance classification accuracy, the system integrates image processing techniques that refine and extract essential features from input images. Various factors, such as lighting conditions, noise, and image quality, can affect the performance of machine learning models. The system applies contrast adjustment, noise reduction, and segmentation to enhance image clarity and isolate regions of interest. The extracted features, including color intensity and texture patterns, are then analyzed to differentiate between normal and abnormal tissues. Automating the feature extraction process minimizes human error and ensures that only relevant image attributes are used for classification. By improving the quality of input data, the system enhances the reliability and accuracy of oral cancer detection.



C. Mobile and Web-Based Application:

The system is designed as a mobile and web-based application to make oral cancer detection accessible to a broader population. Patients and healthcare professionals can upload images of oral lesions for real-time analysis, eliminating the need for in-person consultations in the early screening stages. The machine learning model processes the uploaded images and provides instant classification results, indicating whether further medical evaluation is required. This application is particularly useful for individuals in remote areas with limited access to specialized healthcare. It also enables early screening, reducing the burden on hospitals and clinics by identifying high-risk cases that require immediate attention. Cloud-based data storage ensures secure record-keeping and allows healthcare providers to track lesion progression over time, facilitating better patient management.

D. Improved Early Detection and Patient Outcomes:

Early detection of oral cancer significantly improves treatment success rates and patient survival. Many cases are diagnosed at advanced stages, leading to complex treatment procedures and higher mortality rates. The proposed system provides a solution by enabling the early identification of pre-cancerous lesions, allowing for timely medical intervention. Faster and more accurate diagnosis helps reduce treatment costs by minimizing the need for aggressive therapies such as chemotherapy and radiation. By integrating continuous improvements in machine learning models and expanding the system's dataset, detection accuracy can be further enhanced over time. This system has the potential to evolve beyond oral cancer detection and be applied to other oral diseases, further improving public health outcomes.

V. SYSTEM REQUIREMENTS

A. Hardware Requirements

The system requires a stable hardware configuration to ensure smooth execution of the machine learning model, image processing tasks, and web-based application. The following are the minimum hardware requirements:

- **System:** A personal computer or server capable of handling machine learning computations and web application hosting.
- **Processor:** Intel Pentium i3 Processor or higher for efficient data processing.
- **Hard Disk:** A minimum of 500 GB storage to accommodate datasets, model files, and user data.
- **Monitor:** A 15'' LED display for clear visualization of images and user interface.
- **Input Devices:** Standard keyboard and mouse for user interaction.
- **RAM:** At least 8 GB to support machine learning computations and real-time image processing.

B. Software Requirements:

The software stack consists of programming languages, frameworks, and an operating system to develop and deploy the proposed system.

- **Operating System:** Windows 10 / 11, ensuring compatibility with required libraries and frameworks.
- **Coding Language:** Python 3.10.9 for developing the machine learning model, image processing pipeline, and backend services.
- **Web Framework:** Flask, a lightweight framework for building the backend and handling API requests.
- **Frontend:** HTML, CSS, JavaScript for designing the user interface, enabling seamless interaction between users and the system.

VI. FUTURE WORK

The proposed system provides a machine learning-based solution for oral cancer detection, but there is still significant scope for improvement and expansion. Future work will focus on enhancing the accuracy, scalability, and real-world applicability of the system while integrating additional advanced technologies.Finally, the system can be extended to detect other oral diseases, such as periodontal diseases, bacterial infections, and fungal conditions. By broadening its scope beyond cancer, the system can become a versatile diagnostic tool for oral healthcare, assisting dentists and clinicians in early disease detection and prevention.

VII. CONCLUSION

The proposed system uses machine learning to detect and classify oral cancer early, overcoming the drawbacks of traditional methods that depend on manual examination and biopsies. By using Light Gradient Boosting Machine (LightGBM) and image processing techniques like contrast adjustment, noise reduction, and segmentation, the system improves accuracy and efficiency in identifying cancerous lesions. A mobile and web-based application allows patients and doctors to upload and analyze images in real time, making early detection possible. This system provides a reliable and automated way to detect oral cancer, reducing human error and making healthcare more efficient. With its fast, cost-effective, and precise classification, it has the potential to improve oral cancer screening, leading to better patient care and wider access to healthcare services

REFERENCES

[1] Goswami, B., Bhuyan, M. K., Alfarhood, S., & Safran, M. (2024). Classification of Oral Cancer IntoPre Cancerous Stages From White Light Images Using LightGBM Algorithm. IEEE Access. DOI: 10.1109/ACCESS.2024.3370157

[2] Abati, S., Bramati, C., Bondi, S., Lissoni, A., & Trimarchi, M. (2020). Oral Cancer and Precancer: A Narrative Review on the Relevance of Early Diagnosis. International Journal of Environmental Research and Public Health, 17(24), 9160. DOI: 10.3390/ijerph17249160

[3] Saberian, E., Jenča, A., Petrášová, A., Jenčová, J., Atazadegan Jahromi, R., &Seiffadini, R. (2023). Oral Cancer at a Glance. Asian Pacific Journal of Cancer Biology, 8(4), 379-386. DOI: 10.31557/APJCB.2023.8.4.379

[4] Taheri, J. B., Namazi, Z., Azimi, S., Mehdipour, M., Behrovan, R., & Far, K. R. (2018). Knowledge of Oral Precancerous Lesions Considering Years Since Graduation Among Dentists in the Capital City of Iran: a Pathway to Early Oral Cancer Diagnosis and Referral? Asian Pacific Journal of Cancer Prevention, 19(8), 2103-2108. https://doi.org/10.22034/APJCP.2018.19.8.2103

[5] Borse, V., Konwar, A. N., & Buragohain, P. (2020). Oral cancer diagnosis and perspectives in India. Sensors International, 1,100046. https://doi.org/10.1016/j.sintl.2020.100046

[6] Musulin, J., Štifanić, D., Zulijani, A., Ćabov, T., Dekanić, A., & Car, Z. (2021). An Enhanced Histopathology Analysis: An AI-Based System for Multiclass Grading of Oral Squamous Cell Carcinoma and Segmenting of Epithelial and Stromal Tissue. Cancers, 13(1784). [DOI: 10.3390/cancers13081784].

[7] Tanriver, G., SolukTekkesin, M., & Ergen, O. (2021). Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders. Cancers, 13(2766). [DOI: 10.3390/cancers13112766].

[8] Imbesi Bellantoni, M., Picciolo, G., Pirrotta, I., Irrera, N., Vaccaro, M., Vaccaro, F., Squadrito, F., &Pallio, G.(2023). Oral Cavity Squamous Cell Carcinoma: An Update of the Pharmacological Treatment. Biomedicines, 11(1112). [DOI: 10.3390/biomedicines11041112].

[9] Pan, X., Yang, H., Fan, Y., Zhang, L., et al. (2020). *Multi-task Deep Learning for Fine-Grained Classification and Grading in Breast Cancer Histopathological Images.* Studies in Computational Intelligence. [DOI: 10.1007/978-3-030-04946-1_10].

[10] Qayyum, H., Majid, M., Anwar, S. M., & Khan, B. (2017). *Facial Expression Recognition Using Stationary Wavelet Transform Features.* Mathematical Problems in Engineering, Article ID 9854050. [DOI: 10.1155/2017/9854050