Mrs. G. SANGEETHA LAKSHMI., 2025, 13:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Oral Cancer Detection Using Deep Learning

Mrs. G. SANGEETHA LAKSHMI., Mrs. S. Hemalatha

M.Sc.(IT)., M.Phil., Assistant Professor and Head Department of Computer Applications , D.K.M. College for Women (Autonomous), Vellore

Abstract- Early and precise detection of oral cancer is critical for improving patient outcomes, yet conventional diagnostic methods often involve manual analysis, which can be slow and susceptible to human error. To overcome these limitations, this research introduces an automated detection system that combines deep learning for feature extraction with the Random Forest algorithm for classification. By analyzing medical images, the deep learning component identifies essential features such as texture, color inconsistencies, and irregular tissue structures. These features are then processed by the Random Forest classifier, which utilizes an ensemble of decision trees to enhance classification accuracy and minimize errors. Trained on a dedicated dataset of oral cancer images, the model effectively differentiates between malignant and benign tissues. Experimental findings reveal that this hybrid approach outperforms standard machine learning techniques, offering a faster and more dependable diagnostic tool to aid clinicians in early oral cancer detection and improve patient survival rates.

Keywords- Oral Cancer Detection ,Deep Learning ,Feature Extraction ,Random Forest Classifier ,Medical Image Analysis ,Automated Diagnosis ,Malignant vs. Benign Classification ,Ensemble Learning ,Diagnostic Accuracy ,Early Cancer Detection.

I. INTRODUCTION

Oral cancer ranks among the most widespread and deadly diseases globally, often leading to high fatality rates due to delays in diagnosis. It commonly targets areas such as the lips, tongue, inner cheeks, and throat, with major contributing factors including smoking, alcohol intake, and inadequate oral hygiene. Detecting the disease at an early stage is essential for improving survival rates, as it facilitates prompt medical intervention and more effective treatment strategies. However, conventional diagnostic practices still rely heavily on visual inspection and clinical judgment, which can be inefficient, subjective, and prone to error.

With the rapid evolution of artificial intelligence (AI), particularly in deep learning, medical diagnostics have seen significant improvements in accuracy and automation. Deep learning algorithms are capable of recognizing subtle patterns in medical imagery that may be overlooked by human evaluators. Among these, the Random Forest algorithm has gained recognition for its strength in managing complex datasets and producing dependable classification results. This study introduces a hybrid AI model that utilizes deep learning to extract important image-based features—such as texture, color differences, and structural irregularities—and applies the Random Forest classifier to accurately categorize tissues. The integrated system aims to improve diagnostic precision, lower the rate of false positives, and reduce reliance on manual evaluation, offering a practical solution for early oral cancer detection and improved patient prognosis.

II . LITERATURE SURVEY

Title: Key Statistics on Oral and Oropharyngeal Cancer

- **Author:** American Cancer Society
- **Year:** 2020
- **Summary:** This report highlights critical data on the occurrence, mortality, and survival trends associated with oral and oropharyngeal cancers. It analyzes demographic patterns and risk

profiles to illustrate the global and regional burden of these diseases.

Title:Contributing Factors to Oral and Oropharyngeal Cancer Development

- **Author:** American Cancer Society
- **Year**: 2020
- **Summary:** The study investigates primary risk elements like tobacco, alcohol, HPV infection, and inherited traits that influence the onset of oral cancers. It underscores the role of early detection and lifestyle changes in reducing cancer risks.

Title: Deep Learning Overview and Its Expanding Role in Machine Learning

- **Author:** Dargan et al.
- **Year:** 2020
- **Summary:** This comprehensive review explores how deep learning, including models like CNNs, is revolutionizing various fields, especially healthcare. It emphasizes the technology's capability in image-based diagnostics, making it vital for detecting diseases such as cancer.

Title: Machine Learning Techniques for Identifying Oral Cancer

- **Author:** Prabhakaran & Mohana
- **Year:** 2020
- **Summary:** This research assesses and compares classification models like SVM, Decision Trees, and Random Forest for diagnosing oral cancer. It focuses on the strengths and accuracies of each method in predicting cancerous conditions.

Title: SVM-Based Techniques in Cancer Genomics Analysis

- **Author:** Shujun et al.
- **Year:** 2018
- **Summary:** The paper examines the use of Support Vector Machines (SVM) in genomics for identifying cancer-specific mutations. It presents case studies highlighting SVM's utility in genetic evaluation and oral cancer detection.

III . PROPOSED SYSTEM

The developed system employs a deep learning framework, incorporating the Random Forest algorithm to analyze medical images for oral cancer detection. It combines sophisticated image enhancement and feature extraction techniques to accurately identify malignant tissues. Utilizing a curated set of histopathological images, the data undergoes preprocessing to improve visual clarity and emphasize critical features. Convolutional Neural Networks (CNNs) are used to extract key visual patterns, which are then classified using the Random Forest method to ensure dependable and precise diagnosis. This automated approach supports healthcare providers in identifying cancer at an early stage, minimizing diagnostic errors and contributing to better treatment outcomes.

IV . MODULES

Data Collection & Preprocessing Module

This module is tasked with gathering and preparing the image dataset required for model development and evaluation. The dataset consists of oral cancer images sourced from hospitals, research institutions, and online medical repositories. Since raw images can include noise, artifacts, or lighting inconsistencies, several preprocessing techniques are employed to enhance image quality. These include resizing and normalizing the images for consistency, applying filters like Gaussian or Median to reduce noise, and using data augmentation methods such as rotation, flipping, and zooming to expand the dataset and improve model generalization. The preprocessed images are then forwarded to the next module for feature extraction.

Feature Extraction Module

In this stage, a Convolutional Neural Network (CNN) is utilized to extract deep and meaningful features from the processed images. These features—based on texture, shape, and color are vital for differentiating between healthy and cancer-affected tissues. The CNN applies convolutional layers to detect visual patterns, pooling layers to minimize spatial dimensions, and activation functions like

ReLU to introduce non-linearity. Ultimately, the CNN converts each image into a feature vector, which represents the extracted information in a numerical format. This output is then sent to the classification module.

Feature Selection & Classification Module
This module takes the extracted features and applies the Random Forest algorithm to categorize the images as cancerous or non-cancerous. To improve efficiency and accuracy, feature selection methods are used to discard irrelevant or redundant attributes, ensuring only the most impactful features are retained. The Random Forest model then constructs multiple decision trees and aggregates their results to deliver a final classification. This module handles model training using labeled data, enabling it to recognize patterns indicative of malignancy. It also manages prediction tasks for new, unseen images.

Performance Evaluation Module
Following classification, the system's performance is assessed through several standard evaluation metrics. This module calculates the accuracy to determine the proportion of correct predictions. Additionally, precision and recall metrics help evaluate the model's effectiveness in identifying true cancer cases without generating excessive false results. The Fl score offers a balanced measure that combines precision and recall, while the confusion matrix provides a visual overview of the classification results, including true positives, false positives, true negatives, and false negatives. Insights from this module are used to refine the model and support result display in the user interface.

User Interface & Report Generation Module
This module serves as the front-end interface for users such as doctors and researchers. It enables them to upload medical images for diagnostic analysis via a web or mobile platform. After processing, the system displays classification results identifying the image as cancerous or non-cancerous along with aconfidence score. It also generates comprehensive diagnostic reports for medical review and decision-making. Additionally, the module maintains a database to store images and their corresponding classification results, ensuring data can be retrieved for future reference or analysis.

V. CONCLUSION

The Random Forest algorithm stands out as an effective technique for detecting oral cancer, offering greater accuracy, reliability, and resistance to overfitting compared to conventional machine learning models. By combining multiple decision trees, it achieves robust classification results, even when dealing with complex, high-dimensional datasets or missing values. In the context of medical imaging, Random Forest analyzes critical features—such as shape, color, and texture—to distinguish between healthy and malignant tissues with reduced error rates. Its capability to rank feature importance not only improves model performance but also provides valuable insights into the factors most indicative of cancer, supporting medical experts in their diagnostic process. Although computationally intensive, its scalability and precision make it a strong candidate for automated cancer screening. Integrating this method with deep learning approaches holds potential to further enhance diagnostic accuracy and effectiveness in the future.

VI . RESULT

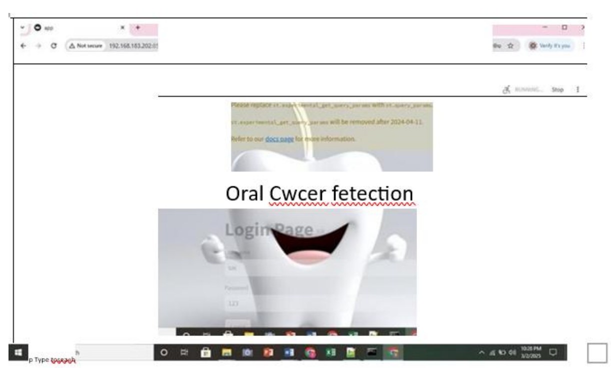

Figure 1: The create for login page

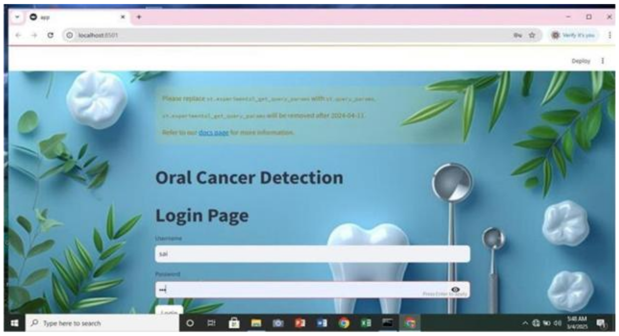

Figure 2: The use for username and password

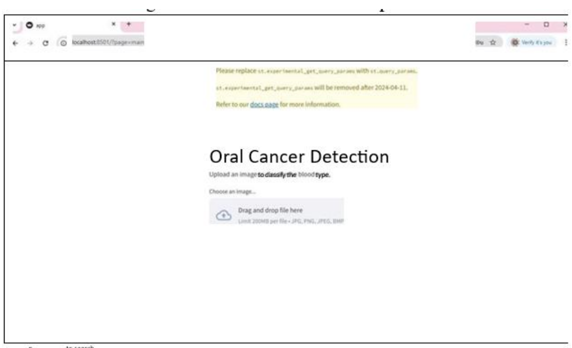

Figure 3: Register Page

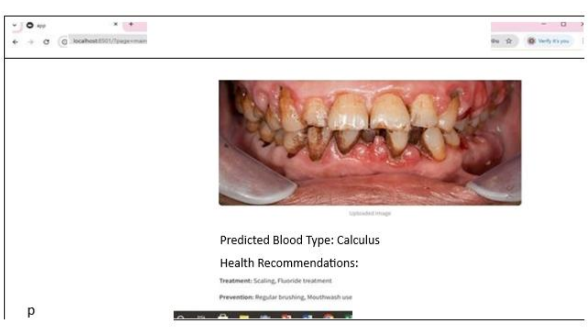

Figure 3: The calculus blood type for oral cancer

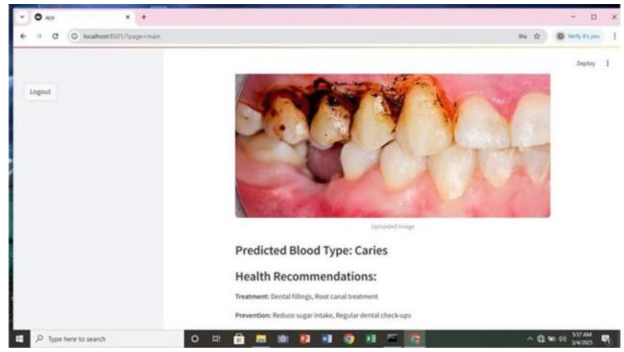

Figure 4: The caries blood type for oral cancer

VII . FUTURE SCOPE

Future developments in oral cancer detection using deep learning and Random Forest algorithms are poised to improve diagnostic accuracy and accessibility. Integrating models like CNNs and transformers can enhance feature extraction, while mobile-based AI applications can support real-time, low-cost screening. High-resolution imaging and personalized risk assessment using patient data may further refine early detection. Cloud-based platforms and AI-assisted clinical tools will enable collaborative, scalable, and efficient diagnostic systems for better patient outcomes.

REFERENCE

1. https://www.cancer.org/cancer/oral-cavity-and-oropharyngealcancer/about/keystatistics.html.KeyStatistics for Oral Cavity and Oropharyngeal Cancers. American Cancer Society.
2. https://www.cancer.org/cancer/oral-cavity-and-oropharyngeal-cancer/causesrisks-prevention/riskfactors.html. Risk Factors for Oral Cavity and Oropharyngeal Cancers. American Cancer Society.
3. Dargan S, Kumar M, Ayyagari M R. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning Arch. Computat Methods Eng 2020; 20: 1071-1092.

4. Prabhakaran R, Mohana J. Detection of Oral Cancer Using Machine Learning Classification Methods. Int J of Elect Eng and Tech 2020; 11(3): 384-393.
5. Shujun H, Nianguang C, Pedro P, Shavira N, Yang W, Wayne X. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics & Proteomics 2018; 15: 41-51.
6. Khanna D, Sharma A. Kernel-Based Naive Bayes Classifier for Medical Predictions. Adv in Int Sys and Compvol 695 Springer Singapore.
7. Lavanya L, Chandra J. Oral Cancer Analysis Using Machine Learning Techniques. ISSN 0974-3154, Volume 12, Number 5 2019: 596-601.
8. Licheng J, Jin Z. A Survey on the New Generation of Deep Learning in Image Processing.1EEE Access 2019; 7: 172231-172263.
9. Daisuke K, Shumpei I. Machine Learning Methods for Histopathological Image Analysis.Compt and Struct Biotec J, 2020; 16: 34-42.
10. Anne H-H. Texture Feature Extraction Methods: A Survey. Journal of IEEE Access,