

Scientiarum: A Multidisciplinary Journal
Vol. 2 No. S1 (2026): Supplement Issue
DOI: 10.54646/SAPARS.2026.57

SAPAR SERVICES LLP
Scholarly Articles Publication And Repository Services

EXPLAINABLE AI FRAMEWORK FOR EARLY ORAL CANCER PREDICTION USING DEEP LEARNING

Navya R[1], Parnapalli Pushpanjali[2]

[1]MCA Student

[2]Assistant Professor

Department of Computer Science and Applications

S.E.A. College of Science, Commerce and Arts

navyanavi2602@gmail.com[1]

pushparoyal777@gmail.com[2]

Abstract

Oral cancer is a major heath concern that effects the thousands of people every year, early detection plays a crucial role in improving treatment outcomes and increasing the chances of survival. Deep learning has shown promise in analysing medical images automatically. Many of these models work like black boxes. This means it's hard to understand how they make decisions. That lack of clarity can be a problem in healthcare, where trust and transparency essential.

This study introduces an artificial intelligence framework for predicting oral cancer early. The goal is to help doctors by providing insights into predictions. The system uses learning and combines information from different sources. It looks at both pictures of the mouth and clinical data. The clinical data includes things like smoking history, use of tobacco and alcohol hemoglobin levels, fever, weight loss trouble swallowing, white or red patches and non-healing sores.

Before any prediction is made there is a validation step. This step checks whether the image actually shows the inside of the mouth. It helps avoid errors caused by unrelated inputs. An EfficientNet model is used to extract features from the images. At the time the structured clinical data is processed separately and then combined with the image features. This fusion allows for predictions based on both visual and patient-specific details.

To make the model’s predictions understandable three XAI methods are used. Grad-CAM highlights areas in the images. LIME explains predictions by approximating the model locally. SHAP provides explanations that show which features contribute most to the outcome. These tools together give a complete picture of why a prediction was made.
Keywords: Oral Cancer Prediction, Explainable Artificial Intelligence, Multimodal Learning, EfficientNet, Medical Image Analysis

Introduction

Oral cancer is a public health concern, especially in groups where tobacco use, alcohol use and other risky habits are common. Oral cancer can affect the lips, tongue, cheeks, gums and other parts of the mouth. How well a patient does depends a lot on how early oral cancer's found. If doctors find mouth spots early they can check the patient sooner and start treatment faster which can help the patient. Usual ways to screen and diagnose rely on looking at the mouth and on the skill of the doctor and then on tests such as a biopsy. These steps take time. Can be hard to use the same way when resources are scarce.

More medical picture collections and new deep learning technology have opened doors for computer helpers that spot cancer. Convolutional neural networks or CNNs and models that use transfer learning can learn picture clues by themselves and decide if an image shows oral cancer or not. Even though these deep learning models can be very accurate it is hard for doctors and patients to see how they come to their decision. That is why it is important to build an AI system that's both correct and easy to understand so it can aid doctors instead of just giving a label.

To fix the lack of clarity we can add Explainable Artificial Intelligence or XAI to deep learning models. Tools like Gradient-weighted Class Activation Mapping or Grad-CAM Local Interpretable Model-agnostic Explanations or LIME and SHAP give ways to see how the model works. Grad-CAM shows the parts of the picture that matter for the decision. LIME explains a prediction in detail. SHAP tells us how much each input feature helps the result.

This paper shows OralAI Detect, a clear-AI system that predicts cancer early by using deep learning and many types of patient data. OralAI Detect looks at mouth pictures. Also uses useful medical facts like whether the patient smokes, drinks etc. Before the

system decides if oral cancer is present it first checks that the picture is suitable for analysis. This step stops the model from looking at pictures.

The main contributions of this work are: (1) a deep learning framework that mixes mouth pictures and medical facts to predict cancer; (2) a picture-validation step that stops bad images from reaching the oral cancer model; (3) the use of Grad-CAM, LIME and SHAP to give clear predictions; and (4) a user-friendly app that shows results and explanations in a simple way. OralAI Detect is meant to help doctors and patients not to replace medical exams.

Literature Review

Deep learning has been increasingly used in the detection and classification of oral cancerous growths using medical images. Convolutional neural networks (CNNs) can automatically extract visual features from oral images and identify patterns associated with suspicious lesions. Previous research has investigated these methods for oral squamous cell carcinoma (OSCC) detection, suggesting that the use of artificial intelligence would support image-based assessment; however, variations in image quality, the size and imaging modalities of the dataset remain significant challenges. Transfer learning has been investigated to improve the performance of medical image classification where access to a large set of labelled data is limited. EfficientNet is a convolutional neural network which balances model depth, width, and resolution to extract image features efficiently. It has been utilized in oral cancer prediction as an approach to learning visual features from oral images. This approach's effectiveness is determined by the training data and the validation methodology.

Explainable artificial intelligence (XAI) has become a key research area in medical prediction systems, as visualizing the decision-making process can improve the interpretability of the results and increase trust in the technology. Grad-CAM provides visual heatmaps, indicating the image regions contributing to the CNN's output, while LIME and SHAP provide explanations of individual predictions and the contribution of features, respectively. However, this information should not be interpreted as evidence of disease.

Recently, research into medical prediction has investigated multimodal learning, referring to the combination of information from multiple data sources. In oral cancer assessment, this might involve integrating image features with clinical variables to provide additional information not available from a single source; however,

combining these inputs must be done carefully to prevent overfitting and misleading predictions.

The proposed OralAI Detect framework builds upon previous research by combining EfficientNetB0-based oral image analysis with structured clinical information. It incorporates Cancer/Non-Cancer prediction and explainability methods such as Grad-CAM, LIME, and SHAP, to make the model's outputs more interpretable. This approach addresses the research need to explore image-based prediction alongside clinical information and model explanations within a single framework.

Related Work

Recent studies have shown that Artificial Intelligence (AI) and deep learning can help detect cancer and classify oral lesions automatically. Convolutional Neural Networks (CNNs) and transfer-learning models have been widely studied because they can learn visual patterns from medical and clinical images. Systematic reviews suggest that deep learning methods work well for identifying cancer and other oral mucosal lesions. Still differences in size, image quality, how images are taken and how results are tested remain major issues [1] [2]. Alongside image data patient details like tobacco use, alcohol consumption, lasting sores, white or red patches trouble swallowing and unexplained weight loss can also help assess cancer risk. Because of this multimodal AI approaches that combine images with structured clinical data have become more popular. These approaches use more than visual clues and can make better predictions by drawing from different types of information.

Another key area of research is Explainable Artificial Intelligence (XAI) which helps make sense of how deep learning models work in healthcare. Many AI models are seen as " boxes" because they give results without explaining how they reached them. Techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) help explain model decisions. Grad-CAM shows which parts of an image affect a prediction. LIME explains predictions locally. SHAP calculates how much each input feature contributes to the result. Earlier work has shown that XAI can be useful in image analysis especially in detecting oral cancer [6]. However many current systems only focus on classifying images. They often don’t include clinical data or they don’t check if an uploaded image is suitable for cancer analysis before making a prediction. The proposed OralAI Detect framework solves these problems

by combining image validation deep learning for image analysis, clinical data processing, multimodal prediction and XAI methods like Grad-CAM, LIME and SHAP. All these tools work together in one system to give interpretable early predictions, for oral cancer.

Research Gap

Even though current studies have shown that deep learning works well for classifying cancer and oral lesions there are still some areas that need more research. Most of the methods use oral images and try to get high accuracy when classifying but they don't pay much attention to combining image information with patient-specific clinical features. Also differences in image quality how the images are. The way datasets are set up can make image-based prediction systems less reliable. Another problem is that an image that is uploaded may not always be an oral image but traditional cancer classification models still try to process it. Many deep learning systems only give a final result without explaining the visual and clinical reasons behind the decision. Even though techniques like Grad-CAM, LIME and SHAP have been studied their use with image validation and making predictions using multiple types of data is still not very common.

To fix these problems the proposed OralAI Detect framework brings together parts in one early oral cancer prediction system. The framework starts with a stage that checks if the image is suitable for analysis then uses learning to classify the image. It also uses features along with the image information to make predictions that use multiple types of data. This allows the system to look at both the details and the patients personal information. In addition Grad-CAM, LIME and SHAP are used to help explain the models decisions. So the research gap that this work is solving is the lack of a complete and easy to understand system that combines image validation, deep learning, clinical features, multimodal prediction and explainable AI, for early oral cancer detection.

Proposed Methodology

The proposed OralAI Detect framework is built as a deep learning system that aims to predict oral cancer at an early stage. This system uses both images and patient-specific clinical information.

First the system collects features that are relevant to oral cancer. These features include tobacco and alcohol exposure, hemoglobin level, fever, weight loss difficulty swallowing, white patches, red patches and non-healing ulcers. The system also receives an oral image.

Before the system makes a cancer prediction the uploaded image goes through an image validation stage. This step checks whether the image is suitable for analysis. Once the image is validated it is preprocessed. The preprocessing includes resizing, changing color format and normalizing the image. After preprocessing the image is fed into the trained deep learning model.

The image model extracts visual features that are important for detecting cancer. Meanwhile the clinical information is turned into a feature representation.

The two representations are then merged using a multimodal fusion process. The result of this fusion is a prediction: Cancer or Non-Cancer.

To make the prediction process clearer the proposed OralAI Detect framework uses Explainable Artificial Intelligence or XAI, techniques. Grad-CAM creates a heatmap that shows which parts of the oral image drive the deep learning prediction. LIME gives an interpretation for each individual prediction. SHAP estimates how much each input feature, the structured clinical information contributes to the final output of the model. By combining image validation, deep learning, clinical feature processing, multimodal fusion and XAI OralAI Detect can offer a prediction and supporting explanations, through its application interface. The OralAI Detect framework is meant to serve as a screening and decision-support tool. It does not replace a clinical examination or a histopathological diagnosis (Figure 1).



Figure 1: *Proposed Methodology*

Experimental Setup

The OralAI Detect framework was built using a mix of learning, image processing, web development and database technologies. The learning parts were developed with Python and TensorFlow/Keras. OpenCV was used for handling image preprocessing and validation tasks. The user interface was created using React.js to allow clinicians to enter details and upload oral images easily. A backend system was set up to manage prediction requests and application functions. MongoDB was used to store user data and prediction results. The trained deep learning models were connected to the application to offer automated cancer predictions along with explanations for each result.

The evaluation of the system uses both image data and clinical information for a multimodal prediction approach. The oral image dataset includes images of both cancer and -cancer cases. The clinical dataset contains features that are relevant to oral cancer diagnosis. Before feeding the data into the model the images are. Normalized. The clinical features are converted into values so they can be used by the model. The datasets are split into training, validation and testing sets to build and test the model effectively. Model performance is measured using classification metrics, like accuracy, precision, recall, F1-score and AUC-ROC.

Result and Discussion

The OralAI Detect framework was tested to see how well it can predict cancer when it uses both oral images and clinical data. The trained deep learning model takes a validated image and the chosen clinical attributes and then gives a Cancer or Non-Cancer prediction along with a confidence score. We measured the system's performance with classification metrics: accuracy, precision, recall, F1-score and area under the receiver operating characteristic curve (AUC). These measures give a picture of how well the model can find both cancer and non-cancer cases.

Besides classification performance we also looked at the explainability part using Grad-CAM, LIME and SHAP. Grad-CAM creates heatmaps that show which parts of the oral image helped the prediction. LIME gives an explanation of the model's choice and SHAP tells how much each clinical feature matters for the final prediction. These explanations make the system easier to understand for users and healthcare

professionals so they can see what drives the prediction. The results show that mixing multimodal learning, with explainable AI can give a clearer early oral cancer scr isions and sho is also used to exa er cases. Be[AP are rev h factors infl

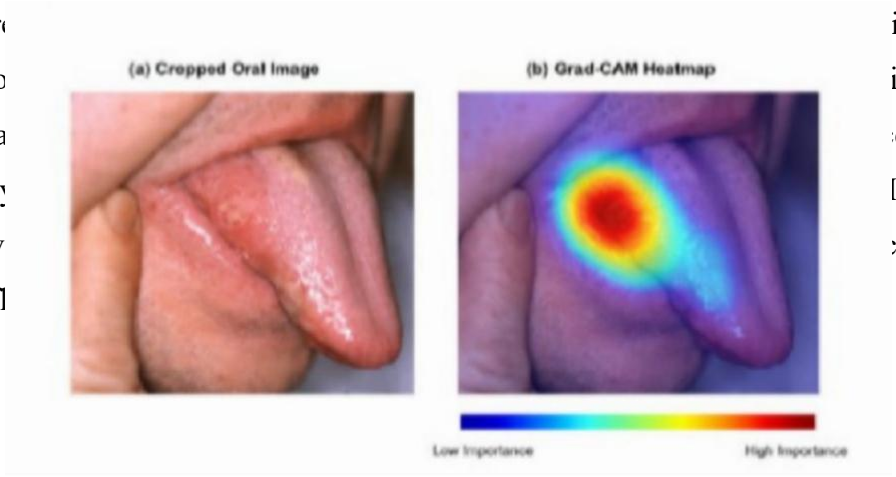

Figure 2: *Grad-CAM visualization of the model prediction highlighting the important region in the oral image*

Challenges and Limitations

The development and implementation of an AI-based oral cancer prediction system involves several technical, data-related, and clinical challenges. One of the major challenges is the availability of a sufficiently large, balanced, and diverse dataset of oral images. Variations in image resolution, illumination, camera quality, lesion location, background, and image acquisition conditions can affect the performance of deep learning models. A limited dataset may also increase the possibility of overfitting and may reduce the model's ability to generalize to images obtained from different sources.

Another challenge is the integration of image-based information with structured clinical information. Clinical features may contain missing, inconsistent, or subjective values, requiring appropriate preprocessing before they can be combined with image features. The selection of relevant clinical attributes is also important because irrelevant or incomplete information may negatively influence prediction performance.

Explainability is another important challenge. Although Grad-CAM can highlight image regions associated with a prediction and LIME and SHAP can indicate the contribution of individual features, these explanations describe model behavior rather than providing definitive clinical evidence. The consistency and reliability of explanations therefore need to be evaluated carefully.

Data privacy and security are additional concerns because the system handles medical images and patient-related clinical information. Appropriate access control, secure storage, and protection of patient information are necessary. Finally, the current system represents a research prototype and should not be considered a replacement for professional clinical examination, biopsy, or medical diagnosis. Independent validation using larger datasets and clinical studies is required before practical clinical deployment.

Future Scope

The proposed OralAI Detect framework provides several opportunities for future enhancement. The first area of improvement is dataset expansion. Future research can incorporate larger, balanced, and geographically diverse oral image datasets collected under different clinical and imaging conditions. Including images from multiple hospitals and healthcare centers could help evaluate the generalization of the model across different patient populations and image acquisition environments.

The classification capability of the system can also be extended beyond the current Cancer and Non-Cancer categories. Future versions could identify different types of oral lesions, including oral potentially malignant disorders and different stages or categories of oral cancer, provided that sufficiently annotated datasets are available.

Another important direction is improving multimodal learning. More advanced architectures could be investigated for combining image features with clinical information. Additional clinically relevant attributes could also be evaluated to determine whether they provide useful complementary information without introducing unnecessary complexity or bias.

Explainability can be further improved by investigating additional XAI techniques and comparing the consistency of Grad-CAM, LIME, and SHAP explanations. Future work could also evaluate explanations with the assistance of dental and medical professionals to determine whether the highlighted regions and important clinical features are meaningful from a clinical perspective.

The system could also be developed into a secure mobile or web-based screening-support application that allows images to be captured using mobile devices. Integration with hospital information systems could enable secure management of patient records and prediction history. Privacy-preserving approaches can additionally be explored to protect sensitive medical information.

Finally, extensive external validation, prospective studies, and evaluation by qualified healthcare professionals would be necessary before considering real-world clinical use. Future research can focus on improving model robustness, reducing false predictions, monitoring model performance over time, and ensuring that AI remains a decision-support tool under appropriate clinical supervision.

Conclusion

This research presented OralAI Detect, an explainable artificial intelligence framework developed for early oral cancer prediction using deep learning and clinical information. The proposed framework integrates oral image analysis with relevant patient clinical features to perform Cancer and Non-Cancer classification. EfficientNetB0 is employed for extracting meaningful visual representations from oral images, while clinical information is incorporated to provide an additional source of predictive information.

A major feature of the proposed framework is the incorporation of Explainable AI techniques. Grad-CAM provides visual information about the regions of an oral image that contribute to the model prediction, while LIME and SHAP provide feature-level explanations for individual predictions. These explanations are intended to improve the transparency and interpretability of the prediction process and help users understand the factors associated with a model output.

The system also provides patient registration, authentication, prediction history, image upload, clinical information management, and administrative functionalities, resulting in an integrated prototype rather than an isolated classification model. By combining deep learning, multimodal information, and explainability, OralAI Detect demonstrates the potential of AI-assisted approaches for supporting early oral cancer screening.

However, the system should be regarded as a computer-assisted research framework and not as an autonomous diagnostic tool. Clinical examination and appropriate medical investigations remain essential for confirming oral cancer. Further validation

on larger and independent datasets is necessary to determine the robustness, generalizability, and practical usefulness of the proposed approach.

References

Alanazi, A., et al. (2022). Intelligent deep learning enabled oral squamous cell carcinoma detection and classification using biomedical images. *Computational and Mathematical Methods in Medicine, 2022*, Article 7643967.

Cimino, M. G. C. A., et al. (2024). Towards explainable oral cancer recognition: Screening on imperfect images via informed deep learning and case-based reasoning. *Computerized Medical Imaging and Graphics*.

Cimino, M. G. C. A., et al. (2025). Explainable screening of oral cancer via deep learning and case-based reasoning. *Smart Health, 35*, Article 100538.

Chu, C. S., Lee, N. P., Ho, J. W. K., Choi, S. W., & Thomson, P. J. (2021). Deep learning for clinical image analyses in oral squamous cell carcinoma: A review. *JAMA Otolaryngology–Head & Neck Surgery, 147*(10), 893–900.

Explainable AI for oral cancer diagnosis: Multiclass classification of histopathology images and Grad-CAM visualization. (2025). *Biology, 14*.

Flügge, T., et al. (2023). Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer. *Scientific Reports, 13*, Article 2296.

Jubair, F., et al. (2022). A novel lightweight deep convolutional neural network for early detection of oral cancer. *Oral Diseases, 28*(4), 1123–1130.

Li, J. W., et al. (2024). Diagnostic accuracy of artificial intelligence assisted clinical imaging in the detection of oral potentially malignant disorders and oral cancer: A systematic review and meta-analysis. *International Journal of Surgery, 110*(8), 5034–5046.

Rokhshad, R., et al. (2024). Artificial intelligence for classification and detection of oral mucosa lesions on photographs: A systematic review and meta-analysis. *Clinical Oral Investigations, 28*(1), Article 88.

Sahoo, R. K., et al. (2024). Diagnostic performance of artificial intelligence in detecting oral potentially malignant disorders and oral cancer using medical diagnostic imaging: A systematic review and meta-analysis. *Frontiers in Oral Health, 5*.

Song, B., et al. (2021). Bayesian deep learning for reliable oral cancer image classification. *Biomedical Optics Express, 12*(10), 6422–6430.

Warin, K., Limprasert, W., Suebnukarn, S., Jinaporntham, S., & Jantana, P. (2022). Performance of deep convolutional neural network for classification and detection of oral potentially malignant disorders in photographic images. *International Journal of Oral and Maxillofacial Surgery, 51*(5), 699–704.

Warin, K., & Suebnukarn, S. (2024). Deep learning in oral cancer—A systematic review. *BMC Oral Health, 24*, Article 212.

Warin, K., et al. (2022). AI-based analysis of oral lesions using novel deep convolutional neural networks for early detection of oral cancer. *PLOS ONE, 17*(8), e0273508.

Yuan, W., et al. (2023). Noninvasive diagnosis of oral squamous cell carcinoma by multi-level deep residual learning on optical coherence tomography images. *Oral Diseases, 29*(8), 3223–3231.