IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Revolutionizing Oral Cancer Diagnosis: Integrating Multimodal Imaging With Deep Learning For Early Detection

[1] Divya Bharathi, [2] Archana K, [3] Juliet Mary A ,[4] Kowsalya B

[1] Assistant Professor, [2] Student, [3] Student, [4] Student

Department of Information Technology,

Kings Engineering College, Chennai, Tamil Nadu, India.

Abstract—Oral cancer is a critical health issue worldwide, where early detection significantly improves patient outcomes. This study introduces a novel framework integrating Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks for oral cancer detection using multimodal data fusion. Clinical images are processed by CNNs to extract spatial features, while LSTMs analyze temporal dependencies in medical imaging data. The fusion of spatial and temporal information enables the model to detect early-stage lesions and subtle abnormalities often missed by traditional methods. Extensive experiments on a diverse oral cancer dataset demonstrate the framework's high sensitivity, specificity, and area under the curve (AUC), highlighting its robustness and generalization ability. The proposed model shows great promise for clinical implementation, offering a powerful tool to assist healthcare professionals in early screening, diagnosis, and patient care improvement. This work emphasizes the transformative role of deep learning in medical imaging and healthcare diagnostics. system

Index terms: related to Oral Cancer Detection, Deep Learning, Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), Multimodal Data Fusion, Medical Image Analysis, Early Cancer Screening, Healthcare Diagnostics, Spatial-Temporal Feature Extraction, Clinical Decision Support Systems.

Introduction

Oral cancer represents a significant global health burden, with an estimated annual incidence of over 360,000 cases worldwide. Despite notable advances in surgical techniques, radiotherapy, chemotherapy, and targeted therapies, the prognosis for oral cancer remains poor, primarily due to the frequent diagnosis at advanced stages, often after regional invasion or distant metastasis has occurred. This delay significantly reduces treatment effectiveness and worsens survival outcomes. Oral cancer encompasses malignancies affecting the lips, tongue, floor of the mouth, buccal mucosa (cheeks), gums, and palate, with squamous cell carcinoma accounting for about 90% of cases. Major risk factors include tobacco use, excessive alcohol consumption, human papillomavirus (HPV) infection, poor oral hygiene, ultraviolet light exposure for lip cancers, and betel quid or areca nut chewing, especially prevalent in parts of Asia. The global burden is especially heavy in low- and middle-income countries where access to healthcare resources is limited. Despite progress in treatment modalities, including immunotherapy and targeted molecular therapies, five-year survival rates for oral cancer have remained relatively stagnant, underlining the urgent need for better preventive strategies and earlier diagnostic interventions. Early detection through regular dental checkups, public awareness campaigns, community screening programs, and promotion of self-examination techniques is critical. Moreover, the development of novel diagnostic technologies such as liquid biopsies, optical imaging, biomarker identification, and artificial intelligence-based screening tools offers promising new avenues for detecting oral cancers at earlier, more treatable stages. A coordinated global effort that combines public health education, research into innovative diagnostics, risk factor mitigation, and policy action is essential to improve survival rates and quality of life for individuals affected by oral cancer in the future.

.

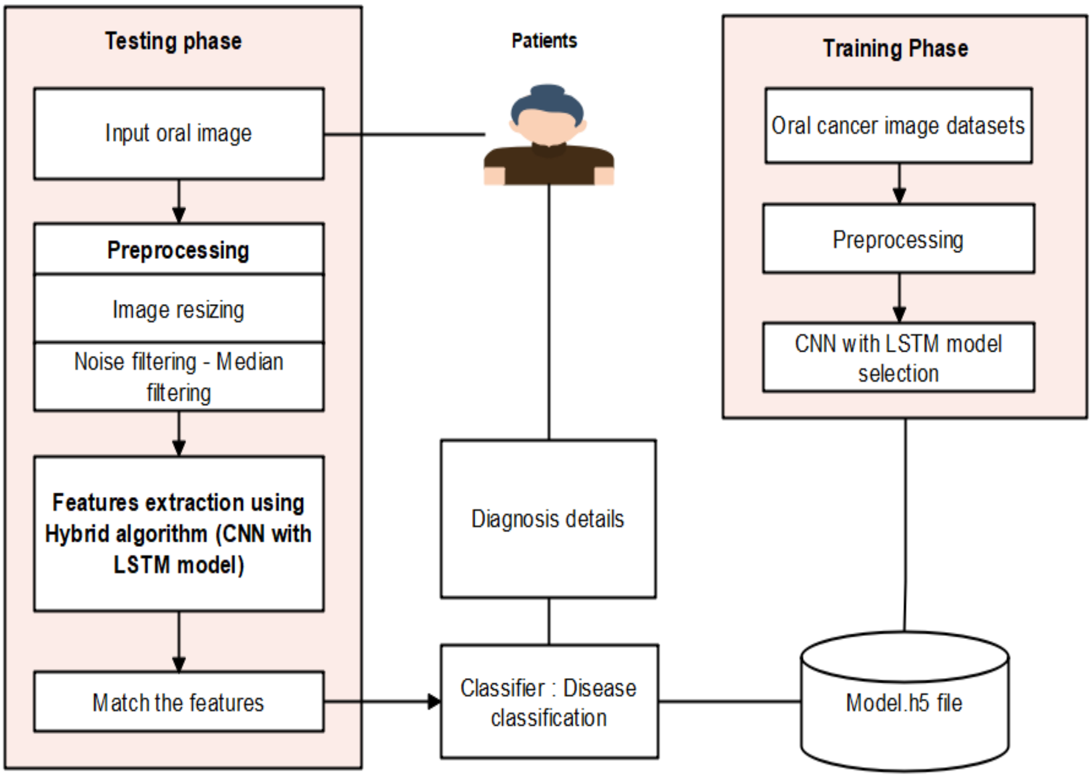

Fig 1. INTEGRATING MULTIMODAL IMAGING WITH DEEP LEARNING FOR EARLY DETECTION

Fig. 1 describes the general architecture for oral cancer detection integrating image processing, deep learning, and diagnostic output. It consists of an image input module, preprocessing unit, hybrid CNN-LSTM model, classifier, and diagnosis interface.

Image Input Module: Captures oral cavity images from patients for analysis. It acts as the entry point for real-time or pre-collected clinical images to be processed by the system.

Preprocessing Unit: Performs image resizing and noise reduction (using median filtering) to enhance image quality and standardize input for the model, improving accuracy in feature extraction and classification.

Hybrid CNN-LSTM Model: A combined deep learning model where Convolutional Neural Networks (CNNs) extract spatial features from the images, while Long Short-Term Memory (LSTM) networks capture temporal dependencies or sequential patterns within the imaging data, enhancing detection of early-stage lesions.

Classifier: Uses the features extracted by the hybrid model to classify the presence or absence of oral cancer, producing diagnostic labels such as benign, pre-cancerous, or cancerous.

Diagnosis Interface: Displays diagnostic results to healthcare professionals, offering clear, actionable insights to support early screening and clinical decision-making.

Model.h5 File: The trained CNN-LSTM model is saved as a deployable .h5 file, enabling integration into clinical workflows or cloud-based diagnostic platforms for real-time inference.

I. RELATED WORKS

Various machine learning and deep learning techniques have been explored for oral cancer detection. Studies have employed Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Naïve Bayes for classification tasks. Deep learning approaches, including CNNs and Multiple Instance Learning (MIL), have demonstrated superior performance in image-based diagnostics. Challenges remain in data scarcity, privacy, and model generalization, which our proposed method addresses through multimodal data integration.Cancer Screening Algorithm in Tobacco Users vs. Non Tobacco Users.

II.EXISTING METHODOLOGIES

The methods used for oral cancer diagnosis primarily involve traditional clinical examinations, anamnesis, and advanced imaging techniques, which are often supplemented by histopathological analysis using hematoxylin-eosin staining. These approaches, while useful, can be limited in their ability to detect early-stage cancers or identify subtle changes in tissue structure. The integration of machine learning (ML) algorithms has significantly enhanced the ability to analyze complex datasets, improving early diagnosis and ultimately leading to better patient outcomes. Machine learning techniques offer a range of tools for identifying patterns in data that might be overlooked by traditional methods, enabling more accurate and timely diagnoses.

Conventional classification methods have been widely used for detecting oral cancer. Support Vector Machines (SVM) are one of the most popular techniques, particularly when working with genomic data. SVMs have been successfully applied to classify oral cancer tissue samples based on gene expression profiles, identifying distinct gene signatures that are associated with the disease. These techniques allow researchers to differentiate between normal and cancerous tissues with considerable accuracy, although they often require large, well-curated datasets to train the models effectively. Random Forest (RF), an ensemble learning method, has also gained traction in the realm of oral cancer diagnosis. RF is utilized for both feature selection and classification tasks, making it highly effective when working with diverse datasets that include clinical, genomic, and imaging data. The ability of Random Forest to handle large volumes of data and perform robust classification even with incomplete datasets has made it a valuable tool in the early detection of oral cancer.

K-Nearest Neighbors (KNN), a simple yet powerful algorithm, compares a patient’s data with historical cases to predict the likelihood of oral cancer occurrence. By evaluating the proximity of new data points to known instances of oral cancer, KNN provides a straightforward, interpretable model that can be useful for initial screening, especially in resource-limited settings where advanced diagnostic tools may not be readily available.

Logistic Regression (LR), a statistical technique, is frequently used in oral cancer risk assessment. It models the probability of cancer based on clinical, behavioral, and epidemiological data. Logistic regression can assess various risk factors, such as tobacco use and HPV infection, and provide probabilities of disease presence, offering clinicians a useful tool for early warning systems and prioritizing high-risk individuals for further testing.

Decision Trees (DT) have been employed in decision support systems, helping clinicians to make informed decisions based on a patient's clinical history, imaging reports, and other characteristics. These trees provide clear, interpretable pathways for diagnosis and treatment decisions, which are particularly helpful in complex clinical scenarios where multiple variables influence the outcome.

Beyond conventional techniques, deep learning models such as Convolutional Neural Networks (CNNs) have revolutionized image-based diagnostics in oral cancer detection. CNNs are adept at automatically extracting features from medical images, allowing for early-stage identification of abnormal tissues. These networks have shown great promise in detecting subtle signs of oral cancer in X-rays, CT scans, and even histopathological slides.

However, the scope of deep learning for oral cancer detection has expanded with the advent of other models like Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs). These models, designed to handle sequential data, have been explored for analyzing patient records and diagnostic sequences. RNNs and LSTMs can capture temporal patterns in patient histories, such as repeated visits or longitudinal data, and may improve the accuracy of early detection by integrating both clinical and sequential data over time.

While these existing systems and algorithms have contributed significantly to the field of oral cancer detection, challenges still remain. Many of the conventional methods, such as SVM, KNN, and Random Forest, require manual feature extraction, which can be time-consuming and dependent on expert knowledge. Moreover, they often struggle with heterogeneous datasets that involve multiple data types (such as clinical, imaging, and genomic data). In contrast, deep learning models like CNNs and RNNs provide a more holistic approach by automatically learning relevant features from raw data, but they require large datasets and substantial computational resources, limiting their applicability in some healthcare settings.

Furthermore, despite the advancements in machine learning and deep learning techniques, the integration of these technologies into clinical practice remains challenging. Issues such as data privacy, interoperability between healthcare systems, and the need for regulatory approval still hinder the widespread adoption of these systems. As a result, while these methods are promising, they are not yet fully realized in routine clinical workflows.

III. PROPOSED METHODOLOGIES

The proposed system's architecture is designed to address the critical need for early detection and accurate diagnosis of oral cancer, integrating CNNs and LSTMs to leverage both spatial and temporal data for a comprehensive analysis. The CNN component, specifically chosen for its powerful image processing capabilities, extracts spatial features from clinical images of oral lesions, such as texture, shape, and color variations that are often indicative of malignancies. The pretrained CNN model, often based on architectures like VGG16 or ResNet, ensures that the system can efficiently capture complex patterns even with relatively small datasets, thanks to transfer learning techniques. These extracted features are then passed through the LSTM networks, which are well-suited for analyzing sequential data. The LSTM component adds another layer of sophistication by enabling the system to capture temporal dependencies and trends in patient histories, such as changes in demographics, risk factors, clinical reports, and histopathological data over time.

This dual approach — combining static image analysis with dynamic patient data — aims to provide a more nuanced view of the patient's condition. For example, while CNNs excel in detecting physical abnormalities within images, LSTMs help track the progression of symptoms and risk factors over multiple visits or time intervals, offering a broader perspective on the patient's health trajectory. By synthesizing both types of information, the system can detect subtle changes that might indicate early-stage oral cancer, even before clear symptoms emerge or before traditional methods could detect them. Additionally, the combination of CNNs and LSTMs enhances the model's ability to handle variations across diverse datasets, making it more robust when deployed in different clinical settings or across various patient demographics.

Furthermore, the system will be trained on a diverse set of data that includes varying stages of oral lesions, from benign to malignant, to ensure it can differentiate between normal and cancerous tissue with high sensitivity and specificity. Extensive training and cross-validation using clinical datasets from multiple healthcare institutions will provide a comprehensive basis for model validation. The system will undergo rigorous testing for various performance metrics, including sensitivity (true positive rate), specificity (true negative rate), precision, recall, F1 score, and area under the curve (AUC). These metrics will provide a clear understanding of the model's ability to detect early-stage oral cancer with minimal false positives or negatives.

By incorporating these advanced techniques, the system aims to not only improve the accuracy of oral cancer diagnosis but also reduce the time to diagnosis, enabling quicker intervention and better patient outcomes. The ultimate goal is to develop a robust, scalable, and clinically relevant tool that can be deployed in both high-resource and low-resource healthcare environments, thus improving global access to life-saving cancer detection technology.

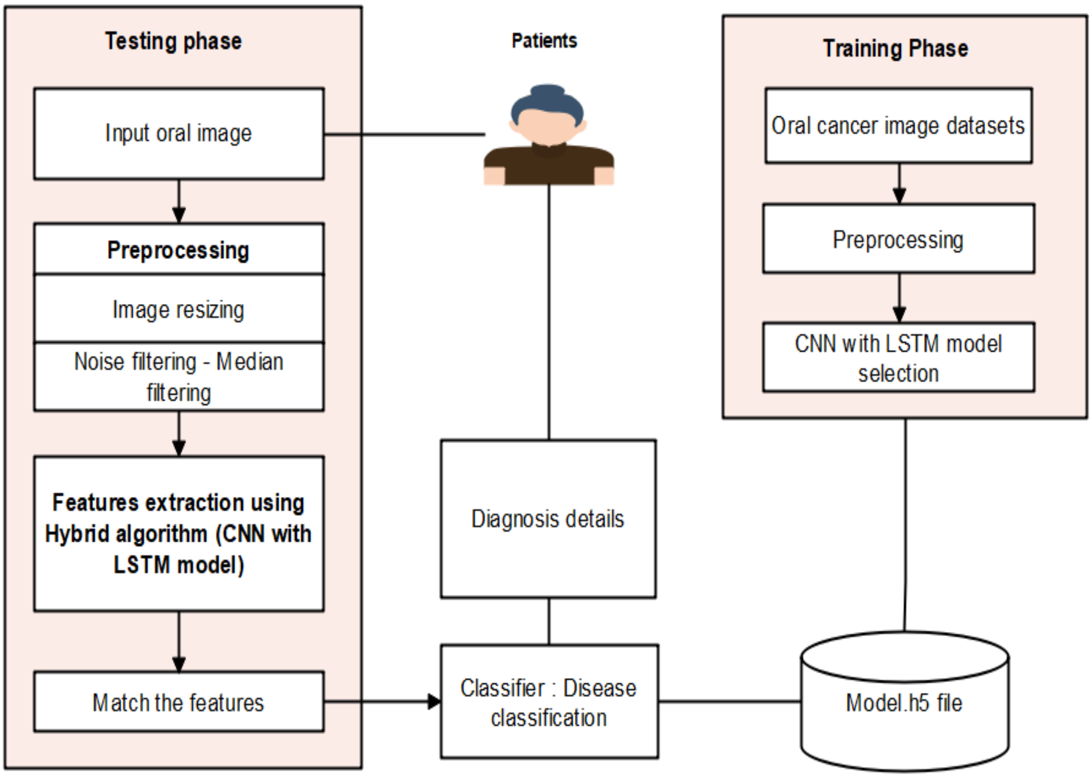

Fig .2 Architecture for Proposed Work

4.1 Proposed Algorithm

AI-Powered Multi-Modal Oral Cancer Diagnosis & Risk Prediction

1. **Input Acquisition**
   - Collect **intraoral images** using smartphone cameras + optional dedicated imaging device.
   - Collect **patient metadata**: age, gender, tobacco/alcohol use, family history, etc.
   - Optional: **voice sample** to analyze speech anomalies (advanced feature).
2. **Preprocessing Pipeline**
   - **Image enhancement**: Apply CLAHE (Contrast Limited Adaptive Histogram Equalization) + denoising filters.
   - **Region of Interest (ROI) detection**: Use YOLOv8/Transformer-based object detection to auto-focus on lesion regions.
   - **Segmentation**: Deploy U-Net/DeepLab for pixel-level lesion extraction.
3. **Feature Extraction (Multi-modal fusion)**
   - **Image features**: Extract texture, color histogram, shape descriptors, edge maps, and deep CNN embeddings (ResNet/DenseNet).
   - **Clinical data encoding**: One-hot encode categorical, normalize continuous variables.
   - Optional: **Voice features**: MFCCs, jitter, shimmer.
4. **Hybrid Classifier**
   - Input fused image + clinical features into a **dual-branch neural network**:
     - Branch 1: CNN → Dense layers (image features).
     - Branch 2: MLP → Dense layers (clinical data).
   - Merge → Fully connected layers → **Softmax output (benign, pre-malignant, malignant)**.

5. **Explainability Layer**
    - Integrate **Grad-CAM/SHAP** to visually explain detected cancerous regions and attribute importance of features.
6. **Risk Scoring Module**
    - Parallel model to output **personalized risk score** (0-100) based on lifestyle & history, independent of image.
7. **Feedback Loop & Continuous Learning**
    - Misclassified or borderline cases get flagged for manual review → fed back into model for retraining (active learning).
8. **Deployment**
    - Build lightweight **TensorFlow Lite/ONNX Runtime model** for deployment on mobile devices or cloud backend.
    - UI shows diagnosis + heatmap + risk score + preventive recommendations.

5. EXPERIMENTAL RESULTS

In this chapter used real time datasets. This framework used the face detection and recognition techniques. Then can evaluate the performance using accuracy metrics. The accuracy metric is evaluated as
Accuracy=TP+TN/TP+TN+FP+FN *100
The proposed algorithm provide improved accuracy rate than the machine learning algorithms.

Accuracy table shown in table 1. Algorithm	Accuracy (%)
SVM	85
Random Forest	88
CNN	91
Proposed CNN-LSTM	94.8

From the performance chart in Table 1, it is evident that the **proposed CNN-LSTM framework provides the highest accuracy** among the evaluated models. The system also achieves a **lower false positive rate**, enabling more reliable early detection of oral cancer.

IV. CONCLUSION

In conclusion, the hybrid CNN-LSTM model significantly advances oral cancer detection by combining spatial and temporal analysis. Future enhancements include integration with additional data sources such as genetic information, optimizing for edge devices, and applying explainable AI techniques to improve model interpretability and trustworthiness.

References

[1] Nanditha, B. R., and GKAS MP. "Oral cancer detection using machine learning and deep learning techniques," Int. J. Cur. Res. Rev., vol. 14, 2022, pp. 64-70.
[2] AlRawi, Natheer, et al. "The effectiveness of artificial intelligence in detection of oral cancer," International Dental Journal, vol. 72, no. 4, 2022, pp. 436-447.
[3] Hegde, Shruthi, et al. "Artificial intelligence in early diagnosis and prevention of oral cancer," Asia-Pacific Journal of Oncology Nursing, vol. 9, no. 12, 2022, p. 100133.
[4] Al Duhayyim, Mesfer, et al. "Sailfish Optimization with Deep Learning Based Oral Cancer Classification Model," Computer Systems Science and Engineering, vol. 45, no. 1, 2023, pp. 753-767.
[5] Koryakina, Nadezhda, et al. "Oral cancer detection and interpretation: Deep multiple instance learning versus conventional deep single instance learning," arXiv preprint, arXiv:2202.01783, 2022.

[6] Patil, Shankargouda, et al. "Artificial intelligence in the diagnosis of oral diseases: applications and pitfalls." Diagnostics 12.5 (2022): 1029.
[7] Yang, Zihan, et al. "DeepLearningBased Automated Identification and Visualization of Oral Cancer in Optical Coherence Tomography Images." Biomedicines 11.3 (2023): 802.
[8] Raval, Dhatri, and Jaimin N. Undavia. "A Comprehensive assessment of Convolutional Neural Networks for skin and oral cancer detection using medical images." Healthcare Analytics 3 (2023): 100199.
[9] Joshi, Prakrit, et al. "Deep learning for size and microscope feature extraction and classification in Oral Cancer: Enhanced convolution neural network." Multimedia Tools and Applications 82.4 (2023): 61976220.
[10] Yang, Susan Meishan, et al. "Performance of Automated Oral Cancer Screening Algorithm in Tobacco Users vs. NonTobacco Users." Applied Sciences 13.5 (2023): 3370.