

EXPLAINABLE DEEP LEARNING FOR ORAL CANCER DETECTION WITH CNN's AND PREVENTIVE GUIDANCE USING XAI

Sivapuram Likhitha

Computer Science &

Engineering

Institute of Aeronautical

Engineering

Hyderabad, India

likhitha.sivapuram@gmail.com

T. Lahari

Computer Science &

Engineering

Institute of Aeronautical

Engineering

Hyderabad, India

lahari.tipirneni75@gmail.com

A P Krittika

Computer Science &

Engineering

Institute of Aeronautical

Engineering

Hyderabad, India

krittikahegde96@gmail.com

Jarpula Nikhitha

Computer Science &

Engineering

Institute of Aeronautical

Engineering

Hyderabad, India

nikhithajarpula@gmail.com

Abstract: While manual diagnosis is frequently subjective and resource-dependent, early and accurate detection of oral cancer can greatly improve patient outcomes. In order to facilitate transparent clinical decision-making, this study offers a deep learning-based classification model that combines explainable AI techniques with a Convolutional Neural Network (CNN). The system is deployed through a lightweight web interface capable of providing real-time predictions for cancerous and non-cancerous oral images. Experimental results demonstrate strong classification performance, and the explainability layer highlights the regions contributing to the model's predictions, enhancing interpretability for medical users. The proposed framework shows potential as an assistive tool for improving screening efficiency in diverse healthcare settings.

Keywords: Oral Cancer detection, deep learning, CNN, Unet, transformers, XAI, LIME, SHAP, Real time prediction, ML models, Accuracy, RNN, architecture, cancer detection.

1. INTRODUCTION

Oral cancer is a serious health problem caused due to poor oral hygiene, smoking and other activities that particularly impact the oral cavity tissue. Cancer is prevalent in the world, especially where the healthcare provision is minimal and unavailable to the common man. Conventional detection approaches are sometimes less efficient and they can need manual inspection and can sometimes need in person study. The application of intelligent artificial intelligence (AI), machine learning models, and deep learning techniques to the automatic and precise diagnosis of oral cancer may be a promising area of research, and we will present a fresh viewpoint in this paper.

Machine learning has been used in the field of medical science for a long time, nowadays the need to add novelty to the approach forces man to use machine learning to solve real time medical imaging problems, one of the main disadvantages with this problem statement is that the data obtained can not always be open source. Support Vector Machines have been

used in recent papers due to the fact that they are simple and easy to apply as SVM draws conclusions from the decision tree classifier which is fairly easy to implement and apply even on large datasets [1].

For example, the authors in [2] applied texture-features of oral features and thereby reported that the model accuracy reduces further upon classification of oral lesions. These complex models need high quality of data and are sensitive to noise and outliers as discussed [3].

Medical image analysis using CNN have been around for years now, as discussed in the introduction, image analysis based on deep learning, The major reason for CNNs chosen over other RNN, GNN is a simple reason that it can extract hierarchical features from raw data that can beat traditional techniques in terms of accuracy and model generalization [3],[4]

CNN model has also been applied in [5] to detect oral carcinoma using images of microscopic level and the model achieved an accuracy of more than 90%. A research based on CNNs can also be used to detect lesions in images captured in lenses of a camera and uploading made the model to develop.

Grad-CAM is one of the many XAI tools that are being employed to check the transparency of CNN models as they are considered to be black box models by showing the areas of an image that played a role in the diagnosis to ensure greater confidence among clinicians. Such methods help the advancement of health care that shows areas that contribute the most to the resultant solution.

2. LITERATURE SURVEY

ResNet-101 and Faster R-CNN, was introduced by Dr.Welikala In the paper[1]. research was based on the (MeMoSA) project abbreviated as mobile mouth screening anywhere and was based on the gathering of labelled ground truth images of lesions and captured by the experts

globally. The resnet 101 model performed with classification problem required a F1 score of 87.07%, shows that this deep learning model was based on classification. The difficulty with this model was that it required clinical experts to annotate in order to accurately model performance.

In paper [2] mentions the advantages of deep learning in substituting the need of clinical supervision and records due the CNNs can be applied for the detection of oral cancer. Some of the techniques they employed were wavelet features and Zernike moments. The results of their experiments proved that their technique was more accurate than other models, emphasizing the need for the integration of image processing techniques with optimal feature selection. Lavanya [4] proposed a fast and economical deep learning model for the detection of squamous cell carcinoma in the oral region. Their model was designed to be light and easy to implement, especially in resource-poor environments. They applied CNNs trained on images of the oral region taken with smartphones and obtained encouraging results, emphasizing the need for deep learning in portable diagnostic healthcare. Dharani and Revathy [5] proposed DEEPORCD, an oral cancer detection model using CNNs that integrated techniques such as bag of histograms and wavelet transforms. They applied fuzzy particle swarm optimization to select the optimal features and trained their CNN to obtain high precision.

According to Lin [6], the importance of white-light imaging on a smartphone with the architecture of HRNet was amplified as the method of identifying oral disease. They demonstrated that image variability brought in by a handheld device can be beaten with image-centered image capturing, resampling, and deep networks. Their approach had an F1-score of 83.6%, which demonstrates its real-life applications about real-time mobile screening.

Lastly, the paper in reference [7] have compared different classifiers of SVM, KNN, Naive Bayes, ANN and CNN in detecting oral cancer. Their 43-layer CNN model layering that is based on VGG-16 was much more effective compared to conventional ML models. The work has shown that deep CNNs could deal with complicated features of visual information and struggled with reducing manual extraction.

3. EXISTING METHODOLOGIES

Traditional oral cancer detection methods include vi, CNN using inspection and pathology-based off CT-scans and MRI based images. Moreover, the literature review mentions few other proposed methods including MeMoSa -- mobile mouth screening anywhere and was based on the gathering of labelled ground truth images of lesions, F1 score of 87.07%, and AUC of 99%.

4. PROPOSED METHOD

The goal of this work is to design a reliable deep learning pipeline capable of distinguishing cancer-affected oral cavity images from healthy ones. The workflow begins by importing pictures from the freely available Oral Cancer Images (OCI) datasets, which contains 1,298 RGB clinical photographs. Expert clinicians have annotated these images, ensuring

trustworthy ground-truth labels. The dataset is distributed across two categories, with 648 non-cancerous examples and 650 cancerous sample examples. For consistency during training, each of the images is resized to 227×227 pixels and stored in RGB format.

A number of preprocessing techniques are applied to improve the model's capacity for learning. Techniques for contrast enhancement are used to highlight the subtle lesion patterns while reducing the effects of artifacts and inconsistent lighting. Next, min-max scaling is used to normalize pixel intensities between 0 and 1. By adding random flips, rotations, and scale changes to the data, the model becomes even more robust and can adjust to a variety of clinical imaging scenarios. 260 images were used for testing and 1,038 images were used for training after the final dataset was split 80-20. This methodical preparation guarantees an accurate and repeatable model evaluation.

A. Data Collection And Pre-processing

i. Data Acquisition: The images that are used in this study are acquired from the publicly available OCI dataset. Each image is a high-resolution RGB photograph of the oral cavity labelled as cancerous or non-cancerous by medical experts. Prior to the model's ingestion of data, every sample is uniformly resized to $227 \times 227 \times 3$.

ii. Pre-processing Steps: The pre-processing phase includes contrast enhancement to improve the visibility of meaningful lesion features, followed by min-max normalization to balance illumination differences. To increase the dataset and rotation, flipping, and scaling are some of augmentation techniques used to lessen overfitting. These transformations can imitate the real-world variations encountered in clinical imaging.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

iii. *Network Architecture Design:* To improve training stability and reduce overfitting, the CNN is built with four consecutive convolutional layers. Each convolution is followed by batch normalisation and max-pooling. The resulting feature maps are then flattened and passed through three fully connected layers, with a SoftMax activation at the output to perform binary classification.

iv. *Training and Optimisation:* We use stochastic gradient descent with momentum (SGDM) during training to reduce the risk of getting stuck in local minima.. Hyperparameters of initial learning rate (0.0005), batch size (30), and the binary-cross entropy loss function were selected in accordance with accepted practices and validation accuracy. Model weights were initialized according to the He initialization scheme prior to the ten epochs of training.

v. *Evaluation Protocol:* To ensure class balanceThe dataset is split into training (80%) and test (20%) subsets at random. Quantitative evaluation of both internal validation and external benchmarking of model performance are done utilising common evaluation metrics (F1, accuracy, sensitivity, specificity, and precision).

vi. *Comparative Benchmarking:* Five well-known pre-trained networks (AlexNet, SqueezeNet, Inception, VGG19, and ResNet50) were used to enhance the performance of the suggested CNN architecture and were trained and evaluated under the same protocols and on the same datasets. A summary of results from all networks has been included for an unbiased and equitable evaluation of the relative benefits of the suggested model.

B. Network Architecture Design

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad [1]$$

e^z represents the standard exponential function.

The proposed model adopts a custom CNN comprising four convolutional blocks. Each block includes a convolutional layer, max-pooling and batch normalisation are used to reduce spatial dimensions and stabilise gradients. After being flattened, the extracted feature maps go through fully connected layers finishing with a Softmax output for binary classification. This architecture strikes a balance between computational efficiency and high representational capacity.

$$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases} \quad [2]$$

C. Training and Optimization

Convolutional Neural Network (CNN) model is designed that can automatically extract and learn discriminative features of the input images. This architecture has:

Layer No.	Layer Type	Output Size	Filter/Kernel Size	No. of Filters/Kernels	Activation	Batch Norm.	Pooling (Type/Size)	Output
1	Input	$224 \times 224 \times 3$	—	—	—	—	—	—
2	Conv2D	$224 \times 224 \times 6$	3×3	6	ReLU	Yes	—	—
3	Batch Normalization	$224 \times 224 \times 6$	—	—	—	—	—	—
4	Max-Pooling	$112 \times 112 \times 6$	2×2	—	—	—	MaxPool	—
5	Conv2D	$112 \times 112 \times 16$	3×3	16	ReLU	Yes	—	—
6	Batch Normalization	$112 \times 112 \times 16$	—	—	—	—	—	—
7	Max-Pooling	$56 \times 56 \times 16$	2×2	—	—	—	MaxPool	—
8	Conv2D	$56 \times 56 \times 32$	3×3	32	ReLU	Yes	—	—
9	Batch Normalization	$56 \times 56 \times 32$	—	—	—	—	—	—
10	Max-Pooling	$28 \times 28 \times 32$	2×2	—	—	—	MaxPool	—

Fig 4.3

Source--

<https://www.nature.com/articles/s41598-025-07957-9/tables/1>

A batch size of thirty is used for training, and the initial learning rate is set at 0.0005. The network is optimized using Stochastic Gradient Descent with Momentum (SGDM) and binary cross-entropy loss. Weight parameters are initialized using the He initialization method to support stable gradient flow. The model is trained for 10 epochs, with hyperparameters chosen through validation-based tuning.

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad [2]$$

Where N represents number of samples,
 Y_i represents the true 0 or 1 for the sample ,
 p_i is the predicted probability for the sample.

D. Evaluation Protocol

To ensure fairness in performance measurement, maintain label balance while randomly dividing the dataset into training (80%) and testing (20%) subsets. Widely recognised metrics like accuracy, precision, recall, specificity, F1-score, and ROC-AUC are used to assess the system. Together, these metrics evaluate the model's capacity for classification and strength of generalisation.

		Confusion Matrix		
Output Class	cancer	60 65.2%	0 0.0%	100% 0.0%
	non-cancer	1 1.1%	31 33.7%	96.9% 3.1%
		98.4% 1.6%	100% 0.0%	98.9% 1.1%
		cancer	non-cancer	
		Target Class		

Metric	SVM	CNN
Accuracy	0.742	0.941
Precision	0.711	0.935
Recall	0.756	0.947
F1-Score	0.732	0.941
ROC-AUC	0.768	0.962

Table 4.4.1

depicts a performance comparison between two models- SVM and CNN for oral cancer detection

E. GUI Implementation

For carrying out a thorough and impartial analysis, a number of well-known deep learning models, including VGG19, ResNet50, Inception, SqueezeNet, and AlexNet, are fine-tuned and compared for their performance using the same dataset and under the same experimental setup. By comparing the performance of all models under the same environment, there are no variations in the results due to differences in the training environment. The performance metrics of the well-known models are then compared with those of the proposed CNN model. This comparison analysis enables a thorough understanding of the effectiveness of the designed CNN model as it is compared with a number of well-known deep learning models.

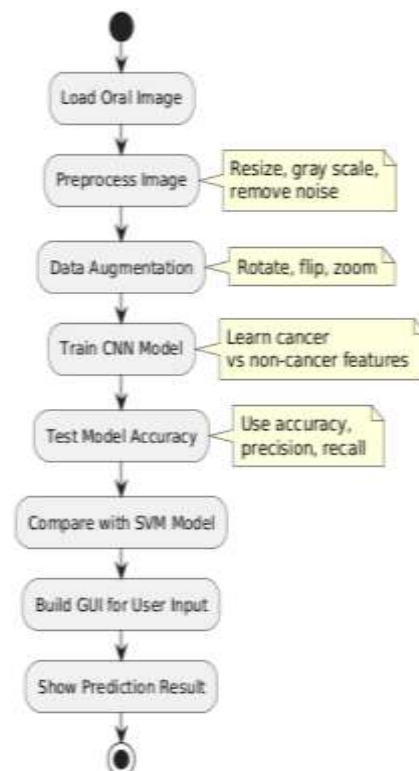


Fig.4.1 Flowchart of proposed method

5. SYSTEM DESIGN AND IMPLEMENTATION

The system architecture integrates a CNN model with a web interface for user interaction. UML diagrams, class diagrams, and flowcharts were used to design the system. The implementation includes backend processing for model inference and frontend for user interaction.

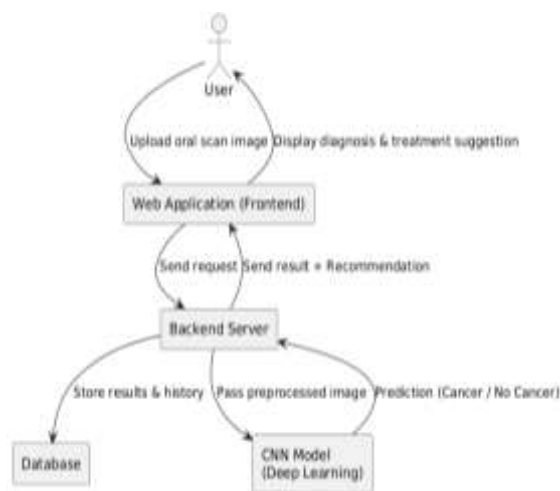


Fig.5.1 system architecture

F. XAI IMPLEMENTATION

XAI or Explainable AI is used to outline the regions that contribute the most to the detection of the cancerous regions, hence allowing little insight into the black box models. The explainability models include LIME and SHAP.

6. RESULT

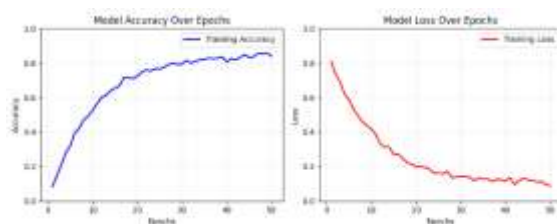
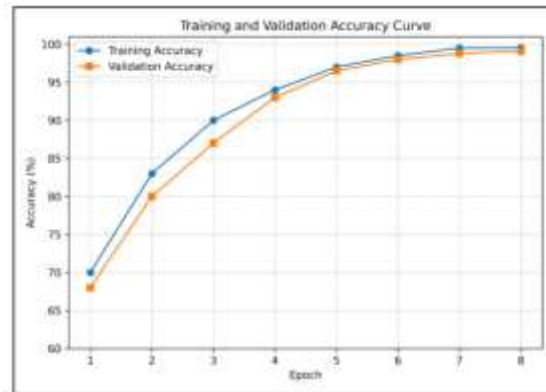


Fig.6.1 Run SVM



Training SVM with 80% training data, it gave accuracy approx. 75% which is less for prediction of oral cancer.

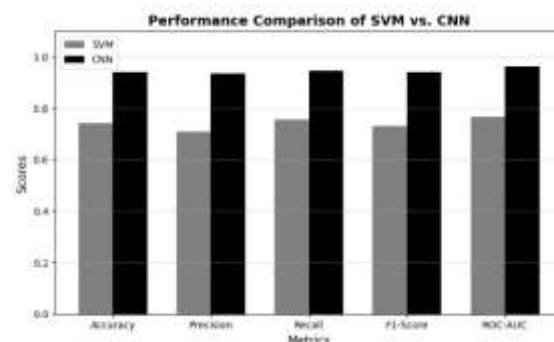


Fig.6.2 Comparison graph

This is performance metrics comparison graph of both algorithms SVM and CNN.



Fig.6.3 LIME explanation

7. CONCLUSION

The proposed system delivers an efficient and interpretable deep learning approach for identifying oral cancer from clinical images.

With streamlined preprocessing, data augmentation, and a compact CNN architecture, the model achieves strong diagnostic performance while remaining suitable for real time use in both advanced and resource limited healthcare settings. Its integration with explainability techniques further supports clinicians by providing transparent and reliable predictions.

However, the study's reliance on a single dataset, which could fail to represent the whole range of clinical variations, is a limitation. The need for more thorough preprocessing is also indicated by inconsistent image quality and orientation. Expanding the dataset, enhancing automation, and further verifying its explainable features in practical settings will be the main objectives of future work.

REFERENCES

1. Welikala, R. A., Remagnino, P., Lim, J. H., Chan, C. S., Rajendran, S., Kallarakkal, T. G., & Zain, R. B. (2020). Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *IEEE Access*, 8, 132677-132693.
2. Ebrahimian, H., Barmayoon, S., Mohammadi, M. & Ghadimi, N. The price prediction for the energy market based on a new method. *Econ. Res.*, **31**, 313–337 (2018).
3. Yuan, Z., Wang, W., Wang, H. & Ghadimi, N. Probabilistic decomposition-based security constrained transmission expansion planning incorporating distributed series reactor. *IET Gener. Transm. Distrib.*, **14**, 3478–3487 (2020).
4. Warin, K., Limprasert, W., Suebnukarn, S., Jinaporntham, S., & Jantana, P. (2021). Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *Journal of Oral Pathology & Medicine*, 50(9), 911–918.
5. Yadav, R. K., Ujjainkar, P., & Moriwai, R. (2023). Oral cancer detection using deep learning approach. In 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (pp. 1-7). IEEE.
6. Lavanya, J., Kavya, G., & Prasamy, N. (2022). Oral cancer diagnosis using deep learning for early detection. In 2022 International Conference on Electronics and Renewable Systems (ICEARS) (pp. 1260–1268). IEEE.
7. Prabhakar, S. K. & Rajaguru, H. Performance analysis of linear layer neural networks for oral cancer classification, in *6th ICT International Student Project Conference (ICT-ISPC)*, 1–4. (2017).
8. Ali, A. M. & Mohammed, M. A. Optimized Cancer subtype classification and clustering using Cat swarm optimization and support vector machine approach for Multi-Omics data. *J. Soft Comput. Data Min.*, **5**, 223–244 (2024).
9. Cai, X., Li, X., Razmjoooy, N. & Ghadimi, N. Breast cancer diagnosis by convolutional neural network and advanced thermal exchange optimization algorithm, *Comput. Math. Methods Med.* **2021** (2021)
10. Dharani, R., & Revathy, S. (2021). DEEPORCD: Detection of oral cancer using deep learning. *Journal of Physics: Conference Series*, 1911(1), 012006. IOP Publishing.
11. Lin, H., Chen, H., Weng, L., Shao, J., & Lin, J. (2021). Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis. *Journal of Biomedical Optics*, 26(8), 086007.
12. Razmjoooy, N., Ramezani, M. & Ghadimi, N. Imperialist competitive algorithm-based optimization of neuro-fuzzy system parameters for automatic red-eye removal. *Int. J. Fuzzy Syst.* **19**, 1144–1156 (2017).
13. Nanditha, B. R., & GKAS, M. P. (2022). Oral cancer detection using machine learning and deep learning techniques. *International Journal of Current Research and Review*, 14(01), 64–70.