RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt26110

OralCANet: Transformer-Based Oral Cancer Detection using Q-Layered Classification

Dasyam Chandra Mouli [a, b, *], Pullela SVVSR Kumar [c], Dasari Haritha [a]

[a] Department of Computer Science and Engineering, University College of Engineering Kakinada, Jawaharlal Nehru Technological University, Kakinada-533003, Andhra Pradesh, India

[b] Vignan's Institute of Information Technology, Vizag, Andhra Pradesh, India

[c] Department of Computer Science and Engineering, Aditya University, Surampalem, Kakinada-533437, Andhra Pradesh, India

* Corresponding Author Email: mouli227@vignaniit.edu.in

DOI: https://doi.org/10.54392/irjmt26110

Received: 04-09-2025; Revised: 16-01-2026; Accepted: 21-01-2026; Published: 29-01-2026

Abstract: Oral cancer is a problem for people all around the world. The thing is, doctors need to find cancer and bad cells. So, people can get better. Usually, doctors look at the mouth and take a sample of the bad cells. This way of doing things is not very good because it is based on what the doctor thinks and it takes a long time. Sometimes doctors make mistakes. Deep Learning is a way of doing things that seems to be working. The old models that use Deep Learning have some problems. They get confused by information and they have trouble in understanding complicated things in pictures of sick cells. Oral cancer and the pictures of cells are what make this hard for Deep Learning to figure it out. This research is trying to fix some problems so, it suggests a technique called OralCANet. This is a kind of technique that has many layers and which processes huge data to retrieve the information in better and accurate way. It uses CNN model called ResNet-101 which is already known for it's functionality that helps in finding the require features like Wavelet Transform and Butterworth Filtering to make the information clearer. To find the parts, it uses a few different methods together like DeepLabV3+, GrabCut and Mask R-CNN to make sure it gets the right areas. The main thing that makes OralCANet special is the Q-Layered classification part. This part is unique because it is the combination of different techniques which produces better result. It uses a Transformer Encoder to look at the picture and finds features that are important everywhere. It also uses a BiLSTM to look at how things are connected in a sequence. It uses an Attention-based Deep Feature Encoding Layer to focus on the features that are bad. OralCANet model was tested on two data sets. One set was of tissue samples. The other set was of mouth pictures. OralCANet did a job and was right 99 percent of the time. It even did better, than models that are currently the best. These results suggest that the Q-Layered integration offers a robust, high-precision tool for automated oral cancer screening, potentially assisting clinicians in making faster, more accurate diagnosis.

Keywords: Oral Cancer, Q-Layered Architecture, Transformer Encoder, BiLSTM, Attention Mechanism, ResNet-101, Medical Image Analysis.

1. Introduction

Oral Cancer, primarily Oral Squamous Cell Carcinoma (OSCC), remains one of the leading challenges in global health and is a major contributor to cancer-related deaths around the world. Oral cancers are defined by the abnormal growth of lesions in the oral cavity, which consists of the lips, tongue, gums, and throat. Oral cancer is caused by several factors. The most common causes are long-term tobacco use, smoking, and high amounts of alcohol consumption and having poor oral hygiene. Studies conducted recently have also revealed that Human Papillomavirus (HPV) and family history of genetic conditions can be some of the most common causes of the disease [1, 2]. In addition, the prognosis for oral cancer is closely related to the stage at which the patient is diagnosed; when precancerous and malignant lesions are detected early enough, the chances of survival and response to treatment will greatly improve. The problem is that early-stage lesions are difficult to detect because they often don't show any symptoms and can look very similar to benign conditions [3, 4]. Traditional methods of diagnosis (i.e., visual examination by a clinician and invasive biopsy of tissue) are frequently accepted as the gold standard for the diagnosis of oral cancer; however, they carry various limitations. Firstly, these procedures take a long time to complete, and they rely on the experience of the clinician performing the procedure, and this makes it difficult to assess the condition reliably

by all clinicians and causes differences between observers and results [5, 6].

In order to overcome diagnostic limitations referred to as “bottlenecks”, there is a growing trend towards utilizing Artificial Intelligence (AI) and Deep Learning (DL) techniques as solutions for the field of medical research. Automated procedures that can assess medical images taken with clinical photography, radiographic scans and histopathology enable efficient and precise identification of pathology. Convolutional Neural Networks (CNNs), which are a particular type of artificial neural network used in deep learning, are becoming established as the most commonly utilized architectures when performing evaluations of medical images, due to their effectiveness for classification and segmentation tasks associated with malignant tumors [7, 8].

In recent years, various studies have successfully demonstrated that alternative versions of CNN versions can provide effective means to differentiate between normal, benign, and malignant tissues with promising results [9, 10]. For example, hybrid AI solutions with an ensemble of models using CNNs have improved the diagnostic accuracy of these models by using large collections of histological images as training datasets [5, 11]. Nevertheless, despite these types of development and spreads of opportunity for this classification/segmentation approach for pathology diagnostic applications, although still in their infancy, there continue to remain many obstacles to be overcome. Generally speaking, traditional CNN architecture is often incapable of performing consistently in medical images due to their high level of intra and inter variability caused by external variables that may obscure important pathological characteristics [12]. Conversely, to date, CNNs appear to excel at extracting local spatial features from images, such as edge detection and texture measurements, whereas they often do not account for the long-range global dependency of multiple areas of tissue or for the contextual relationships that may exist between these different areas, which are critical for accurate classification or grading of tumor malignancy [13].

The need to study the complexities of oral cancer pathology, therefore, requires a more advanced analytical pipeline than what simple image classification can provide. Many of the standard "black-box" DL models often suffer from overfitting when it is applied on small or noisy datasets due to a lack of interpretability on feedback from clinical observations: Clinical interpretation needs are stringently met by other architectural designs that rely heavily on precise lesion segmentation, noise reduction, and complex-feature integration [14, 15]. Many in the field have raised issues with complex architectures being computationally redundant; with simpler pipelines never able to generalize on unseen clinical datasets where image quality varies a lot. With this being said, this approach proposed within the study is a multi-stage approach integrating state-of-the-art image-processing methods and a novel DL design. It will be argued on two important fronts: (1) Pre-processing using the Wavelet Transform and Butterworth Filtering has been adopted to suppress high-frequency noise artifacts and augmentation of structural details [16, 17] (2) Accurate segmentation through the combined use of DeepLabV3+, GrabCut, and Mask R-CNN has been developed to implement the ROI, so as to eliminate confusing background objects [18, 19] and instead a new classification system that pushes the limits beyond traditional feature extraction.

In this study, proposed OralCANet, a new type of method that uses the Q-Layered classification. With most models, features are typically combined via concatenation. However, with the Q-Layered system, three different sets of features can work together in a way that maximizes the number of features used to create an integrated feature representation. These three components include: 1) a Transformer Encoder, which uses the attention mechanism to enable the model to capture both the spatial dependencies of the features and their long-term dependencies over the entire image; 2) the use of Bi-directional Long Short Term Memory networks for sequence modeling, which enables the capture of how features are related to one another and how those relationships change with time; and 3) a Deep Feature Encoding layer using an Attention Mechanism, which dynamically weighs the features that were extracted, allowing the model to focus on the patterns associated with Malignancy while ignoring irrelevant features. Because of this architecture, we are able to create a model that is highly sensitive to small precancerous changes, while still being robust to false positives.

2. Literature Survey

Vayadande *et al.* [13] proposed ML and DL techniques implemented with AI for the detection of skin and oral cancer. Different methodologies such as CNNs, Transfer Learning, Hybrid models, and ADM have been evaluated in the context of their utility in medical imaging and histopathological analysis. Finally, the proposed approach achieved a high degree of accuracy (>90%) in differentiating malignant and benign lesions in both skin and oral cancers. Hemalatha *et al.* [14] proposed Fragment Jaya Whale Optimizer with Deep Convolutional Neural Network (FJWO-DCNN) model that detects oral cancer with classification. In this work, the theoretical features were extracted from the input image, which helps for better classification. Finally, the proposed FJWO-DCNN produced superior performance. Jagadesh *et al.* [15] developed the Particle Swarm Optimization and AI-Biruni Earth Radius (PSOBER) to identify oral cancer early on. To enhance the segmentation, Salt and Pepper noise detection is employed. Lastly, the classification results display the

tumor partitioning and feature extraction performance. Akbar *et al*. [20] introduced the hybrid algorithm that combines with various DL algorithms. In this context, the feature vectors used to create the compact vector, which significantly influences the results. The suggested method achieves 97.80%, 95.13%, 93.91% and 94.17% accuracy for several datasets. Sharma *et al.* [21] presented the new RNN-CNN model to detect the skin cancer cells. The proposed RNN-CNN mainly divides the skin cells based on melanoma and focal carcinoma. Finally, the detection and classification show 98.34% accuracy. Previously, it was 85.34%, which is low for the existing CNN approach. Huang *et al*. [22] presented a novel approach that improve the CNN version for detecting and classifying oral cancer detection. The combination of Seagull Optimization (SO) and Particle Swarm Optimization (PSO) algorithms shows a superior accuracy of 96.94% compared with other models. The different performance metrics include precision of 94.56%, recall of 91.66%, and F1S of 88.65%. The main drawback of this approach is the poor accuracy in region detection. Haq *et al*. [23] presented the FPSO-CNN with feature extraction based on Transformer, which improves diagnostic accuracy. For training and evaluating the model, oral lesion image dataset is considered, which employs transfer learning and attention mechanisms for learning microscopic features with high discrimination abilities in classification among malignant and benign lesions. Ultimately, you will get the result as accuracy−94.8%, specificity (SPC): 97.2% and F1-score: 95%. Shah *et al.* [24] introduced an automated oral cancer detection that finds and classifies the premalignant lesions from the given MRI images. Used MATLAB for image preprocessing, highlighting the dissimilarity among normal and abnormal images through high red values. The GLCM is used to extract the features needed for the final step. Proposed approach obtained a decision value of 89.12%. Feyzullah Temurtas *et al*. [25] presented a detection model for thyroid cancer using ML algorithms. The model contributes to the pathology of thyrotoxicosis. Thus, the proposed theory describes a highly effective model that identifies both the vaunted state and the polygenic alterations affecting the shape and composition of the thyroid. Lastly, obtained results of the method will be observed and displayed in terms of change in classification rate. Shanmuga Sundari et al. [26] presented a novel DL approach, XceptionAttnV1, that increases the cancer detection rate. In Pre-processing CLAHE is used for image enhancement in feature representation and contrast of the image. Proposed approach has implemented, but overfitting enhances the translation performance. However, it also leans more towards the data and the overfitting models. The proposed approach's accuracy, F1-score, recall, specificity, and precision are 96.59%, 96.60%, 96.60%, 96.46% and 96.60%, respectively. The literature review is further given in Table 1.

Table 1. Summary of Literature review

Author (Year)	Method/ Model	Data set	Key Findings	Limitations
Vayadande *et al.* [13] (2024)	CNN + Transfer Learning	Skin & Oral Cancer images	Accuracy > 90% in distinguishing malignant/ benign.	Computationally intensive; limited to binary classification.
Hemalatha *et al.* [14] (2022)	FJWO-DCNN (Fragment Jaya Whale Opt.)	Oral Cancer Data set	Superior performance in theoretical feature extraction.	Segmentation accuracy was not the primary focus.
Haq *et al.* [23] (2023)	FPSO-CNN + Transformer	Oral lesion images	Acc: 94.8%, Specificity: 97.2%.	High specificity but requires large training data.
Proposed Method	OralCANet (Q-Layered)	DS1 & DS2	Acc: 99%, F1: 99%. Integrates Segmentation & Attention.	High computational cost due to multi-stage pipeline.

3. Pre-trained ResMed-101 Model with Transfer Learning

DL architectures were used in medical image classification, particularly with CNNs exhibiting superior performance. Deep Residual networks (ResNets) are the new architectures that can learn more complex hierarchical features while reducing the vanishing gradient problem. It is an extensive variant of the ResNet family which is widely used in image classification and has proven effective in applications like medical imaging [27]. To detect oral cancer a Pre-trained ResNet101 model using Transfer learning has implemented in this study. Transfer learning allows a model trained on a large dataset (ImageNet) to adapt to a particular medical imaging task with less available labeled data. ResNet-101 contains 101-layer DRN classifies the cancerous and non-cancerous categories.

The proposed method improves the efficiency of the model in discriminating malignant and non-malignant oral lesions by fine-tuning the top layers of ResNet-101 while preserving the information captured in features obtained from natural images [28].The initial layers of ResNet-101 capture the advanced image features such as textures and edges. Final classification layer used with the fine-tuned deep layers.

The mathematical model of the proposed approach is represented as:

Step 1: ResNet-101 Forward Propagation is represented as:

$$y = F(x, \{W_i\}) + x \quad (1)$$

Were, a-input; $\mathrm{F(a,\{W_i\})}$- residual mapping; b-output.

Step 2: Optimization Algorithm: In this step, the Adam optimizer used for update and it is represented as,

$$\theta_{(t+1)} = \theta_t - \eta\left(\widehat{m_t}/\left(\sqrt{(\widehat{v_t})} + \epsilon\right)\right) \quad (2)$$

Step 3: Fine-Tuning with Loss Function: In this step, the Cross-Entropy is implemented for classification task and it is represented as,

$$L = -\sum_{i=1}^{C} y_i \, log(p_i) \quad (3)$$

Step 4: Gradients Update: The two layers were updated by measuring the gradients only for θ_{new}, and fix the θ_{base} constantly.

$$\theta_{(\mathrm{new})} \leftarrow \theta_{(\mathrm{new})} - \eta\left(\partial \mathrm{L}/\,\partial\theta_{(\mathrm{new})}\right) \quad (4)$$

Step 5: Transfer learning Update: The fine-tuning applied for last two layers by initializing the frozen layers to obtain the pre-trained weights.

Let $\mathrm{f(a;\,\theta)}$ represents the model functions and weights; θ_{base} represents the pre-trained weights and θ_{new} represents the trainable parameters.

$$f(a;\theta) = g\left(h\left(a;\theta_{(base)}\right);\theta_{(new)}\right) \quad (5)$$

$\mathrm{h(a;\,\theta_{base})}$ Represents the feature extrcation from frozen layers. $\mathrm{g(\theta_{new})}$ Represents the new trainable layers.

Fine-tuning the top layers of ResNet-101 using Transfer Learning involves loading a pre-trained ResNet-101 model and freezing the lower layers representing general features such as edges and textures. The final layers that capture high-level task-specific features are replaced with a new classifier-for example, a fully connected layer with softmax activation for classification. We perform the training of the model via cross-entropy loss, updating just with respect to the new classifier θ_{new} and not updating the backbone weights θ_{base}. In training, the data used to obtain the predicted values, measures the loss and backpropagates the trainable layers through an optimizer like SGD or Adam if required. A new model takes advantage of the learned features of the ResNet-101 and applies it to the new dataset, allowing greater convergence with drastically less training.

4. Pre-processing Techniques-Wavelet Transform and Butterworth Filtering

In this section, the role of pre-processing techniques was conversed which plays a major role in filtering the noise and improves the feature selection and classification. Here, the Wavelet Transform is implemented first and then Butterworth Filtering is implemented.

4.1 Wavelet Transform

Wavelet Transform (WT) is a powerful model for Medical Image Analysis in the Detection of Oral Cancer. Traditional medical imaging is essential for oral cancer diagnosis, while noise, blurred contrast and useless background information in images make the deep learning-based detection model inaccurate. Wavelet Transforms (WTs) is a crucial mathematical model and preprocessing technique to address such issues like improving image quality and enhancing feature extraction for classification and segmentation, which facilitates in detection of oral cancer [16].

4.2 Butterworth Filtering (BF)

Butterworth Filtering (BF) is a noise reduction model that filters noise to improve the quality of an image [17]. It is the most significant filter that retains the required features from the input images. This model primarily focused on suppressing high-frequency noise and improving the contrast of the input image. It mainly relies on high-frequency noise removal while preserving critical structures, such as texture and edges in tumor areas. Using BF highlights the abnormal regions and makes anomalies look more conspicuous to deep learning-based diagnostic models.

Furthermore, this filtering in the spatial frequency domain allows for a smooth transition between frequencies. Hence, hybrid features result in misinterpretation of the output images, which are not generated. It effectively increases segmentation, feature extraction and classification accuracy, thereby enhancing the reliability of automated oral cancer detection systems.

5. An Ensemble Segmentation

Accurate segmentation of cancerous lesions is challenging because of lighting, texture variation and differences in the morphology of lesions. Here presented an ensemble segmentation approach in which DeepLabV3+ is used in conjunction with background removal by GrabCut and Mask R-CNN to improve the accuracy of the segmentation of oral cancer lesions. DeepLabV3+ is a traditional DL model with strong

semantic segmentation performance [18]. GrabCut is mainly used for foreground extraction based on graph cuts. Mask R-CNN is a powerful model for instance segmentation [19]. The main aim of this is to filter the noise while enhancing both the accuracy and contrast of the segmentation of the lesions. Proposed architecture consists of several stages. In the first stage, the background is removed using GrabCut and Mask R-CNN, retaining only the region of interest (ROI) and excluding irrelevant background details. Next, DeepLabV3+ is used to segment the detected lesion accurately. The combination of the benefits of each technique such as DeepLabV3+ for deep feature extraction, GrabCut for edge refinement and Mask R-CNN for instance-level segmentation improved the segmentation accuracy. This work will contribute to the reliability of automated oral cancer detection by improving segmentation accuracy, which would help better in clinical diagnosis and treatment planning by doctors. Finally, using experimental analysis, the performance of Ensemble Segmentation was compared with existing approaches applied to benchmark dataset images.

6. Q-Layered: A Novel Classification Approach

The Q-Layered model is the integrated approach combined with Transformer Encoder, BiLSTM and an Attention-based Deep Feature Encoding Layer for effective feature learning in Oral Cancer Detection. First, high-quality medical images, such as histopathological or fluorescence images are obtained and then they are preprocessed through noise reduction, normalization, augmentation etc., which is done to maximize the generalizing ability of a model. In some situations, segmentation techniques (such as DeepLabV3+) can be utilized to isolate the cancerous regions. After preprocessing the images, the Transformer Encoder obtains deep spatial features by splitting the images into several patches, converting the patches to embedding vectors and applying self-attention layers that capture the global dependencies of the patches. The extracted features are then fed into a Bidirectional long short-term memory (BiLSTM) network to improve the feature learning of sequences while retaining the spatial relationship and contextual information from frames, in addition to both the previous and next frames. An Attention-based Deep Feature Encoding Layer then places more value on significant areas of the image, allowing the model to concentrate on meaningful cancerous features while minimizing distractions from the background. A fully connected neural network classifies the refined feature representation through a softmax layer that predicts per class probability (Normal, Pre-cancerous or Cancerous) that can be optimized using cross-entropy loss. Explainable AI (XAI) post-processing techniques based on the Grad-CAM algorithm are usually employed to give visual justification for predictions, therefore increasing the interpretability of the model from the clinician's standpoint. Furthermore, it can be included in CAD frameworks to assist physicians in better decision making. Redesigning of the Q-Layered approach allows for greater accuracy, powerful computational feature representation and increased configurational translation thereby providing a novel deep-learning framework would be well suited for oral cancer detection. The proposed Q-Layered Architecture (Figure1) effectively integrates hierarchical feature extraction with adaptive learning mechanisms, enabling improved discrimination of complex oral lesion patterns and supporting robust diagnostic performance.

Step 1: Transformer Encoder (TE): The TE is the initial Step for feature extraction of input image:

The input image belongs to oral cancer $A \in R^{H\times W\times C}$ converts into patch embeddings and it is measured as:

$$A_p = f_{(\text{patch})}(A) \in R^{(N\times d)} \quad (6)$$

Where: A_p represents the patch embeddings,

N represents the total patches; d represents the embedding dimension.

The self-attention method helped to obtain the global relationships:

$$C = A_P W_C, D = A_P W_D, E = A_P W_E \quad (7)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left((QK^T)/\sqrt{(d_k)}\right)V \quad (8)$$

Where W_C, W_D, W_E represents the weights:

The output of the final transformer A_T is:

$$X_T = LayerNorm\left(A_p + Self - Attention(C, D, E)\right) \quad (9)$$

Step 2: BiLSTM used for consecutive Feature Learning the output transformer A_T is transmitted to a BILSTM for the learning of consecutive patterns:

$$h_t^{(\rightarrow)} = \sigma\left(W_f A_T^t + U_f h_{(t-1)}^{(\rightarrow)} + b_f\right) \quad (10)$$

$$h_t^{\leftarrow} = \sigma\left(W_b A_{T_t} + U_b h_{t-1}^{\rightarrow} + b_b\right) \quad (11)$$

$$H_t = [h_t^{\rightarrow}; h_t^{\leftarrow}] \quad (12)$$

Where

$h_t^{\rightarrow}$ and $h_t^{\leftarrow}$ represents the forward and backward LSTM hidden states, H_t concatenets the final output of BiLSTM

Step 3: Attention-Based Deep Feature Encoding Layer In this step, the attention mechanism initializes the significant weights to BiLSTM outputs:

$$\alpha_t = \frac{exp(W_\alpha H_t)}{\sum_j exp(W_\alpha H_j)} \quad (13)$$

$$F_{encoded} = \sum_t \alpha_t H_t \quad (14)$$

Step 4: Final classification layer for Oral Cancer Detection The final classification is obtained by using the softmax layer:

$$\hat{b} = softmax(W_O F_{encoded} + b_O) \quad (15)$$

Where: W_O and b_O represents the final parameters, $\hat{b}$ represents the final prediction over oral cancer classes.

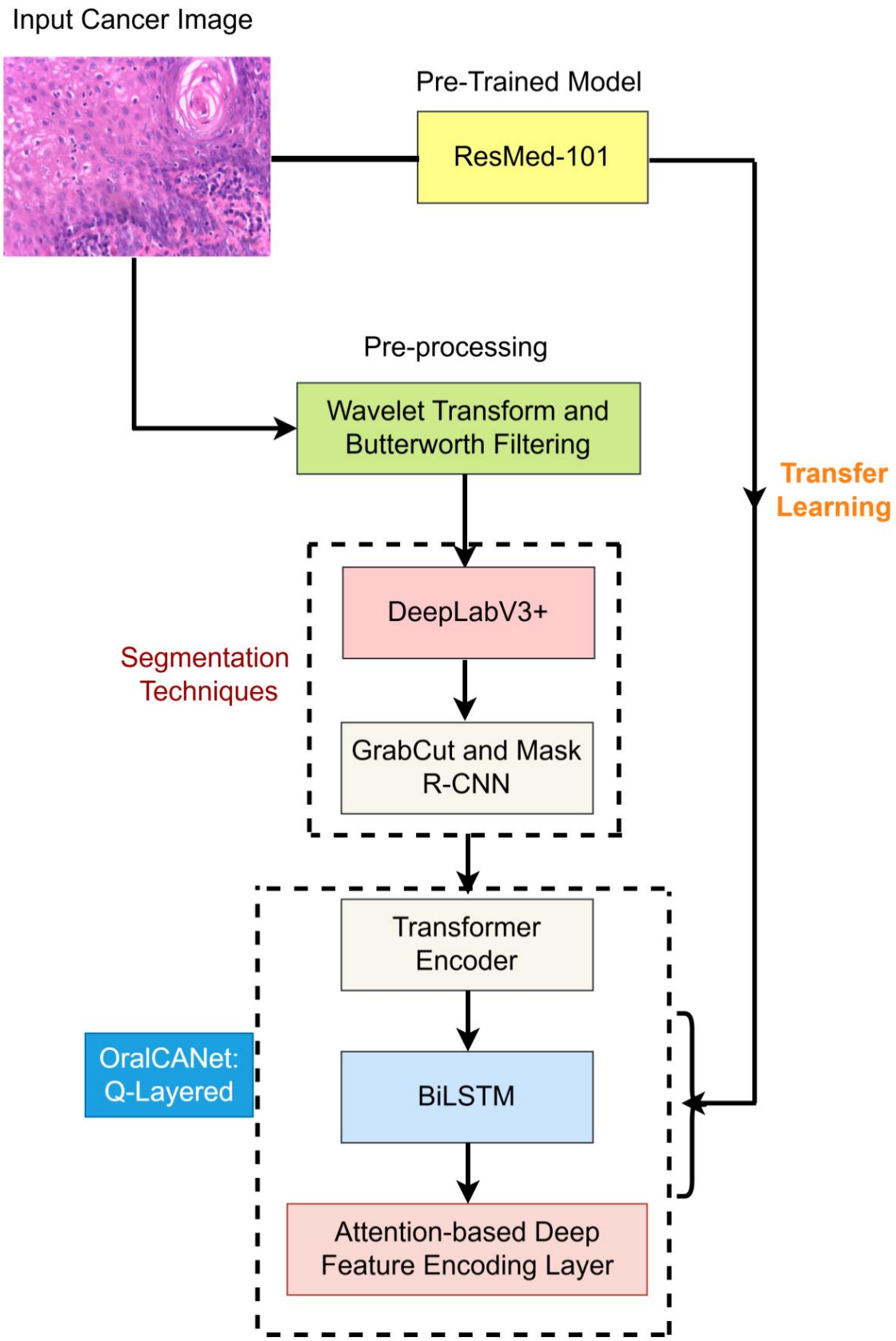

Figure 1. Q-Layered Architecture

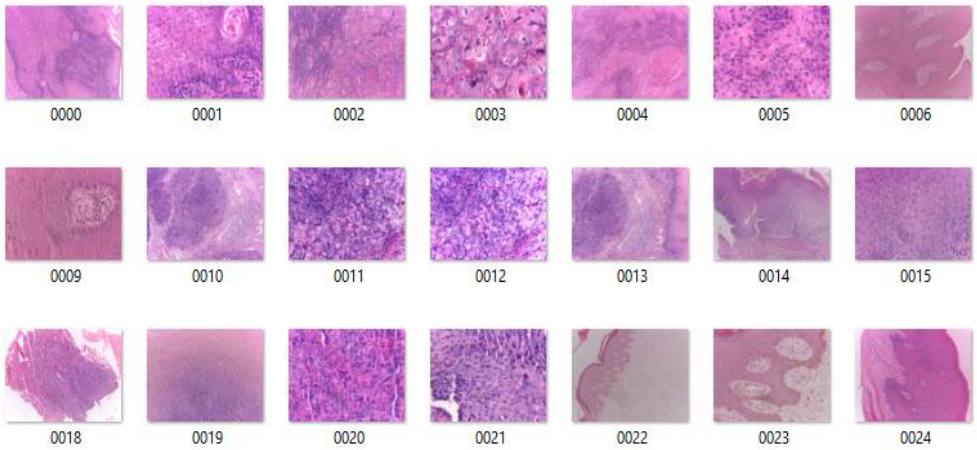

Figure 2 Sample Oral Cancer Images belongs to Dataset1 (DS1).

Figure 3. Sample Oral Cancer Images belongs to Dataset2 (DS2).

7. Dataset Description

The experiments were conducted using two publicly available oral cancer datasets. Dataset-1 consists of histopathological image patches obtained from the Oral Diagnosis Project (NDB) of the Federal University of Espírito Santo (UFES), Brazil [32], comprising 178 training images and 100 testing images. Overall dataset distribution used for training and testing is illustrated in Figure 2. Dataset-2 was collected from an open-source oral cancer repository and includes 1231 training samples and 1000 testing samples, of which 750 correspond to cancerous images and 250 to non-cancerous images. Representative sample images from this dataset are shown in Figure 3.

8. Results and Discussions

The metrics most efficiently indicate the strength of the algorithms performed on Oral cancer datasets. The confusion matrix plays the major role in binary classification. In this context, the count sample values are obtained from the algorithm implementation. The algorithm receives the predicted values, and every image has label values. The following parameters measures the performance of algorithms:

$$Sensitivity\ (Sn) = \frac{TP}{TP+FN} \quad (16)$$

$$Specificity\ (Sp) = \frac{TN}{TN+FP} \quad (17)$$

$$Precision\ (P) = \frac{TP}{TP+FP} \quad (18)$$

$$Accuracy\ (Acc) = \frac{TP+TN}{TP+FP+TN+FN} \quad (19)$$

$$F1S = 2 * \frac{P*S_n}{P+S_n} \quad (20)$$

The comparisons between different ML algorithms and the proposed Q-layered algorithm are based on the parameters above. The performance of different ML classifiers applied to the data set (DS1) without any pre-processing, segmentation on the different key performance metrics even without pre-processing for DS1 and segmentation is shown in Table 1: the NB results in poor performance in terms of Acc (0.5691) and F1S (0.5621) which indicates the existence of high levels of misclassification. This small increase

was likely due to SVM, which reached an Accuracy of 0.65 and an F1 score of 0.64. The other ML approach, LightGBM, outperformed quite dramatically on Acc (0.895) and F1S (0.906), showing how effectively boosting these models is even without pre-processing. The Q-layered model proposed outperformed all other approaches by obtaining the highest accuracy measure of 0.93 and the highest F1-score of 0.931, indicating a better predicting capability. These results suggest that traditional ML models, including NB and SVM, are adversely affected if pre-processing is not applied. Still, it supports the fact that ensemble-based models, including LightGBM and the Q-Layered Model, have a strong classification performance.

The ML algorithms perform quite differently on DS2 without preprocessing and segmentation (with cross-validation) as shown in Table 2. NB outperforms the rest with the highest Se (0.888), Sp (0.876), and Acc (0.882) scores, while NB suffers the most with low Se (0.56), Sp (0.563) and ACC (0.5612), suggesting that it does not possess a good classification capability at all. SVM shows moderate improvement, better Pre (0.651) and ACC (0.648), but still lags without data preprocessing. LightGBM beats these models extremely well with a good Acc of 0.921 and an F1S of 0.916 due to its gradient boosting technique. Q-Layered Model outperforms all other models, with Acc (0.956) and F1S (0.959) indicating it has the best classification ability. This idea indicates that while traditional models with no preprocessing cannot interpret inputs, LightGBM and Q-Layered Models can handle raw data without preprocessing.

Table 3 explains the comparative models, such as NB and SVM, are not good at multi-class classification, as the Sn, Pre, and F1S are lower. SVM performs the weakest among the four models, with only 0.312 F1S, which illustrates that the SVM model is not well-balanced for an imbalanced class. On the other hand, LightGBM demonstrates strong performance, achieving an accuracy of 0.98 across all evaluation metrics. Exceptionally, the proposed Q-layered model outperforms all other models by attaining near-perfect scores of 0.99, highlighting its superior capability to handle complex patterns while maintaining high classification reliability. These results clearly indicate that advanced learning models outperform traditional approaches, as summarized in Table 4.

Table 2. Comparison of ML algorithms without pre-processing and segmentation for DS1

	Sn	**Sp**	**P**	**Acc**	**F1S**
NB	0.58	0.57	0.56	0.5691	0.5621
SVM	0.64	0.64	0.64	0.65	0.64
LightGBM [30]	0.89	0.907	0.905	0.895	0.906
Q-Layered Model	0.93	0.924	0.941	0.93	0.931

Table 3. Comparison of ML algorithms without pre-processing and segmentation for DS2

	Sn	**Sp**	**P**	**Acc**	**F1S**
NB	0.56	0.563	0.571	0.5612	0.578
SVM	0.613	0.626	0.651	0.648	0.638
LightGBM	0.884	0.891	0.901	0.921	0.916
Q-Layered Model	0.934	0.945	0.951	0.956	0.959

Table 4. Evolution of ML algorithms with multi-class classification DS1

	Sn	**Sp**	**P**	**Acc**	**F1S**
NB	0.527	0.76	0.555	0.711	0.532
SVM	0.356	0.677	0.281	0.60	0.312
LightGBM [30]	0.98	0.98	0.98	0.98	0.98
Q-Layered Model	0.99	0.99	0.99	0.99	0.99

Table 5. Evolution of ML algorithms with multi-class classification DS2

	Sn	Sp	P	Acc	F1S
NB	0.531	0.77	0.56	0.722	0.54
SVM	0.37	0.656	0.291	0.62	0.321
LightGBM [30]	0.98	0.97	0.98	0.98	0.98
Q-Layered Model	0.99	0.99	0.99	0.99	0.99

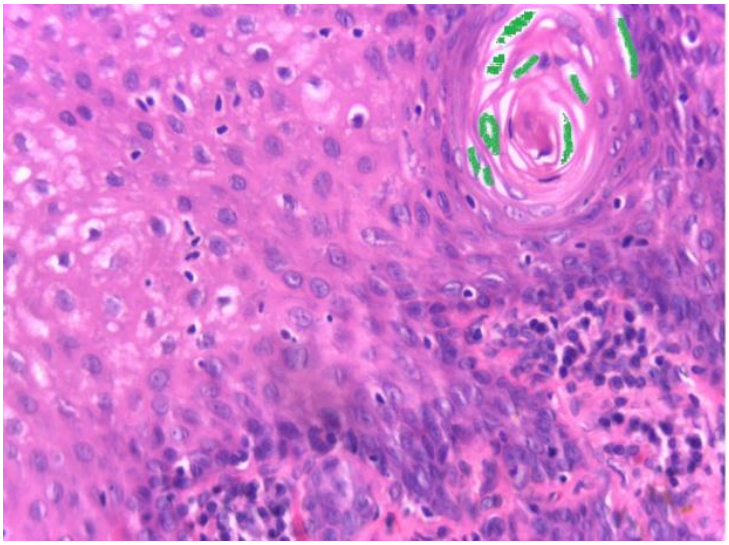

Figure 4. Final Output for Oral Cancer Detection DS1.

Figure 5. Final Output for Oral Cancer Detection DS2.

For the multi-class classification (DS2), if we analyze the performance of the different ML algorithms, SVM underperforms compared to XGBoost and GCN, which achieved a good Sn (0.37) and Pre (0.291) as well as good Acc (0.62) explained in Table 5. NB performs better than SVM, displaying only moderate Sn (0.531), Sp (0.77) and Acc (0.722) but poor Pre (0.56). In all metrics, LightGBM highly improves both of those (0.98), indicating a good generalization. The Q-Layered Model improved upon LightGBM and achieved almost perfect classification (0.99 across all metrics), making it the best-performing model for DS2. Performance of the proposed OralCANet is graphically illustrated in Figure 4, where it clearly surpasses traditional CNN architectures, such as standard ResNet-101 as well as recent hybrid models based on the experimental results. The superior accuracy of 0.99 can be attributed to the effective cooperation enabled by the Q-Layered architecture. While conventional CNNs perform well in extracting local features such as edges and textures, they often struggle to capture global contextual information related to tissue morphology. In contrast, the proposed model leverages a Transformer Encoder to embed long-range dependencies, enabling a

comprehensive understanding of spatial relationships among cells across the entire image patch. From a clinical perspective, the high sensitivity value of 0.99 indicates that OralCANet has strong potential as a reliable screening tool, significantly reducing the risk of false-negative diagnosis an essential requirement in cancer detection. However, despite its promising performance, the deployment of the model presents challenges due to the complexity of its processing pipeline. The multi-stage framework, which integrates wavelet transformation, ensemble segmentation and the Q-Layer, demands substantial computational resources when compared to lightweight architectures such as MobileNet. As illustrated in Figure 5, this computational overhead may limit immediate deployment on edge devices unless extensive optimization strategies, such as model pruning are employed.

9. Conclusion

This study deals with OralCANet, a methodology that will help doctors to locate or detect oral cancer. A new system has been created, developed and formulated by a research team, which employs computers in looking at the images. It does jobs better than other manual methods as it examines the images in such a way. It originated from the initial brainstorming sessions, where the inventors of OralCANet conceived the method of facilitating a computer in seeing. This technique is called the Q-Layered classification block. OralCANet goes through the detail and looks at the big picture in locating oral cancer inside the matter of a minute. This integration of the Transformer Encoder with BiLSTM and Attention provided a more discriminative representation for cancerous lesions and resulted in state-of-the-art performance metrics on each of the two diverse datasets (Accuracy, Sensitivity, and Specificity > 99%). Several issues arise in the course of this study. The datasets used in this study, DS1 and DS2, can show how well the model works, but they are not huge. Hence, we are less certain if the model should work over datasets with generally diversified features from many places.The process of using the model has steps and this can cause delays when we try to use the model on small medical devices that people carry with them like mobile medical devices because it needs a lot of computing power. Future work will focus on Validating the model on larger, more diverse external cohorts to ensure robustness.

References

[1] M.Z.M. Shamim, S. Syed, M. Shiblee, M. Usman, S.J. Ali, H.S. Hussein, M. Farrag, Automated Detection of Oral Pre-Cancerous Tongue Lesions using Deep Learning for Early Diagnosis of Oral Cavity Cancer. The Computer Journal, 65(1), (2020) 91–104. https://doi.org/10.1093/comjnl/bxaa136

[2] H. Myriam, A.A. Abdelhamid, E.M. El-Kenawy, A. Ibrahim, M.M. Eid, M.M. Jamjoom, D.S. Khafaga. Advanced Meta-Heuristic Algorithm based on Particle Swarm and Al-Biruni Earth Radius Optimization Methods for Oral Cancer Detection. IEEE Access, IEEE, 11, (2023) 23681–23700. https://doi.org/10.1109/access.2023.3253430

[3] R.A. Welikala, P. Remagnino, J.H. Lim, C.S. Chan, S. Rajendran, T.G. Kallarakkal, R.B. Zain, R.D. Jayasinghe, J. Rimal, A.R. Kerr, R., Amtha, K. Patil, W.M. Tilakaratne, J. Gibson, S.C. Cheong, S.A. Barman, Automated Detection and Classification of Oral Lesions using Deep Learning for early detection of oral cancer. IEEE Access, 8, (2020) 132677–132693. https://doi.org/10.1109/access.2020.3010180

[4] G.A.I. Devindi, D.M.D.R. Dissanayake, S.N. Liyanage, F.B.A.H. Francis, M.B.D. Pavithya, N.S. Piyarathne, P.V.K.S. Hettiarachchi, R.M.S.G.K. Rasnayaka, R.D. Jayasinghe, R.G. Ragel, I. Nawinne, Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted early Detection of Oral Cancer. IEEE Access, IEEE, 12, (2024) 124375–124390. https://doi.org/10.1109/access.2024.3454338

[5] M. Das, R. Dash, S.K. Mishra, A.K. Dalai, An Ensemble Deep Learning Model for Oral Squamous Cell Carcinoma Detection using Histopathological Image Analysis. IEEE Access, 12, (2024) 127185–127197. https://doi.org/10.1109/access.2024.3450444

[6] P.M. Conforti, G. Lazzini, P. Russo, M. D'Acunto, Raman Spectroscopy and AI Applications in cancer Grading: an Overview. IEEE Access, 12, (2024) 54816–54852. https://doi.org/10.1109/access.2024.3388841

[7] M. Aharonu, L. Ramasamy, A Multi-Model Deep Learning Framework and Algorithms for Survival Rate Prediction of Lung Cancer Subtypes with Region of Interest using Histopathology Imagery. IEEE Access, 12, (2024) 155309–155329. https://doi.org/10.1109/access.2024.3484495

[8] R. Chavva, S. Jeba Priya, Mathu, Oral Cancer Detection Using Deep Learning. In 2024 International Conference on Science Technology Engineering and Management (ICSTEM), Coimbatore, India. https://doi.org/10.1109/ICSTEM61137.2024.10561172

[9] B. Goswami, M.K. Bhuyan, S. Alfarhood, M. Safran, Classification of Oral Cancer into Pre-Cancerous Stages from White Light Images Using LightGBM Algorithm. In IEEE Access, IEEE, 12, (2024) 31626-31639. https://doi.org/10.1109/ACCESS.2024.3370157

[10] P. Kalaivani, P. Iyyanar, C. Rajan, R. Harshini Priya, P. Janani, A.S. Jayasudha. (2024) Oral Cancer Detection Using Deep Learning. In 2024 2nd International Conference on Artificial

Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), Namakkal, India. https://doi.org/10.1109/AIMLA59606.2024.10531555

[11] I.U. Haq, M. Ahmed, M. Assam, Y.Y. Ghadi, A. Algarni, Unveiling the Future of Oral Squamous Cell Carcinoma Diagnosis: An Innovative Hybrid AI Approach for Accurate Histopathological Image Analysis. In IEEE Access, IEEE, 11, (2023) 118281-118290. https://doi.org/10.1109/ACCESS.2023.3326152

[12] Y. Xu, Y. Hong, X. Li, M. Hu, MedTrans: Intelligent Computing for Medical Diagnosis Using Multiscale Cross-Attention Vision Transformer. In IEEE Access, IEEE, 12, (2024) 146575-146586. https://doi.org/10.1109/ACCESS.2024.3450121

[13] K. Vayadande, T. Kamble, A. Padole, K. Mukkawar, S. Kurumbhatte, M. Patil. (2024) AI Driven Detection of Skin and Oral Cancer: A Survey of Machine Learning and Deep Learning Approaches. In 2024 International Conference on Sustainable Communication Networks and Application (ICSCNA), IEEE, Theni, India. https://doi.org/10.1109/ICSCNA63714.2024.10864094

[14] S. Hemalatha, N. Chidambararaj, R. Motupalli. (2022) Performance Evaluation of Oral Cancer Detection and Classification using Deep Learning Approach. In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE, Chennai, India. https://doi.org/10.1109/ACCAI53970.2022.9752505

[15] T. Jagadesh, P. Kamalesh, A. Kishore, V. Lokin, B. Jaiprakash. (2024) Oral Cancer Detection Using Convolutional Neural Networks. In 2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India. https://doi.org/10.1109/ICONSTEM60960.2024.10568599

[16] U. Sipai, R. Jadeja, N. Kothari, T. Trivedi, R. Mahadeva, S.P. Patole, Performance Evaluation of Discrete Wavelet Transform and Machine Learning Based Techniques for Classifying Power Quality Disturbances. In IEEE Access, IEEE, 12, (2024) 95472-95486. https://doi.org/10.1109/ACCESS.2024.3426039

[17] S.S. Daud, R. Sudirman, Butterworth Bandpass and Stationary Wavelet Transform Filter Comparison for Electroencephalography Signal. In 2015 6th international conference on Intelligent Systems, Modelling and Simulation, IEEE, (2015) 123-126.

[18] M. Zafar, J. Amin, M. Sharif, M. A. Anjum, G. A. Mallah, S. Kadry, DeepLabv3+-Based Segmentation and Best Features Selection Using Slime Mould Algorithm for Multi-Class Skin Lesion Classification. Mathematics, 11(2), (2023) 364. https://doi.org/10.3390/math11020364

[19] D.R. Jiménez, L. Casanova-Lozano, Sergi Grau-Carrión, R. Reig-Bolaño, Artificial Intelligence Methods for Diagnostic and Decision-Making Assistance in Chronic Wounds: A Systematic Review. Journal of Medical Systems, 49(1), (2025) 29. https://doi.org/10.1007/s10916-025-02153-8

[20] S. Akbar, A. Raza, T.A. Shloul, A. Ahmad, A. Saeed, Y.Y. Ghadi, O. Mamyrbayev, E. Tag-Eldin, PAtbP-EnC: Identifying Antitubercular Peptides using Multi-Feature Representation and Genetic Algorithm-based Deep Ensemble Model. IEEE Access, IEEE, 11, (2023) 137099–137114. https://doi.org/10.1109/ACCESS.2023.3321100

[21] A.K. Sharma, S. Tiwari, G. Aggarwal, N. Goenka, A. Kumar, P. Chakrabarti, T. Chakrabarti, R. Gono, Z. Leonowicz, M. Jasinski, Dermatologist-Level Classification of Skin Cancer using Cascaded Ensembling of Convolutional Neural Network and Handcrafted Features Based Deep Neural Network, IEEE Access, IEEE, 10, (2022) 17920–17932. https://doi.org/10.1109/ACCESS.2022.3149824

[22] Q. Huang, H. Ding, Navid Razmjooy, Oral Cancer Detection using Convolutional Neural Network Optimized by Combined Seagull Optimization Algorithm. Biomedical Signal Processing and Control, 87(Part B), (2024) 105546–105546. https://doi.org/10.1016/j.bspc.2023.105546

[23] I.U. Haq, M. Ullah, K. Muhammad, S.W. Baik, Deep learning Techniques for Oral Cancer Diagnosis. Computational Intelligence in Cancer Diagnosis, Elsevier eBooks, (2023) 175–193. https://doi.org/10.1016/B978-0-323-85240-1.00015-8

[24] P. Shah, N. Roy, Pinakin Dhandhukia, Algorithm Mediated Early Detection of Oral Cancer from Image Analysis. Oral Surgery Oral Medicine Oral Pathology and Oral Radiology, 133(1), (2021) 70–79. https://doi.org/10.1016/j.oooo.2021.07.011

[25] F. Temurtas, K. Gorur, O. Cetin, I. Ozer, Machine Learning for Thyroid Cancer Diagnosis. In Elsevier eBooks (2023) 117–145. https://doi.org/10.1016/b978-0-323-85240-1.00011-0

[26] T. Shanmuga Sundari, M. Maheswari, Automatic Oral Cancer Detection using Deep Learning Techniques. Biomedical Signal Processing and Control, 106, (2025) 107731. https://doi.org/10.1016/j.bspc.2025.107731

[27] T. Thakuria, T. Rahman, D.R. Mahanta, S.K. Khataniar, R.D. Goswami, T. Rahman, L.B. Mahanta, Deep learning for Early Diagnosis of Oral Cancer via Smartphone and DSLR Image

Analysis: a Systematic Review. Expert Review of Medical Devices, 21, (2024) 1189-1204. https://doi.org/10.1080/17434440.2024.2434732

[28] S.A. El-Ghany, M. Azad, M. Elmogy, Robustness Fine-Tuning Deep Learning Model for cancers diagnosis based on histopathology image analysis. Diagnostics, 13(4), (2023) 699. https://doi.org/10.3390/diagnostics13040699

[29] L.M. De Lima, M.C.R. De Assis, J.P. Soares, T.R. Grão-Velloso, L.A.P. De Barros, D.R. Camisasca, R.A. Krohling, Importance of Complementary Data to Histopathological Image Analysis of Oral Leukoplakia and Carcinoma using Deep Neural Networks. Intelligent Medicine, 3(4), (2023) 258–266. https://doi.org/10.1016/j.imed.2023.01.004

[30] B Goswami, B., Bhuyan, M.K., Alfarhood, S., Safran, M. (2024). Classification of Oral Cancer into Pre-Cancerous Stages from white Light Images using LightGBM Algorithm. IEEE Access, IEEE, 12, 31626–31639. https://doi.org/10.1109/access.2024.3370157

Authors Contribution Statement

D. Chandra Mouli: Conceptualization, Methodology, Analysis, Writing, Review and Editing. Pullela SVVR Kumar: Data Curation, Investigation, Review and Editing D. Haritha: Formal Analysis, Visualization, Supervision, Review and Editing. All the authors read and approved the final version of the manuscript.

Funding

The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?

Yes