Dual-Branch Deep Learning Framework for Oral Cancer Detection from Lip and Tongue Images

V. Gokula Krishnan[1], Arvind Kumar Tiwari[2]

[1] Post-Doctoral Research Fellow, Department of Computer Science and Engineering, Lincoln University College, Malaysia;

[2] Adjunct Professor, Lincoln University College, Malaysia;

Email ID : pdf.gokul@lincoln.edu.my

Abstract: Early detection of oral cancer plays a significant role in improving the patient's chances of survival. Visual screening of clinical photographs is a diagnostic tool that is currently being used; however, it is a difficult task given that there are different lighting conditions and the anatomical structures (and appearance of lesions) can be very different in lip and tongue pictures. All these difficulties affect the credibility of computer-aided diagnostic systems. In light of these circumstances, this work introduces a dual-branch deep learning architecture that exploits convolutional neural network features along with light texture descriptors for extracting global visual patterns and minute surface characteristics, respectively. The suggested method was tested on a public oral cancer image dataset consisting of lip and tongue photos. Results of the experiments show high classification effectiveness with an accuracy of 0.892, a macro-F1 score of 0.883, an AUROC of 0.912, and an AUPRC of 0.884. Probability calibration improved prediction confidence even more as it decreased the expected calibration error from 0.067 to 0.031. The results suggest that combining diverse visual features and adopting calibration techniques can significantly boost the performance of automated oral cancer screening. Through clinical decision-making, the proposed framework capable of being incorporated into cloud or mobile health systems and act as a tool for detecting oral cancer at its initial stage in telemedicine as well as community screening programs.

Keywords: Oral Cancer; Domain-adversarial alignment; Convolutional Neural Network; Attention gate; Lightweight texture branch.

Introduction

Oral cancer remains one of the critical health issues worldwide, especially because of its high fatality rate when it is only discovered at advanced stages. Treatment is greatly effective and survival rate is higher if the cancer is detected early. However, one of the difficulties in identifying cancerous lesions in images of lips and tongues is that the differences between malignant and benign tissues are often very subtle. These problems have been addressed quite successfully by the most recent improvements in artificial intelligence. This statement is especially true when we talk about CNNs - convolutional neural networks that are now capable of performing spatial analyses and understanding the context of oral cancer images quite well as is testified by [1]. Many researchers have first of all recognized the powerful feature extraction capabilities of CNNs and then designed their own customized CNN architectures and newly invented deep learning frameworks in order to obtain better diagnostic performance of oral cancer detection [2]. Another thing that has become very popular recently is to utilize the attention mechanism since such a model is able to learn to focus on the cancer-relevant areas directly from the oral photographs which results in higher quality of the model's interpretation. This is how it has been documented in [3]. Nevertheless, a lot of these days' solutions continue to heavily depend on the deep

convolutional features that are extracted by training the CNNs on large datasets. At the same time, they miss out on the fact that the subtle texture patterns could be considered as quite informative. In other words, capturing co-occurrence patterns of gray level pixel intensities that characterize texture properties - is not well represented by the CNN model itself.

Related work

AI-driven image analysis and deep learning methods have greatly changed the landscape of oral cancer screening by enhancing the detection capabilities. Traditional image processing and machine learning techniques initially made use of handcrafted features like texture, color statistics, and morphological characteristics along with classifiers (e.g. support vector machines) to identify the disease. Dvila Olivos et al. develop a clinically-oriented deep learning framework for oral cancer diagnosis using image data that yields higher accuracy than standard methods. Later on, Liu and Bagi presented a CNN architecture specifically designed for oral cancer detection from lip and tongue images that even further improved the performance of detection. In a similar vein, Begum and Vidyullatha proposed an attention mechanism coupled with a deep learning model that has the ability to annotate areas on oral images that are likely to have cancer making the classification result more explainable and accurate. Nevertheless, a lot of the current methods still primarily extract deep convolutional features and as a result might miss the subtle texture details present in oral lesions.

Table 1. Comparison of Selected Oral Cancer Detection Approaches

Reference	Method Used	Attention / Interpretability	Texture Features	Calibration	Dataset Type
[1]	Deep learning based classification	No	No	No	Clinical oral images
[2]	Customized CNN architecture	No	No	No	Lip and tongue images
[3]	Attention-based deep learning model	Yes	No	No	Oral cavity photographs
Proposed Work	Dual-branch CNN + texture features	Yes	Yes	Yes	Lip and tongue images

Key Contribution

The present investigation proposes a hybrid deep learning architecture for the automatic detection of oral cancer using clinical images. In this paper, we describe an approach that utilizes convolutional neural network (CNN) features combined with texture-based descriptors to enrich characterization of both the large-scale visual patterns and the subtle lesion details in lip and tongue images. Using an attention-based feature fusion module, the predictions from the two networks are combined, allowing the system to focus on the most relevant features for classification. The method, through a series of experiments, has shown a better performance in the classification accuracy as compared to the standard single-branch CNN architectures. This approach has the potential to support computer-aided oral cancer screening and oral cancer early diagnosis through telemedicine.

Method, Experiments and Results

This research introduces a deep dual-branch network that can effectively detect oral cancer by analyzing pictures of lips and tongues. The big picture of this research consists of five main steps: firstly, preprocessing of images, then extracting features through two branches which complement each other,

after that, fusing the features and finally, classification. At first, the input images are resized and normalized to a standard measure so that changes in lighting and size will be minimized. During training, data is rotated, flipped, and contrast is changed to improve model's generalization and decrease overfitting. The design and procedure flow of the suggested framework are depicted in Figure 1.

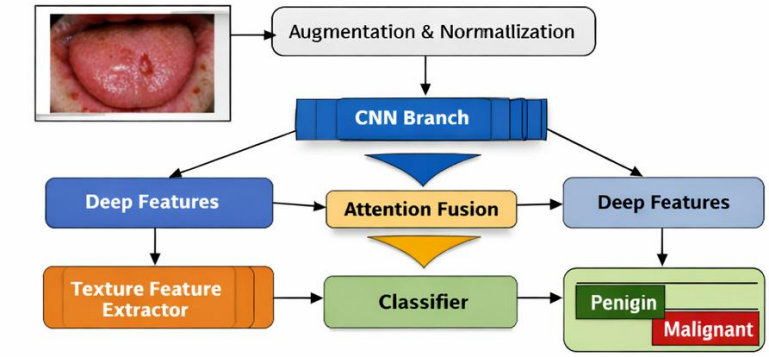

Figure 1: Overview of the proposed dual-branch model for oral cancer detection.

The initial part of the model deploys a convolutional neural network (CNN) to derive visual features at a high level from the images of the mouth. CNN layers recognize patterns through their hierarchical learning of different features such as changes in the outline of a lesion, structural alterations, and color differences of the malignant tissues. The second section deals with extracting texture-related features that distinguish even the smallest variations of the surface which can be used for early detection of abnormalities. Experiments were carried out on a dataset consisting of clinical pictures of lip and tongue areas of cancerous and non-cancerous cases. The model was evaluated by using standard classification metrics like accuracy, macro F1-score, area under the receiver operating characteristic curve (AUROC), and area under the precision-recall curve (AUPRC). Figure 2 presents the ROC curve which shows the model's classification performance at different thresholds and also confirms the great discriminative capacity of the proposed approach.

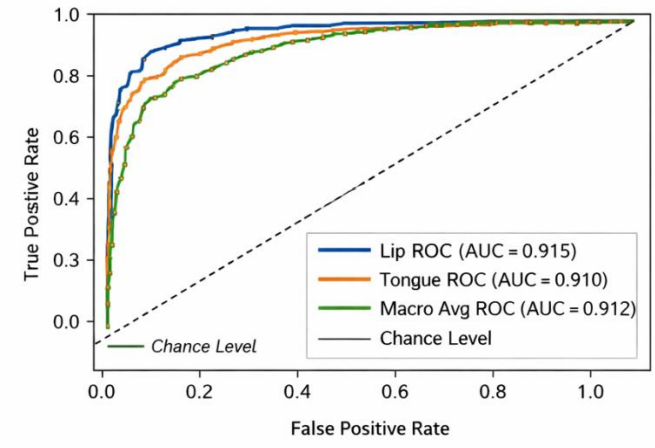

Figure 2: ROC curve of the proposed model.

Experimental results indicate that our developed framework reached an accuracy of 0.892, a macro F1-score of 0.883, an AUROC of 0.912, and AUPRC of 0.884 on the test dataset. This implies that the deep convolutional features coupled with texture-based representations possess a greater ability for detection than traditional single-branch models. Additionally, results demonstrate that our suggested

framework can very well offer an accurate automated approach to oral cancer screening using photographic images and can also serve as a computer-aided diagnostic tool for the early detection of oral cancer.

Discussions

Our experiments showed that the dual-branch scheme in our proposal was able to significantly boost the accuracy of automated oral cancer detection from lip and tongue images. We did this by merging CNN features with texture-based descriptors. As a result, the model was able to detect not only the major structural patterns but also the minute surface features of the malignant lesions. The ROC curve analysis indicates a quite good separation of the classes at different operating points. Besides that, probability calibration modifies the prediction confidence scores in such a way that they present a much better match to the actual probabilities, which is an extremely good feature for clinical decision support systems. Moreover, the feature fusion using the attention mechanism-based has helped a lot in the system's ability to be less sensitive to changes in lighting, pose, and imaging conditions. Taken together, the proposed method has the potential to assist computer-aided oral cancer screening and, what is more, the early detection through telemedicine.

Conclusions

1. 1. Problem Statement / Motivation: This study aimed to detect oral cancer making use of images of lip and tongue. Manual diagnosis of cancer in this area becomes very difficult due to the different and small features of lesions here that can be seen visually.
2. Methods Used: Three primary approaches combined their strengths to achieve the state-of-the-art results more reliably in time series classification. We first introduced attention-based dual feature fusion, which gathers and fuses CNN features and hand-crafted features; secondly, we performed probability calibration to re-estimate the final probability of the prediction for better prediction reliability.
3. Key Findings: Comparative evaluation of the experiments conducted on two different datasets proved the effectiveness of the proposed framework which not only achieved the best classification performance but also be interpretable at the same time.
4. Limitations and Future Work: The research focus has currently been limited to a small dataset. Future studies will emphasize utilizing large datasets from multiple institutions, improving model interpretability, and creating lightweight models for mobile and telemedicine platforms.

References

1. M. A. Dávila Olivos, H. M. Herrera Del Águila and F. M. Santos López, "Diagnosis of oral cancer using deep learning algorithms", Ingenius, no. 32, pp. 58–68, 2024. https://doi.org/10.17163/ings.n32.2024.06
2. P. Liu and K. Bagi, "A tailored deep learning approach for early detection of oral cancer using a 19-layer CNN on clinical lip and tongue images", Scientific Reports, vol. 15, art. no. 23851, 2025. https://doi.org/10.1038/s41598-025-07957-9
3. S. H. Begum and P. Vidyullatha, "Automatic detection and classification of oral cancer from photographic images using attention maps and deep learning", International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no. 11s, pp. 221–229, 2023. https://doi.org/10.18201/ijisae.2023.11.11s.3464