International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 496-504, May-2026
https://doi.org/10.47001/IRJIET/2026.105069

AI-Powered Oral Cancer Detection System Using Deep Learning

[1]Ashphak Khan, [2]Anamika Patil, [3]Nikita Borse, [4]Ashwini Shinkar

[1,2,3,4]Department of Computer Engineering, D. N. Patel College of Engineering, Shahada, Dist-Nandurbar, Maharastra, India

***Abstract* - Oral cancer is a life-threatening disease where early detection is critical for effective treatment and improved patient survival rates. Traditional diagnostic methods are often invasive, expensive, and dependent on expert interpretation, which limits accessibility in resource-constrained settings. This study proposes an AI-based deep learning framework utilizing Convolutional Neural Networks (CNNs) to analyze clinical and histological images of oral lesions for automated detection and classification. The model is trained on a curated dataset to identify distinguishing patterns between benign and malignant cases with high accuracy. By automating the diagnostic process, the system aims to assist healthcare professionals in early identification of oral cancer, reduce diagnostic delays, and alleviate the workload on medical experts. The proposed approach has the potential to enhance screening programs, particularly in underserved regions, and contribute to improved health outcomes in oral oncology.**

Keywords: Deep learning; Images Classification; Oral Cancer; Feature Extraction.

I. INTRODUCTION

In conventional practice, a pathologist investigates the histopathological images of oral mucosa, i.e. the study of tissue samples of the affected area by a microscope. The pathologist visually examines the image, and the keratinized area and this manual assessment process fully depend on the pathologist's expertise and experience [1]. But, this manual process is time-consuming and prone to diagnostic errors, which can be solved by using deep learning technology which automates the image classification of cancer cells. It provides an accurate classification of extensive image dataset by training the algorithm with the experts' knowledge [2].

Deep learning has been applied to several applications, particularly in the field of medicine for medical image analysis [3]. Deep learning provides an advanced classification technology enabling calculation models consisting of several processing layers to learn the data representations used for oral cancer image classification. The algorithm is trained using expert knowledge, and the trained network is tested using the remaining dataset, which provides the exact feature map that results in highly accurate image classification [2]. The network arrangement and lack of proper training dataset results in the network not precisely producing the required feature map with dimension reduction, which could overfit the network. The solution to reduce this problem is using a convolutional neural network classifier. This classifier is further modified to generate the exact feature map from the input dataset learning required for accurate oral cancer classification. A current study of Deep Learning uses various algorithms and techniques for better feature extraction and image classification. For accurately classifying oral cancer images, it is necessary to train the algorithm using the proper set of the input dataset with a less complex neural network. The current state of art using Inception v3 convolutional neural network provides classification accuracy of 87.02% [4]. Research on this network has identified overfitting problem which cannot produce the exact required feature map of classification accuracy. In order to overcome this [5], introduced autoencoder architecture for extracting characteristic features from the input data but still suffers from reconstruction error. Therefore, current studies still have an area that needs be improved.

The contribution o of this paper summarised in the below points • Increase the classification accuracy and speed for image classification of oral cancer.

- This research aims to provide highly accurate feature map, input dimension reduction and high classification accuracy by decreasing overfitting.
- This study proposes an autoencoder architecture that extracts characteristic features from the input dataset, which regenerates the input data from those features by learning a network [5]. In addition, the input is passed for de-convolution and de-pooling to extract the optimized image input. This helps extract more complex image features with higher-order structures.

II. LITERATURE REVIEW

The main aim of this review is to conduct a survey of different available papers and find the new solution to improve the existing system. The following section provides a review of different papers and provides an idea about the existing solution to the problem. [2] modified the pre-trained

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 496-504, May-2026

https://doi.org/10.47001/IRJIET/2026.105069

Convolutional Neural Network, the Google Net Inception V3 CNN architecture by using the regression-based partition convolution and subsampling layer to use optimum data into the training network. The author improved the state of art's solution [6]. Their structure processes complex data which produced classification accuracy of 94.5% with specificity 0.98 and sensitivity of 0.94, which is greater than another base classifier in a single phase of training data set. But the classifier has not been designed to extract the feature map of fewer variation data which need to be considered to increase the classification accuracy.

[7] employed 50 CBCT 3D image dataset verified by the expert for identifying periapical cyst and keratocystic odontogenic tumor (KCOT) lesions for exact classification. They offered the solution to the problem performing segmentation on Cone Beam Computed Tomography (CBCT) images using the viewer software. The author marked the lesional volume of interest and calculated order statistics for each CBCT dataset which was not marked in the previous[8] solution. This solution identifies the best classifier that is SVM with the accuracy percentage of 96% and F1 score of 96%, which is the best amongst other classifiers. The classifier performance has been increased by decreasing the size of the feature vector. BCT dataset needs to be enhanced with different types of dental pathologies to enhance accuracy. [4] studied a different classification approach for classifying Confocal Laser.

Endomicroscopy (CLE) images and considered patch probability fusion method to outperform other conventional approaches. The author improved the state of art's solution [9]. In this method, the author designed the network with LeNet-5 network, 2 convolutional layers with different filter sizes with max-pooling layer, only fully connected layer and output softmax output. This design has increased the classification accuracy to 87.02% at a sensitivity of 90.71% and specificity of 83.80%. However, the classification can be improved further by adding entities like precursor affected region of cancer and adding histopathology process to assess the tissue.

[5] experimented using a different network architecture and extended the network to process larger pathological input images to evaluate the local phenotypic feature and their distribution in the tissue. This solution has improved the current state of art solution [10]. They offered a solution to implement deep convolutional autoencoders to extract the characteristic features from the given inputs. Further, they built a sample based on three autoencoders and one classification reducer to evaluate larger pathological images classifying the transcriptome subtypes. The complex feature extraction increased with the larger input image size, and hence the accuracy of classification also increased to 98.89%. However, it is hard to differentiate the statistical distribution of cellular features in bigger tissue input. Thus, the latest approach can be developed for the differentiation of various tissue types. [11] modified the AlexNet architecture to dropout 1 layer from the fully connected layer and used 5 layers as convolutional layers and 2 as fully connected layers and used dropout probability of 0.5 to make the classification appropriate for the small-size data analysis which is the improved solution [12]. It helped to quickly analyze the minimum number of clinical images that are ready for the training of dataset. This resulted in an accuracy of 78.2% which result in matches to that of the 2 experienced radiologists study results. While the diagnostic performance of the deep learning system had an accuracy of 78.2% on using the arbitrarily sized square image patches as an input, there was no consideration of impact of datasets and input image size sampling on the training. But this model failed to operate in real time due to manual image segmentation.

[13] has added the detection operation in segmentation operation, which alone focuses on a particular reason containing an organ and segments this particular organ from others. This solution has improved the current state of the art solution [14]. This solution has reduced the overall detection time to 16s and segmentation time to 30s for labelling of a single image input which is suitable for the clinical workflow and is more optimized than other solutions such as Fully Convolutional Neural Network (FCN). Although the solution has improved the sensitivity from 0.997 to 1 for the most organs, the input is only taken from the non-contrast CT scanning. Which limited the accuracy of detection and segmentation. Thus, the current Organs-at-risk detection and segmentation network (ODS net)) can be improved by considering both the non-contrast and contrast CT images. [15] have fused the autofluorescence and white light images information into three different image channel and feed the information into the deep learning neural network. This solution has improved the current state of art solution [16]. The small size network with five different convolutional networks is used to train the small data sets to reduce the network complexity. Hence, this system provides the average accuracy of 86.9% with sensitivity of 85% and specificity of 88.7% when applied in a 4-fold cross validation. Which improved the performance in detecting the oral cancer and provided the cost effective solution to many smartphone users for further test and treatment.

[17] used transfer learning by using the greater sample dataset on the radiographic dataset with known biopsy for preparing training input data set. They prepared test data using 50 ameloblastoma images and 50 Keratocystic Odontogenic Tumor (KCOT) images to overcome the limitation of

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 496-504, May-2026

https://doi.org/10.47001/IRJIET/2026.105069

minimum training dataset. The author pre-trained the VGG-16 (16-layer CNN) network in ImageNet and developed the gradient weighted class activation maps (Grad-CAM) to identify the unequal regions on panoramic digital X-ray images. This solution has improved the current state of art solution [18]. It provides an accuracy of 83.0 % and time to analyze the images is 38 seconds as compared to 23.1 minutes time taken by the oral and maxillofacial surgeons for classification [18]. However, the lateral X-ray radiographs have not been considered for image input. Both frontal X-ray and lateral view Xrays should be used as input to enhance accuracy, and the learning approach needs to be improved to be applied effectively in the clinical domain. [19] sampled the input by converting the large image into the spectral patches, which is used in the convolutional neural network for image classification.

They evaluated the performance by counting the steps and using the cross-validation method, which is also the benchmark for the proposed solution. This solution has improved the current state of art solution [20]. The author provided the solution of using 37-fold, leave-one-out external-validation which showed the reliability of classification and suggested that it can be used for any new patient. While this solution provides a reliable method of classifying the classifying normal and cancerous tissue from the hyperspectral image with better accuracy but the classification is totally based on the few data samples. Thus this problem can be solved by using the larger data set with more patient HSI data.

[21] has modified the convolutional neural network (CNN) of AlexNet to fc-CNN model by modifying 3 dense layers at the end to operate it on a sliding window. This solution has improved the current state of art solution [22]. It helped to process the arbitrarily large input which uses Region of Interest (ROI) as input. The accuracy gained by Active Learning (AL) has been improved by 3% as compared to Random Learning [23], without requiring any additional cost to training. While the classification accuracy of active learning applied in 3 iterations is increased by 3% over random learning, stain normalization on training has not been considered in the experiment. Thus, the proper normalization of data needs to be done to avoid the generalization errors. [24] improved the current predictive model by considering the tumour depth as a measuring factor. The author applied recursive feature elimination to determine the most important feature to optimize the classifier performance. Which improved the current state of art solution [25]. They offered the solution by developing the classificaiton algorithms to predict pathological lymph node metastasis which maximized the area under the receiver operating characteristic curve (AUC) to 0.840 using decision forest tree algorithm. While the classification performance has been improved to 0.840 as compared to another predictive model, the quality of input data set has not been considered in the experiment. The algorithm needs to consider the quality of data in which the actual depth of invasion (DOI) is measured according to the accepted standard.

[23] used their own Fully Conventional Neural Network (FCN), CN24 and the initialization of the weights was done by pretraining the open-source ILSVRC2012 dataset for improving the training data set. This solution has improved the current state of art solution [26]. They also divided the whole image size to 384*384 region tiles for better processing and reduction in memory. Hence, they used a pixel-wise classification of the imaged tissue, which helped to increase the accuracy in overall recognition rates of 75% and 83%. While the overall recognition rates of the input images have been improved, but there is still doubt if pixels of head and neck cancer and other tissues using multimodal images alone is enough for the classification. The other types of pathological images of the tissues should be considered while conducting the experiment. [27] enhanced the current deep convolutional neural network (DCNN) to improve the average validation accuracy. They modified VGG-16 CNN by removing the first fully-connected layers and adding a new layer with three units to convert the output to a three-class probability. This solution has improved the current state of art solution [28]. They used a softmax layer to convert the output of the fully connected layer and used VGG-16 parameters pretrained with IMAGENET for transfer learning. It provided the best average validation accuracy of 60.7%, 64.7%, and 68.0%, for the image size of 56, 112, and 224, respectively. While the average validation accuracy has been improved for 2D computed Tomography images [5], 3D images need to be considered for the better accuracy for ternary classification. The author has only investigated the effect of small size images on the output, but they need to determine the effect of bigger image size.

[29] designed an adapted convolutional neural network (CNN) architecture to the characteristics of mass spectra. They introduced the analysis tool based on the sensitivity of the input-output relationship to allow interpretation in the input domain. This solution has improved the current state of art solution[30]. It revealed the model artifacts that distinguished lung and primary pancreas tumour. The current solution used the Imaging mass spectrometry (IMS) data but has not considered the pixel-to-pixel variation in data. The source data might trigger some technical issues like misalignment or variation of sample data of different patients. Thus the experiment can include the study of design which involves data from various institutions, patients or operating devices the data should be labelled and aligned with the design choices. [31] has modified the VGGNet to design the patch-based

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 496-504, May-2026

https://doi.org/10.47001/IRJIET/2026.105069

DCNN architecture with data augmentation process. They combined T2-weighted imaging (T2WI), diffusion-weighted imaging [24], and apparent diffusion coefficient (ADC) and used one high b-value DWI image, which achieved better diagnostic accuracy than using three different bvalues. The performance has been improved with the area under curvature value of 0.944 with 95% confidence interval compared to the traditional prediction model. But the author has not considered the data set size and parameter initialization, which need to be considered for better performance.

III. PROPOSED SYSTEM

- **Data Collection and Preprocessing:** Collect oral images from medical datasets or clinical sources. Preprocess them by resizing, normalization, and data augmenta7tion.
- **Model Development:** Train and test machine learning models like CNN to classify normal and cancerous images.
- **Web Integration:** Develop a Django-based web app to upload images, run predictions, and display results with visual explanations like heatmaps or highlighted regions.
- **Model Training:** Supervised ML algorithms including CNN, SVM, Random Forest.
- **Evaluation:** Accuracy, precision, recall, and F1-score metrics are used to assess model performance

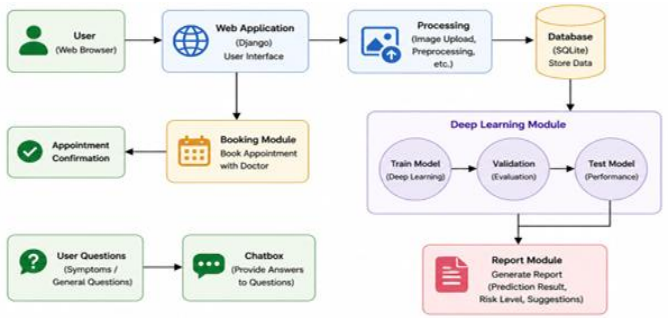


Figure 1: AI-Powered oral cancer detection system

Image Pre-processing: The suspected carcinogenic regions were resected after image acquisition. Noise removal or filtering is used to enhance the quality of the images, which are subsequently used for image recognition [4]. The two-fold data augmentation is applied to the image to enrich the data provided to the classifier by random rotation of the image. Since CLE data is 16 bit and the Inception v3 model only accepts 8-bit image input, dynamic compression is applied to scale the image. Feature extraction: In this stage, the Google Net Inception V3 Convolutional Neural Network architecture is applied to extract the important features [4]. First, the input data are scanned by convolutional filters; then the image is down-sampled by a max-pooling layer, which is finally passed to encoding layer, where un-pooling and de-convolution are done to generate the input data to minimize the difference between input and output data (Antonio et al., 2018) as shown in Fig 2. The sparsity penalty function is introduced to enhance the efficiency of the feature extraction and to compress the information in the autoencoder as shown in Fig 3. Essential features like lesion patch size and microscopic features are extracted as a feature map for the oral cancer classification. The overfitting problem is minimized by adding the penalty. The transfer learning helps decrease the training time and size of the training data set. The learned knowledge from the pre-training is applied to this dataset. Also, it dramatically reduces the number of parameters in the network, eliminating a large number of irrelevant parameters. Classification: From the Google Net Inception V3 CNN architecture, the final dense layer and the softmax layer are replaced with a new two-node dense layer and subsequent softmax layer which .classify the images as per the training [4].

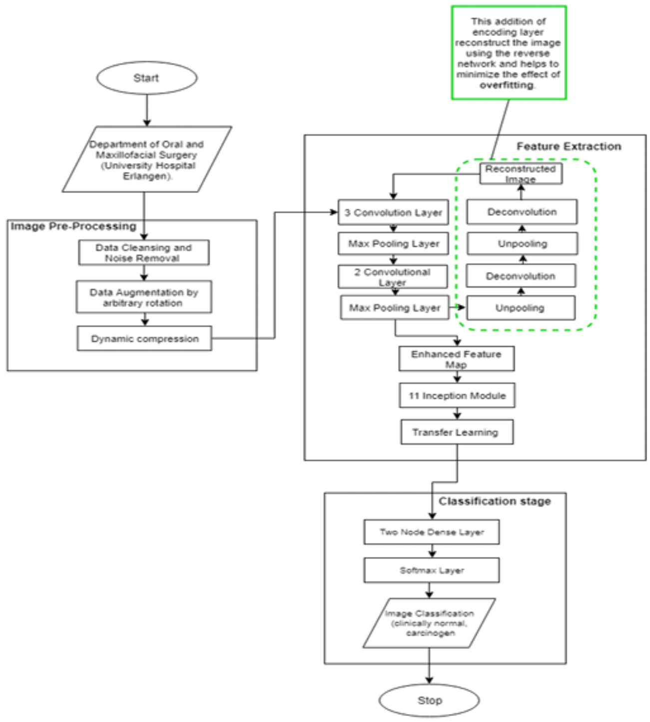


Figure 2: Block Diagram of the proposed deep learning system for oral tumor classification using autoencoder architecture

The different convolutional and pooling layers are usually used to extract more abstract feature representations in moving through the network. This two nodes dense layer and subsequent softmax layer classify the output of the neural network. The output is the probability distribution of all classes, which uses loss function to measure the difference between the actual output and the target output.

CNNs specialize in analyzing visual patterns in medical images using filters and convolution operations. For oral

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 496-504, May-2026

https://doi.org/10.47001/IRJIET/2026.105069

cancer detection, CNNs help identify abnormal tissue patterns, lesion shapes, color variations, ulcer formations, and irregular cell structures from oral cavity images. They are particularly effective in detecting early signs of oral cancer by analyzing image features such as texture, boundaries, and suspicious regions in medical and intraoral images.

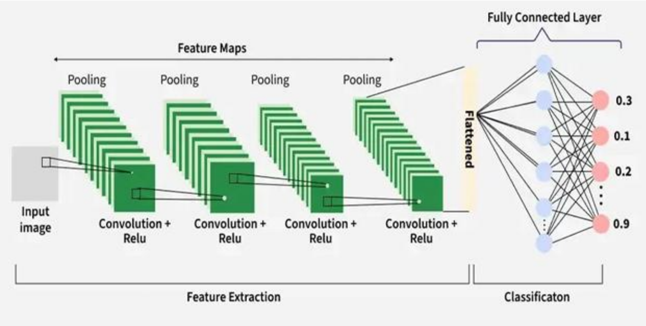


Figure 3: CNN Architectuer in AI – Oral Cancer

Enhancing feature extraction efficiency and compressing the information in auto-encoder has increased the accuracy by adding the sparse penalty, as shown in equation.

$$Xii,jj = \mathrm{R} + \lambda s\ S$$

Where,

λs is a weight constant
R is the optimisation factor
S is the sparsity constant

For features extraction, Eq equation 3:

$$lS = 1n\sum(-rjencodelogrjencode)\ j=1$$

Here, $rjencode$ is the output intensity of filter j in the encoding layer relative to their total summation. x is the input q and l are the number of nodes in the input and encoding layers. λs is a weight constant The total network can be optimized by reducing the difference between the input and the output by using the reversed network in which the information is also encoded to enhance the feature extraction, as shown in equation 4. The features can be extracted during the process of optimising the input parameters by reducing the reconstruction error [32], which is given by the cost function as:

Finally, equation 1 [2] has been enhanced by us by modifying sparsity constrain suggested by [32] which is expressed mathematically below as equation:

$$qlEEj,k=f(b+\sum\sum Wi,z*MXii,jji=0$$

Where, 'b' is the bias factor z=0)

'W' is the weight of each neuron in layer j-1 for the input connection
MXii,jj is the modified input from the nodes
'q' and 'l' represents the size of the input matrix of shared weights of
W, 'ii' and 'jj' are the indexes of the input activation at position (j+i, k+z).

We proposed one equation. The accuracy value of the model is comparatively low due to extracting irrelevant characteristics. The feature extraction efficiency is enhanced, and the information in autoencoder is compressed by introducing the sparsity penalty, as shown in equation. This helps in minimizing overfitting of the network, but still, reconstruction error is generated due to the iteration in training data which is shown in equation (3). The reconstruction error is reduced by modifying equation 3 with the help of equation (5). The difference between the input and the output in the network can be reduced by modifying equation (5) with the help of equation (6) to obtain the final modified reconstruction error equation (7). This equation (7) aim is to decrease the magnitude of the weights that helps prevent overfitting where weight decay parameter λ is introduced. Equation (2) is modified to obtain the equation (8) which is modified sparsity constrain to reduce the reconstruction error. Hence, with the help of equation (8), the feature extraction process is enhanced as shown by equation (9) to generate the final feature map, which improves the overall system classification accuracy.

IV. RESULTS AND DISCUSSION

Python 3.6.0 with Sliderunner were used for the implementation of the model using 32 different CLE images samples from different datasets of varying age groups. Sliderunner was used to cell annotations in whole slide images. Two convlutuonal layers were use the first convulational with 64 filters of (5x5)px, followed by (3x3) px max pooling layer. The second convolutional layer with 32 filters followed by (3x3)px max pooling layer. The sample data consist of images from various stages of radiotherapy treatment i.e. pre-treatment, mid-treatment, and post-treatment. The data samples exclude the participant such as pregnant women, patients younger than 18 years and patients whose size and weight would not allow scanning. Thesample dataset used in this training is a free open source, just as the folder that can be downloaded from the cancer imaging archive. The CNN with inception module is used for the implementation based on a deep learning algorithm. There are three groups of sample images taken from the three different anatomical locations such as vocal cord, oral cavity and samples from both vocal fold and oral cord area. Those three different anatomical images have been tested for the classification accuracy as shown in Tables 3, 4 and 5. We used

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 496-504, May-2026

https://doi.org/10.47001/IRJIET/2026.105069

1.6 GHz Octa Core i5 processor with 8 GB RAM memory for the experiment. The image quality is at an accepted level, and we configured our server and data storage to support our purpose. The patch from the whole image is extracted by using the file Patch extraction (and randomization) (extractPatches.py) in python. The file cellVizio is read, and the patches without annotated artifacts are extracted as per the value entered in the CLE database. The microscopic feature to determine the presence of lesion is extracted from the CLE image to classify it as a carcinogenic image of oral cancer. The classifier was trained and tested using the train.py file in which the numpy array with images probability is generated. We used 33% of the dataset for testing the network using 10 fold cross-validation. The patch probability calculation is done by equation 10 in Result Analysis table; Finally, the classification is done based on the patch probability function, which is based on the areas of the image that are covered multiple times by patches. The library used to calculate the confidence range for Python was Numpy. The accuracy is calculated using the confidence range shown by equation 11. During the feature extraction stage, the modified CNN extracts the important features from the image input based on the labelled data, which helps exact classification. See figure 4.

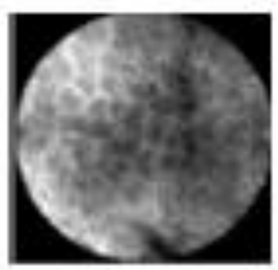


Figure 4: Input image after passing to Convolutional Neural Network and max pooling layer

The output from the convolutional layer and the max pooling layer is again passed to the autoencoder where the de-convolution and un-pooling are done to reconstruct the image and encode the information which is again passed to the same convolutional layer to generate the better feature maps as shown in the figure 5 below:

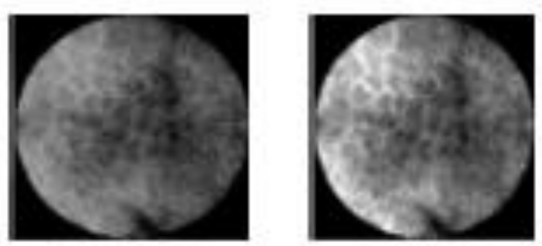
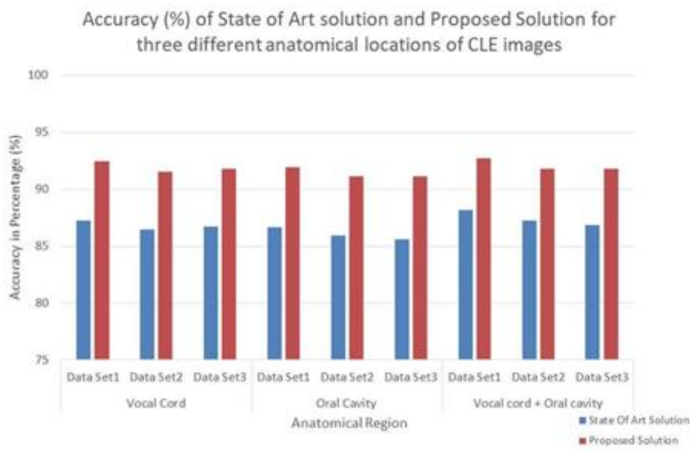
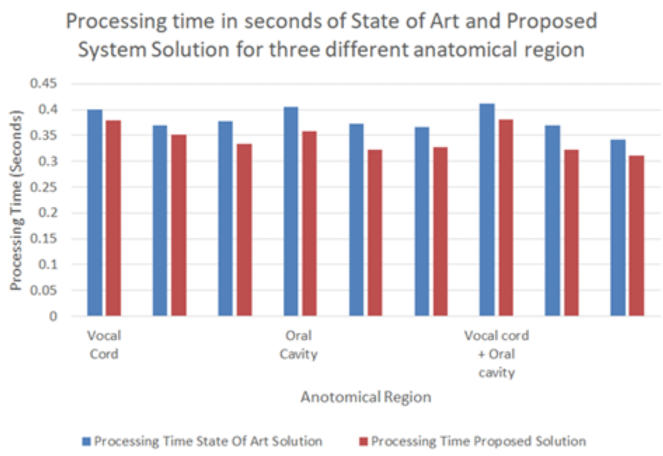


Figure 5: Image reconstruction after passing through reverse network for un-pooling and de-convolution



Bar graph shows the processing in percentage for three different anatomical locations of squamous cells of CLE images. The blue color indicates the accuracy of State of Art solution while orange colour indicates the accuracy of the proposed solution. (1) First couple, second couple and third couple of bar graph shows the average accuracy for vocal cord region for two different data sets (2) fourth couple, fifth couple and sixth of bar graph shows the average accuracy for oral cavity region for two different data sets (3) seventh couple, eight couple and ninth of bar graph shows the average accuracy for both vocal cord and oral cavity region for three different data sets.



Bar graph shows the processing time in seconds for three different anatomical locations of squamous cells of CLE images. The blue color indicates the accuracy of State of Art solution while orange color indicates the accuracy of proposed solution. (1) The first couple.

The results illustrate that there is an overall improvement in accuracy and processing time as compared to the state of art solution for the classification of images. The proposed system enhances the classification accuracy to 92% with the help of modified Convolutional Neural Network which is better than state of art solution. It processes the reconstructed image

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 496-504, May-2026

https://doi.org/10.47001/IRJIET/2026.105069

which requires less training than the original image to learn the feature. Similarly, processing time decreases by 20 to 30 milliseconds with the help of transfer learning mechanism that records the previous processing learning and finetunes the new image data set through the learning. The patch probability score is used to measure the accuracy whereas execution time is used to calculate the processing time of the system. Processing time is measured in second and accuracy is measured in percentage.

The current Convolutional Neural network using inception V3 module layer is modified using autoencoder which is used in feature extraction stage that enhances the performance of the proposed system. Python programming language was used for the implementation, which has minimized the processing time and increased the classification accuracy of the model. The use of autoencoder, reconstruction and enhancement of the image and encode the information which is then again used as an input for extracting significant features. The purpose of the two dense layer and softmax layers in a convolutional neural network for the classification has minimized the processing time by avoiding the random generation of the weight matrix. In conclusion, the Convolutional Neural Network combined with the autoencoder and dense layer has enhanced the CLE image classification with an overall improvement of accuracy by 5% and processing time by 20 ~ 30 milliseconds.

There are various techniques and algorithms that have been implemented for classifying CLE images. However, many alternatives and techniques have been researched to improve the accuracy and processing time of the classification. The limitation of the current best solution has been solved in this research with improved accuracy of 92% against the current accuracy of 87%. The system has reduced the processing time to 0.343 seconds from 0.380 seconds. This improvement in accuracy and processing time is a result of implementing the autoencoder in the feature extraction process to generate the important feature map hence solving the overfitting issue. The proposed system has better accuracy and reduced processing time in all different sample image groups of oral section such as oral cavity, vocal cord and both regions. See Table 1.

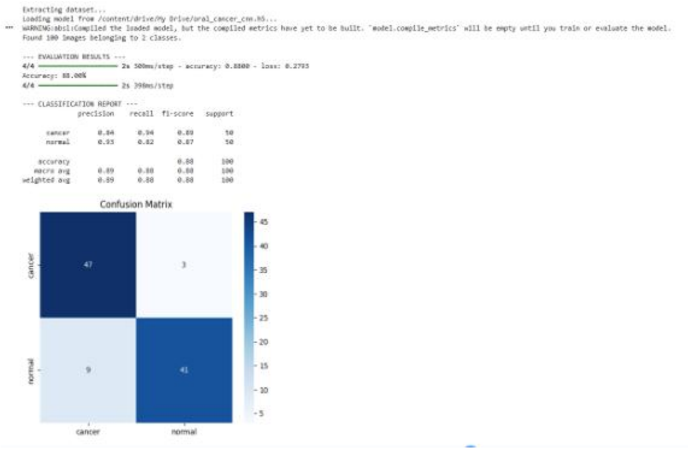


Figure 6: Training and Testing Model Result

Table 1: Comparison table between Proposed Solution and State of Art Solution

	Proposed Solution	State of Art Solution
Name of the solution	Inception V3 with Autoencoders	Inception V3 CNN
Accuracy	Improved the feature map by reducing the input dimension. Provides an accuracy improvement in classification of 92%	Provides an accuracy of 87%
Processing Time	Decrease in processing time from 0.380 milliseconds to 0.343 milliseconds	Provides a processing time of 380 milliseconds
Proposed equation	$EE_{j,k=}f(b + \sum_{i=0}^{h}\sum_{z=0}^{l} W_{i,z} * MX_{ii,jj})$	$E_{j,k=}f(b + \sum_{i=0}^{q}\sum_{z=0}^{l} W_{i,z} * X_{ii,jj})$
Contribution 1	Autoencoder helps to minimize the different between the input and the output by sending the output from max pooling to the reverse network which help to optimize the total network	The state of art does not consider the optimisation of inputs and calculates the energy as the original value.
Contribution 2	It helps to encode the information by introducing sparsity penalty for the optimization function.	The state of art select only one variable from a group of highly correlated variables and ignores others.

IV. CONCLUSION

The accurate classification of oral cancer tumour in confocal laser endomicroscopy (CLE) images is essential for the early diagnosis of oral cancer. Deep learning technology has been successfully implemented for the identification and classification of oral squamous cell carcinoma. The purpose of this paper is to improve the accuracy and to reduce the processing time of squamous cell carcinoma classification. The autoencoder has been developed by combining it with the Inception V3 module, which is adopted in the second-best solution (Antonio et al., 2018). It adds a sparsity penalty function that helps to minimize the overfitting issue. This solution improves the feature extraction process by reducing the number of parameters in the network by reconstructing the image after passing to reverse module. Hence, it enhances the classification accuracy by 5% on average and reduced the processing time by 20~30 milliseconds on average. For future work, a large amount of CLE image dataset can be used for more feature extraction from CNN network to train the module effectively. In future work, a large amount of CLE image dataset can be used for training and testing process of CNN. The network can even be run with the multi-variate dataset with 3D visualisation to produce the exact classification result.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 496-504, May-2026
https://doi.org/10.47001/IRJIET/2026.105069

Appendix:

RISA	Reconstruction independent subspace anal-ysis (RISA).
CNN	Convolutional Neural Network
FCN	Fully convolutional neural networks
VOI	Volume of Interest
LBP-TOP	Rotation invariant uniform pattern local binary pattern on three orthogonal planes
ReLU	Activations on the convolutional layers are rectified linear units
fc-CNN	Fully convolutional CNN

REFERENCES

[1] D. K. Das, S. Bose, A. K. Maiti, B. Mitra, G. Mukherjee, and P. K. Dutta, "Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis," *Tissue and Cell*, vol. 53, pp. 111-119, 2018/08/01/ 2018.

[2] P. R. Jeyaraj and E. R. Samuel Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm," *Journal of Cancer Research and Clinical Oncology*, 2019/01/03 2019.

[3] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68-71, 2018/06/01/ 2018.

[4] M. Aubreville et al., "Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning," *Sci Rep*, vol. 7, no. 1, pp. 11979-11979, 2017.

[5] V. A. A. Antonio, N. Ono, A. Saito, T. Sato, M. Altaf-Ul-Amin, and S. Kanaya, "Classification of lung adenocarcinoma transcriptome subtypes from pathological images using deep convolutional networks," (*in eng), International journal of computer assisted radiology and surgery*, vol. 13, no. 12, pp. 1905-1913,2018.

[6] D. K. Jain et al., "An approach for hyperspectral image classification by optimizing SVM using self organizing map," *Journal of Computational Science*, vol. 25, pp. 252-259, 2018/03/01/ 2018.

[7] E. Yilmaz, T. Kayikcioglu, and S. Kayipmaz, "Computer-aided diagnosis of periapical cyst and keratocystic odontogenic tumor on cone beam computed tomography," *Computer Methods and Programs in Biomedicine*, vol. 146, pp. 91-100, 2017/07/01/ 2017.

[8] W. De Vos, J. Casselman, and G. R. J. Swennen, "Cone-beam computerized tomography (CBCT) imaging of the oral and maxillofacial region: A systematic review of the literature," *International Journal of Oral and Maxillofacial Surgery*, vol. 38, no. 6, pp. 609-625, 2009/06/01/ 2009

[9] C. Jaremenko et al., "Classification of Confocal Laser Endomicroscopic Images of the Oral Cavity to Distinguish Pathological from Healthy Tissue," *in Bildverarbeitung für die Medizin* 2015,

[10] Berlin, Heidelberg, 2015: Springer Berlin Heidelberg, pp. 479-485. M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," *in Computer Vision – ECCV* 2014,

[11] Cham, 2014: Springer International Publishing, pp. 818-833. Y. Ariji et al., "Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, 2018/10/15/ 2018.

[12] J. Sun et al., "Computed tomography versus magnetic resonance imaging for diagnosing cervical lymph node metastasis of head and neck cancer: a systematic review and metaanalysis," *OncoTargets and therapy*, vol. 8, pp. 1291-1313, 2015.

[13] S. Liang et al., "Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning," *European Radiology*, vol. 29, no. 4, pp. 1961-1967, 2019/04/01 2019.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," (*in eng), IEEE Trans Pattern Anal Mach Intell*, vol. 38, no. 1, pp. 142-58, Jan 2016.

[15] B. Song et al., "Automatic classification of dual-modalilty, smartphone-based oral dysplasia and malignancy images using deep learning," *Biomed. Opt. Express*, vol. 9, no. 11, pp. 53185329, 2018/11/01 2018.

[16] K. H. Awan, P. R. Morgan, and S. Warnakulasuriya, "Evaluation of an autofluorescence based imaging system (VELscope™) in the detection of oral potentially malignant disorders and benign keratoses," *Oral Oncology*, vol. 47, no. 4, pp. 274-277, 2011/04/01/ 2011.

[17] W. Poedjiastoeti and S. Suebnukarn, "Application of Convolutional Neural Network in the Diagnosis of Jaw Tumors," *Healthcare Informatics Research*, pp. 236-241, 2018.

[18] I. Sturm, S. Lapuschkin, W. Samek, and K. R. Muller, "Interpretable deep neural networks for single-trial EEG classification," *(in eng), Journal of neuroscience methods*, vol. 274, pp. 141145, Dec 1 2016.

[19] M. Halicek et al., "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," *Journal of biomedical optics*, vol. 22, no. 6, pp. 60503-60503, 2017.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 496-504, May-2026
https://doi.org/10.47001/IRJIET/2026.105069

[20] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of biomedical optics*, vol. 19, no. 1, pp. 10901-10901, 2014.

[21] J. Folmsbee, X. Liu, M. Brandwein-Weber, and S. Doyle, "Active deep learning: Improved training efficiency of convolutional neural networks for tissue classification in oral cavity cancer," *in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 770-773.

[22] L. Yisheng, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, Traffic Flow Prediction With Big Data: A Deep Learning Approach. 2014, pp. 865-873.

[23] E. Rodner et al., "Fully convolutional networks in multimodal nonlinear microscopy images for automated detection of head and neck carcinoma: Pilot study," *Head & Neck*, vol. 41, no. 1, pp. 116-121, 2019/01/01 2019.

[24] A. M. Bur et al., "Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma," *Oral Oncology*, vol. 92, pp. 20-25, 2019/05/01/ 2019.

[25] R. K. De Silva, B. S. M. S. Siriwardena, A. Samaranayaka, W. A. M. U. L. Abeyasinghe, and W. M. Tilakaratne, "A model to predict nodal metastasis in patients with oral squamous cell carcinoma," *(in eng)*, *PloS one*, vol. 13, no. 8, pp. e0201755-e0201755, 2018.

[26] S. Heuke et al., Multimodal nonlinear microscopy of head and neck carcinoma - toward surgery assisting frozen section analysis. 2016.

[27] M. Nishio et al., "Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning," *PloS one*, vol. 13, no. 7, pp. e0200721e0200721, 2018.

[28] F. Ciompi et al., "Towards automatic pulmonary nodule management in lung cancer screening with deep learning," Scientific reports, vol. 7, pp. 46479-46479, 2017.

[29] J. Behrmann, C. Etmann, T. Boskamp, R. Casadonte, J. Kriegsmann, and P. Maaβ, "Deep learning for tumor classification in imaging mass spectrometry," *Bioinformatics*, vol. 34, no. 7, pp. 1215-1223, 2017.

[30] K. He, X. Zhang, S. Ren, and J. Sun, Deep Residual Learning for Image Recognition. 2016, pp. 770-778.

[31] Y. Song et al., Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. 2018.

[32] Y. Feng, L. Zhang, and Z. Yi, "Breast cancer cell nuclei classification in histopathology images using deep neural networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 2, pp. 179-191, 2018/02/01 2018.

Citation of this Article:

Ashphak Khan, Anamika Patil, Nikita Borse, & Ashwini Shinkar. (2026). AI-Powered Oral Cancer Detection System Using Deep Learning. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(5), 496-504. Article DOI https://doi.org/10.47001/IRJIET/2026.105069

****