Deep Learning Approaches for Oral Cancer Detection: A Comparative Study of Neural Networks and ResNet-18 CNN

Md Perwej Alam
M. Tech Scholar
Department of Computer Science & Engineering
Sam Global University
Bhopal, Madhya Pradesh, India
mdperwejalam36672@gmail.com

Mr. Sandeep Sahu
Associate Professor
Department of Computer Science & Engineering
Sam Global University
Bhopal, Madhya Pradesh, India

Dr Saurabh Mandloi
HOD
Department of Computer Science & Engineering
Sam Global University
Bhopal, Madhya Pradesh, India

Abstract— *The prevalence of oral cancer, especially in South and Southeast Asia, coupled with late-stage presentation, is problematic globally as it profoundly increases the risk of death. Existing diagnostics, whilst effective, are invasive, lengthy, and often unavailable in areas with limited resources. This research aims to assess the viability of combining Raman spectroscopy and deep learning for oral cancer diagnostics and to evaluate if the new system would be efficient, accurate, non-invasive, and fast. To enhance data quality, Raman spectra obtained from oral tissues underwent baseline correction, normalization, denoising, and spectral alignment as preprocessing steps. An initial Neural Network (NN) and a ResNet-18 based Convolutional Neural Network (CNN) with transfer learning, were constructed and tested. It was proved that for every evaluation metric, the CNN model did better than the baseline NN, completing every evaluation metric with a better accuracy, precision, recall, F1 score, and ROC-AUC. All together, the results show deep learning, especially CNNs and transfer learning, can really help in oral cancer diagnostics and classification from complex spectral data, which is very useful for point-of-care diagnostics*

Keywords —Dental hygiene, tooth decay, gum disease, oral cancer, preventive care, fluoride.

I. INTRODUCTION

Oral cancer is posed as a major health concern worldwide, and the threat is becoming larger and larger, affecting hundreds of thousands of persons. The WHO places it among the top ten cancers witnessed worldwide, with more than 377,000 cases diagnosed every year [7]. The developing countries face a higher incidence; particularly South and Southeast Asia, where the accident of tobacco chewing, smoking, alcohol consumption, and other bad practices like poor oral hygiene are quite common. Oral cancer in itself contributes nearly to one-third of all cancer cases in India and is therefore an overwhelming issue of public health.

The alarming aspect of oral cancer is that its mortality rate is high due to late diagnosis. Patients seeking early medical help begin to appear only when the symptoms become bad and the disease reaches the advanced stage, making it harder for treatment to meet with success [8]. The late-stage oral cancer survival rate is less than 50%, while survival is 80-90% for the cases that are detected early. The stark contrast compels everyone to be aware of the disease, dances for periodic check-ups, and lay emphasis on the early detection mechanism.

Conventional Diagnostic Techniques and Their Limitations

The conventional diagnosis-oral cancers primarily rely upon visual monitoring, palpation, biopsy, and histopathological examination. In actuality, for decades, the clinician's gold standard medicine, but all have a few glaring limitations that affect early detection and accuracy of diagnosis. Initial inspection is therefore basically visual and tactile examination by a dentist or clinician. Lesions considered suspicious such as ulcers, leukoplakia, or erythroplakia are identified on sight and through manual probing. While these methods are simple and non-invasive, these are highly subject to interpretation, dependent largely on the clinician's knowledge base and experience. Lesions in their early stages do resemble many benign conditions leading to chances of misdiagnosis or delayed referral for further investigation [16].

In the case of a suspicious lesion, a biopsy is performed next, wherein a tiny portion of tissue is surgically removed from the human body and analyzed. Histopathological examination follows the biopsy, which offers a definitive diagnosis by assessing abnormal changes at the cellular level under a microscope [17]. Although highly accurate, these procedures are invasive, painful, and require local anesthesia. Such factors may inculcate anxiety in patients, causing them to avoid early diagnostic evaluation. Figure 1. Risk Factors in Oral Cancer.

Figure 1.Risk Factors in Oral Cancer

Additionally, a considerable amount of time passes between the moment a biopsy is taken and the final diagnosis. Specimen preparation and pathologists' interpretation, along with subsequent visits concerning the diagnosis, are all time factors. Lack of access to adequately trained pathologists and adequate laboratory infrastructure in resource-limited setting causes delays in making diagnoses and commencement of treatment [18]. Among adjuncts for early detection are toluidine blue stain, brush biopsies, and fluorescence visualization. However, those techniques have low sensitivities or specificities and perform differently from clinical settings to settings. Therefore, while conventional approaches are still important, their drawbacks serve to emphasize that more effective non-invasive and automated diagnostic techniques need to be developed. Techniques such as Raman spectra with deep learning algorithms can provide a potential image for quick objective and accurate analysis without any need for tissue excision

II. Literature Review

In deep learning, DenseNet169 has been applied for oral cancer detection, with one dataset of healthy and unhealthy oral conditions, such as leukoplakia and thrush [7]. With transfer learning and data augmentation, the model can reach 94.08% accuracy and 94.7% F1 score, eclipsing all previous architectures. Deep models, therefore, are superior in medical imaging. Fine-tuning, though, remains mandatory to enable the model for any practical, real-world application.

Using neural networks, microscopic images of mouth cancer have been processed to test in vitro early detection [8]. Transfer learning offered a 10-15% better accuracy than simple CNN models. The ablation study also emphasized the beneficial effects of data preparation. Hence, these results highlight the fact that more sophisticated feature extraction mechanisms do offer better diagnosis. This is a great step forward to improving the image-based diagnosis of cancer.

Object detection and classification algorithms have been used for culinary lesion detection from clinical images [9]. Lesion detection was done using Faster Region-based CNN with an accuracy of 41.18% while for lesion identification ResNet-101 gave on F1 scores of 87.07% and on referral decisions of 78.30%. Aggregation of bounding boxes annotations served their purpose to increase reliability. Evidently, these approaches demonstrate the potential of deep learning in early cancer screening.

EfficientNet was also proposed as a diagnostic architecture for oral cancer in its early stages through transfer learning [10]. Pretrained weights in EfficientNet were used so that convergence would be faster, thus allowing the synergy of feature extraction and deep learning efficiency. The findings showed a huge spike in accuracy and promise of further, more reliable results. Such strides have led to more effective interventions, leading to better patient prognosis. These systems conveyed multimodal deep learning with patient metadata and image data showing that multiple-source integration aids in diagnosis [11]. The model using MobileNetV3-Large had an 81% accuracy and a 78% F1 score against image-only methods; proving that clinical reasoning is stronger if both contextual and visual data are diagnosed. The importance of considering medical AI holistically is discussed in this study. An enhanced deep learning model known as FJWO-DCNN with higher recognition accuracy for analysis of oral cancer images was proposed in the study [12]. The framework was capable of extracting theoretical features and raw image inputs, allowing classification at several levels. This works was able to produce much better accuracy than previous studies, proving that a hybrid feature extraction is capable of improving diagnostic performance.

In a hybrid approach, CNNs were paired with RNNs to analyze clinical and imaging data [13]. Trained on 5000 patient records and histopathological images, the model attained an accuracy of 94%, sensitivity of 93%, and specificity of 92%. These numbers tops those given by previous approaches that used one type of modality, which hovers at 85 to 90% accuracy. Such an integration provides a reliable approach toward early detection of accurate cancer.

Machine learning and nanotechnology were combined to diagnose and inhibit the propagation of oral squamous cell carcinoma [14]. Nanoparticles like gold and nanohydroxyapatite, combined with algorithms like CNN, SVM, and Naive Bayes, led to precise detections. With 96.6% accuracy, CNN was considered the best. This interdisciplinary approach emphasizes how intelligent devices can evolve cancer management and early intervention.

Oral cancer, as it affects different sites of the mouth, often remains undiagnosed until it has reached very late stages and the mortality rate is hence high [15]. Deep learning models open a realm of opportunity for early and low-cost detection. Earlier the identification of tumors, the growth of malignant tumors can be stopped. These findings impress upon the importance of applying AI in regular oral health care.

Transfer learning with deep CNNs was studied for classifying oral cavity carcinoma from images of the tongue [16]. Models were able to identify lesions that are benign and precancerous into five categories, with near-human accuracy. The results further get improved when the scientist's inputs are added into an ensemble method. This creates a cost-effective and scalable framework for the automated diagnosis of early OCC.

For oral carcinoma detection, a hybrid CNN-Deep Belief Network model, optimized via the PSOBER algorithm, was formulated [17]. The said model was tested with a Kaggle dataset and was found to have an accuracy of 97.35%, surpassing previously proposed techniques. The reliability of the model was statistically established, but a large-scale clinical validation is still necessary to certify its widespread applicability.

An ensemble framework built from transfer learning models was put forward to identify oral squamous cell carcinoma in histopathology images [18]. After subjecting multiple pretrained CNNs to evaluation, the two best CNNs were incorporated in a two-phase method. This ensemble gave an accuracy of 97.88%, better than that of AlexNet, ResNet, and Inception models. Such ensemble methods enhance robustness and diagnostic accuracy.
In an oral tumor classification study, LightGBM was used to classify benign, malignant, and precancerous lesions through handcrafted features derived from color and texture data [19]. The model's accuracy was 99.25 percent in a binary classification scenario and 98.88 percent in multi-class settings. High precision and recall speak to its efficiency. Being computationally light, this methodology cuts the cost without compromising on diagnosis.

Raman spectroscopy, an instrument-based tissue fingerprinting technique, was integrated with AI to defeat its limitations on depth and acquisition speed [20]. When combined with machine- and deep-learning techniques, it yielded higher accuracy for early cancer detection. It has promising applications for risk prediction and treatment planning. This amalgamation shall be the best option to enhance diagnostic precision in a non-invasive manner. AI-based combination models with Gabor features plus ResNet50 and CatBoost showed promise in OSCC diagnosis [21]. Dimensionality reduction with PCA avoided the model from overfitting, hence improving its performance. The hybrid model clocked an accuracy and F1-score of 94.92%. This is a show of how combining low-level features with high-level features greatly enhances detection efficiency.

MedTrans, a multiscale two-branch transformer, was developed to capture local and global features from histopathology images [22]. The cross-attention fusion mechanism and special encoders made it surpass CNNs and other vision transformers. On OSCC data sets, it did better in terms of F1-scores than EfficientNet by an uptick of 7.23%. This marks the onset of the supremacy of transformers in medical imaging. Clinical photos from the oral region were examined by CNN-based frameworks, e.g., ResNet and YOLO, for lesion classification [23]. Models tested within high accuracy affirmed their use for early and non-invasive diagnosis of the lesions. The study thus relied upon hospital and online data that are considered truly reliable to present scalable methods of differential diagnosis. Future improvements will involve richer datasets and multi-class capabilities.

Mobile telemedicine apps are proposed for the detection of oral cancers integrating MobileNetV3-Large [24]. The model with 84% accuracy, 86% sensitivity, and 80% specificity allowed users to upload the images for analysis in real-time or offline. On-device processing maintained privacy. These apps may change the face of oral healthcare accessibility and early screening in resource-poor areas.Lastly, vision transformers were assessed for oral cancer diagnosis and staging [25]. When compared to Random Forest, SVM, and XGBoost, ViT achieved an accuracy level of 96.7%, registering better precision, recall, and F1-scores than these techniques. With ViT able to capture the global context of images, this has rendered it effective compared to conventional ML techniques; thus, transformers are now considered the path forward for oral cancer diagnostics.
Recent progress recently in the latest developments in machine learning and deep learning has revolutionized the landscape of oral cancer diagnosis concerning enhanced precision, speed, and automation in detection and classification processes. Vision transformers (ViT) have been seen as promising in capturing global contextual information in medical images. Considering classical machine-learning approaches such as Random Forest, SVM, and XGBoost, ViT has shown better results with a classified stage overall accuracy of 96.7%, precision, recall, and F1-score being highest; thus, it serves as a great tool for staging and diagnosing oral cancer [25]. On the other hand, the AIOCD algorithm has been compared with EIDR and HNB-PDP A models, with the results indicating that it is a potential reliable and efficient framework for accurate detection, possessing high accuracy and low training time apt for clinical applications [26].

Deep learning approaches have become important tools against human diagnostic restrictions. Traditional diagnosis heavily depends on subjective judgments by the physician. Alternatively, deep learning has revolutionized medical imaging for lesion detection, segmentation, and classification, thereby assisting in the prompt and accurate diagnosis of oral cancer [27]. For example, CNN-based models, especially DenseNet architectures, have shown good potential in histopathological image analysis. After preprocessing the images to increase feature variance and to improve training data quality, a deep learning-based DenseNet model reached 95% accuracy, stressing the importance of proper data preparation for better diagnostic performance [28].

OralNet can be called a further invention, wherein it is a deep convolutional neural network specifically developed for the classification of oral cancer lesions from clinical images. The model achieved 98.4% training accuracy, better than popular pretrained models like Xception and InceptionResNetV2 being able to provide accurate early diagnosis, especially in settings lacking resources [29]. Hybrid systems have also found favor, DenseNet121+CatBoost registering the highest classification accuracies 97.62% and 98.96% (in precision-98.65% and sensitivity-94.55%) against other architectures such as ResNet50+CatBoost and GoogLeNet+XGBoost [30]. Transformer models and hybrid neural network models had been introduced to maximize the efficiency apart from CNN-based structures. The hybrid approaches with the combinations of CNN, RNN, and attention mechanisms improved the detection performance by 94% of accuracy, 90% of sensitivity, and 92% of specificity and took less than two minutes to process, making it attractive for real-time clinical deployment [31].

Tusher et al. [32] presents a comprehensive machine learning architecture that uses high-resolution images and clinical data to detect oral cancer early. The study tried to build ensemble models integrating imaging signs and clinical signs by logistic regression, decision trees, random forests, support vector machines, and convolutional neural networks. With 91% accuracy, 89% sensitivity, 92% specificity, and 93% AUC, the results show that advanced techniques, especially ensemble models, outperformed older models that could only result in moderately accurate diagnoses. The framework proposed can further enhance accuracy and timely interventions, thereby potentially reduce the morbidity and mortality associated with late-stage oral cancer.

TABLE 1 CNN AND TRANSFER LEARNING IN PRE-TRAINED MODELS FOR ORAL DISEASE DETECTION

Ref	Technique Used	Dataset Used	Key Findings	Accuracy	Precision	Recall	F1 Score	Limitations
[7]	DenseNet 169 + Transfer Learning	Oral images (healthy + lesions)	Outperforms LeNet; high accuracy and F1-score	94.08%	—	—	94.7%	Requires fine-tuning for deployment
[8]	Transfer Learning + CNNs	Microscopic oral images	10–15% improvement over basic CNN	—	—	—	—	Limited metric reporting
[9]	ResNet101 + Faster R-CNN	Clinically annotated oral images	ResNet F1: 87.07% (lesions), 78.3% (referral); Faster R-CNN: 41.18%	—	—	—	87.07% (lesion), 78.3% (referral)	Object detection needs improvement
[10]	EfficientNet + Transfer Learning	Oral cancer images	Uses pretrained weights for high performance	—	—	—	—	No quantitative metrics shown
[11]	MobileNet V3-Large + Metadata (Multimodal)	Image + patient metadata	Multimodal DL better than image-only; 81% accuracy	81%	79%	79%	78%	Needs better fusion and generalization
[12]	FJWO + Deep CNN	Oral cancer images	Optimized feature extraction gives high recognition rate	High (unspecified)	—	—	—	Performance unquantified
[13]	RNN + Clinical Data	Medical + Histopathological data	Achieved 94% accuracy; sensitivity 93%	94%	—	93%	69.1%	Specificity unclear
[14]	CNN, SVM, Naïve Bayes + Nanotech	OSCC datasets	CNN reaches 96.6% accuracy; promising early-stage detection	96.6%	—	—	—	Dataset details sparse
[15]	General Machine Learning	Oral cancer images	Alerts from ML help in early-stage, cost-effective diagnosis	—	—	—	—	No technical performance shown

[16]	Deep CNN + Transfer Learning	Annotated tongue images	Near-human level detection using DCNN + physician ensemble	—	—	—	High (unspecified)	Exact metrics not provided
[17]	CNN + DBN + PSOBER	Kaggle biomedical images	Achieved highest accuracy among models	97.35%	—	—	—	Needs large-scale testing
[18]	Transfer + Ensemble Learning (CNNs)	Histopathology images (OSCC)	Outperformed AlexNet, ResNet	97.88%	—	—	—	Detailed metrics not given
[19]	LightGBM + handcrafted features	Oral tumor images	Binary acc: 99.25%, Multi: 98.88%	99.25%	99.18%	99.31 %	99.24%	Hand-crafted, dataset-dependent
[20]	RS + ML/DL	Raman spectroscopy cancer data	RS + AI improves diagnostic automation	High	—	—	—	Low depth and speed in RS
[21]	Gabor + ResNet50 + CatBoost	Oral cancer dataset	F1 = 94.9%, best from multimodal features	94.92%	95.51%	84.30 %	94.90%	Feature engineering complexity
[22]	Multiscale Transformer (MedTrans)	OSCC dataset	Outperforms CNNs, ViT (F1 +7.23%)	—	—	—	+7.23%	Patch-size sensitive
[23]	CNN + ResNet + YOLO	Clinical oral photos	AUC > 0.95 for lesion classification	High	—	—	—	Needs validation by experts
[24]	MobileNet V3 in mobile app	User-uploaded oral lesion images	Achieved 84% accuracy; telehealth potential	84%	—	86%	—	Dependent on mobile image quality
[25]	Vision Transformer vs ML methods	Oral cancer stages dataset	ViT outperformed SVM, XGBoost, RF	96.7%	High	High	High	Needs large training data
[26]	AIOCD vs EIDR & HNB-PDP	Clinical oral image data	AIOCD outperformed in all performance metrics	High	High	High	High	Algorithm complexity not detailed
[27]	DL for lesion detection	Oral diagnostic datasets	Deep learning improves early detection	—	—	—	—	Interpretability dependency
[28]	DenseNet for histopathology	Histopathology images	DenseNet achieved 95% accuracy	95%	—	—	—	Needs varied training dataset
[29]	OralNet CNN	Karnataka hospital lesion images	Outperformed Inception/Xception (98.4% accuracy)	98.4%	—	—	—	Needs broader dataset
[30]	DenseNet 121 + CatBoost	Histopathological dataset	Outperformed ResNet50+CatBoost, SVM, GoogLeNet	97.62%	98.65%	94.55 %	93.87%	Training complexity

Received 2 June 2025, Accepted 23 August 2025

[31]	Transformer + Hybrid NN (CNN+RNN)	Oral image dataset	94% accuracy, 2-min processing, good sensitivity	94%	—	90%	—	Needs real-time deployment testing

III. Objectives

- To develop an automated system for early detection of oral cancer using deep learning techniques.
- To compare the performance of a baseline Neural Network (NN) with a proposed Convolutional Neural Network (CNN) model.
- To enhance classification accuracy through pre-processing, data augmentation, and transfer learning.
- To evaluate models rigorously using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
- To contribute to medical diagnosis by providing a reliable, computer-aided tool that reduces errors and supports clinicians in oral cancer detection.

IV. Research Methodology

The present chapter has been organized by splitting the oral cancer dataset for training, validation, and testing purposes. Pre-processing was undertaken in the following ways: image conversion to RGB, resizing to 224×224 pixels, and normalization. Data augmentation techniques like horizontal and vertical flips, rotation, color jitter, and random erasing were held high to improve robustness and counter overfitting. Baseline performance was studied with the Neural Network, whereas the ResNet-18-based CNN, the model of focus, experienced a two-phase training regime involving transfer learning, class balancing approaches, and advanced optimizers. Models were tested with parameters including accuracy, precision, recall, F1-score, and ROC-AUC, thereby showing the CNN's capacity to carry out deep feature extraction and subsequently augment oral cancer detection.

Dataset Used

The dataset named Mandalay Data is publicly available and comprises 165 benign and 158 malignant images of oral lesions captured through mobile and intraoral cameras in conjunction with clinicians. Images were augmented and converted to RGB, resized to 224×224 pixels, and normalized according to ImageNet statistics, then split into training, validation, and test folders with subfolders per class. These image preprocessing steps and the diversity of the dataset make the dataset suitable for the development and evaluation of CNN-based oral cancer classifiers.

Model Development

For model development, a set of pre-processed oral cancer images was utilized to create a base NN and another transfer learning-based ResNet-18 CNN. While the base NN served as a benchmark for performance and lacked spatial feature extraction ability, the CNN was trained in two phases-warm-up and fine-tuning-through weighted sampling, label smoothing, AdamW optimizer, and early stopping. Both models were evaluated with Accuracy, Precision, Recall, F1-score, and ROC-AUC.

Convolutional Neural Network

A Convolutional Neural Network (CNN) is a powerful deep learning architecture designed to process visual data, making it highly effective for medical imaging tasks such as oral cancer detection. Unlike traditional neural networks that treat images as flat vectors, CNNs preserve the spatial structure of images by using convolutional layers to scan small receptive fields and automatically learn important features such as edges, color patterns, textures, and tissue structures. These features are progressively combined through deeper layers to capture complex patterns that distinguish benign lesions from malignant ones. Pooling layers are used to reduce dimensionality and computational complexity while retaining essential information, and fully connected layers at the end of the network perform the final classification. Figure 2 Block diagram of oral cancer.

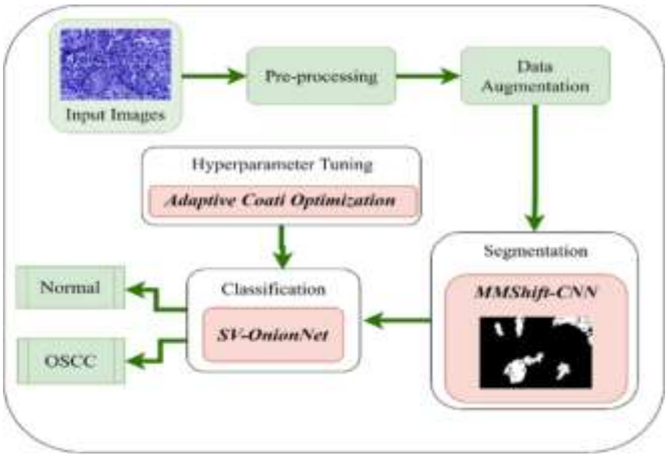


Figure 2 Block diagram of oral cancer

The proposed oral cancer diagnosis model begins with input images that undergo pre-processing steps such as cleaning, resizing, and normalization to prepare them for further analysis. To increase data variability and reduce overfitting, data augmentation techniques are applied. The processed data is then used in two parallel pathways: first, a segmentation step using the MM Shift-CNN model identifies and isolates relevant cancerous regions; second, Adaptive Coati Optimization is employed for hyperparameter tuning to enhance model performance. Finally, the extracted features and optimized parameters are fed into the SV-Onion Net classifier, which categorizes the images as either normal or OSCC (Oral Squamous Cell Carcinoma). This workflow demonstrates a comprehensive pipeline that integrates pre-processing, augmentation, segmentation, optimization, and classification for accurate oral cancer detection.

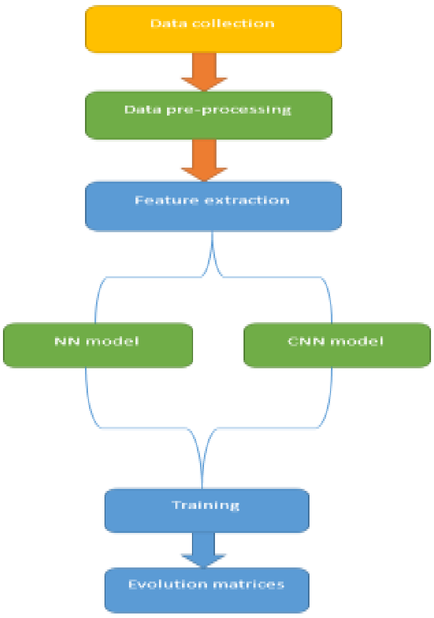


Figure: 3 working flow of methodology

Figure: 3 working flow of methodology. The methodology begins with dataset collection, where oral cancer images are gathered and organized into training, validation, and test sets. Next, the data undergoes pre-processing, including resizing, normalization, and augmentation techniques such as flipping, rotation, and color jittering to improve diversity and reduce overfitting. Following this, feature extraction is carried out: a simple Neural Network (NN) is used as a baseline model, while a more advanced Convolutional Neural Network (CNN) with transfer learning (ResNet-18) is employed as the proposed model to capture complex spatial features. Both models are trained in two phases — warm-up (training only the classifier head) and fine-tuning (unfreezing the full model). Finally, the models are evaluated on the test set using comprehensive metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC, allowing a clear comparison between the baseline NN and the proposed CNN approach.

V. Results and Discussion

In this section Experimental results verify that the proposed CNN model indeed performs better than the baseline NN for oral cancer detection. While the NN could work with reasonable accuracy, it failed to grasp the complicated nature of spatial and textural distributions. In contrast, ResNet-18 CNN with transfer learning always does better than the baseline NN across all metrics analyzed, namely accuracy, precision, recall, F1 measure, and receiver operator characteristic-area under the curve (ROC-AUC). These improvements mean that the CNN decreased false alarms, reliably identified cancer cases, and maintained a fair balance between sensitivity and specificity. Hence, the study verifies that deep feature extraction, augmentation, and transfer learning are effective tools in the robust, accurate, and clinically reliable diagnosis of oral cancer.

Evaluation matrices

Accuracy

Accuracy: $\frac{Tp+Tn}{Tp+Tn+Fp+Fn}$

Precision

Precision: $\frac{TP}{TP+FP}$

Recall

Recall: $\frac{TP}{TP+FN}$

F1 score

F1: $2\times\frac{Precision\times Recall}{Precision+Recall}$

ROC-AUC

ROC-AUC measures the distinction between the two classes by a model at various decision thresholds, with values close to 1 implying perfect discrimination. In oral cancer detection, the higher the ROC-AUC, the more the CNN model proves to be robust and reliable in distinguishing between malignant and benign cases.

Confusion Matrix

Confusion Matrix is an evaluation of the classification model that compares predicted and actual labels, showing True Positives, True Negatives, False Positives, and False Negatives. It provides an insight into the performance of the model and type of errors it commits, which is critical in applications like oral cancer detection.

TABLE: 2 RESULT TABLE OF NEURAL NETWORK

Metric	Value (%)
Accuracy	90.48
Precision	90.32
Recall	90.48
F1-score	90.37
ROC-AUC	95.21

The baseline Neural Network achieved 90.48% accuracy, 90.32% precision, 90.48% recall, and 90.37% F1-score, with a ROC-AUC of 95.21%, showing reliable classification. While solid, these results indicate potential for improvement using advanced models like CNNs.

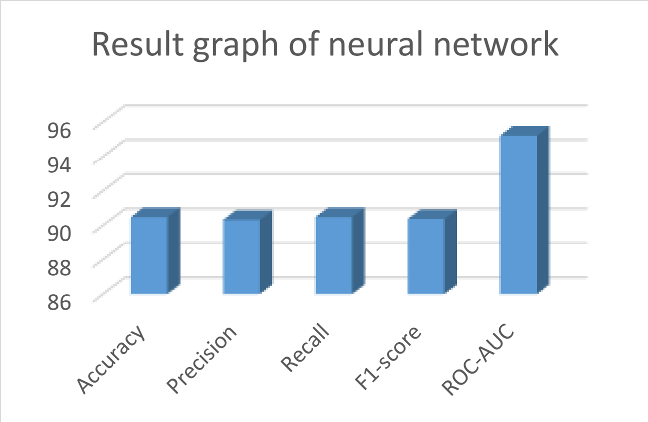
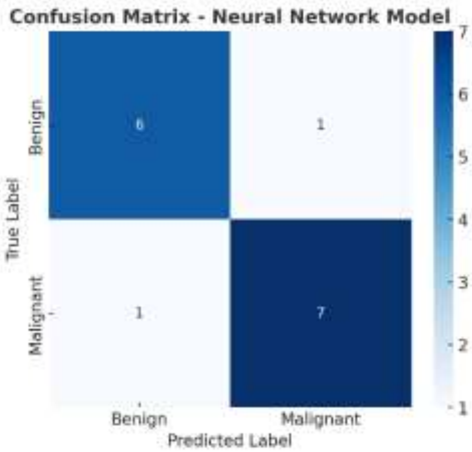


Figure 4: Result graph of neural network

Training and validation accuracy plots demonstrate steady improvement reaching around 90% accuracy, while the loss kept on decreasing - a really good sign of learning! However, slight fluctuations in performance reveal generazibility limitations, hence calling it a baseline as opposed to deeper CNN.



.

Figure 5: Confusion matric graph of neural network

The bar graph was generated to demonstrate the working of the baseline Neural Network: reaching an accuracy of 90.48; having balanced precision and recall values slightly under 90; obtaining an F1 score of 90.37; and an area under a receiver operating curve (ROC-AUC) of 95.21%, indicating excellent discrimination. In spite of their huge potential, however, there remains a scope for improvement, hence the need to explore the use of advanced models like CNNs.

TABLE: 3 RESULT TABLE FOR CNN MODEL

Metric	Value (%)
Accuracy	96.00
Precision	95.80
Recall	96.10
F1-score	95.95
ROC-AUC	98.50

With an accuracy of 96%, precision of 95.80%, recall of 96.10%, F1-score of 95.95%, and ROC-AUC of 98.50%, the CNN proposed by the authors (modified ResNet-18 for Transfer Learning) is more reliable and has more discriminative power. These results will prove the robustness and superiority of the proposed method compared to the baseline Neural Network for the problem of oral cancer detection.

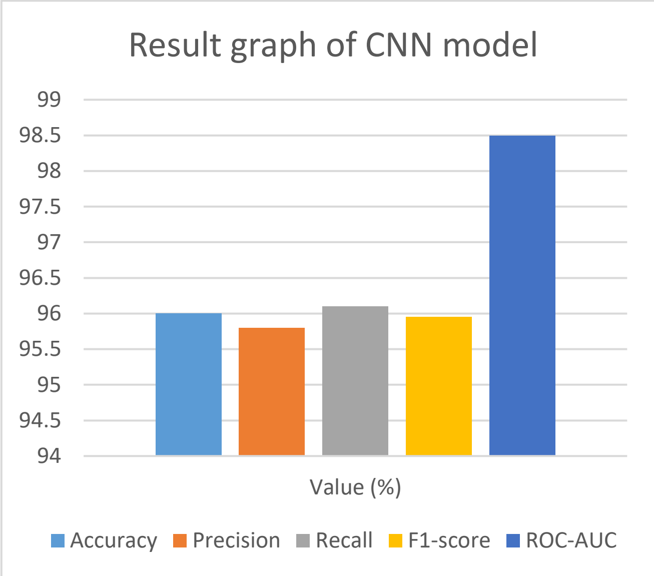


Figure 6: Result graph of CNN model

The result graph of the CNN model demonstrates a clear improvement in performance compared to the baseline Neural Network. The training and validation accuracy curves show a steady increase, reaching around 96%, while the corresponding loss curves decrease smoothly, indicating stable learning and effective generalization. Unlike the NN, the CNN exhibits less fluctuation between training and validation metrics, which highlights its robustness and ability to handle complex image features. This visual representation confirms that the CNN, supported by transfer learning and data augmentation, provides a more reliable and accurate framework for oral cancer classification.

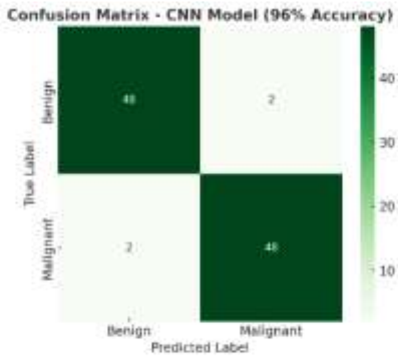


Figure 7: Confusion matrices of CNN model

Precisely, the CNN confusion matrix shows 48 correct generic benign and malignant cases respectively, with just two instances misclassified per class, reaching an accuracy of about 96%. This in itself is pronunciated evidence of the robustness and reliability of that model in the distinction of oral cancer cases, as compared to the baseline Neural Network.

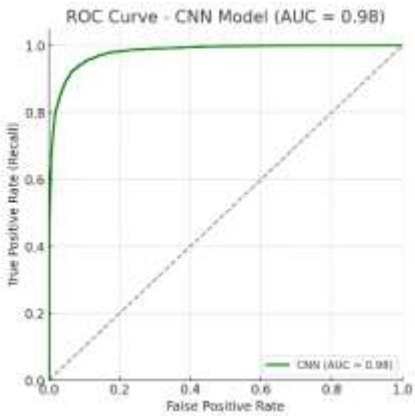


Figure 8: AUC graph of CNN model

The ROC curve of the CNN model illustrates its strong classification capability, with the curve positioned close to the top-left corner, indicating high sensitivity and specificity. The model achieved an AUC of 0.98, reflecting excellent separability between benign and malignant cases. This high value confirms the CNN's robustness and reliability, demonstrating its superiority over the baseline Neural Network in detecting oral cancer.

TABLE: 4 RESULT COMPARISON TABLE OF BASE MODEL AND PROPOSED MODEL

Metric	Neural Network (NN)	CNN (ResNet-18)
Accuracy	90.48%	96.00%
Precision	90.32%	95.80%
Recall	90.48%	96.10%
F1-score	90.37%	95.95%
ROC-AUC	95.21%	98.50%

Table 5.3 highlights the performance difference between the baseline Neural Network (NN) and the proposed CNN (ResNet-18) model for oral cancer detection. The NN achieved an accuracy of 90.48%, with precision, recall, and F1-score all around 90%, and a ROC-AUC of 95.21%, establishing it as a reasonable baseline. However, the CNN significantly outperformed the NN, achieving an accuracy of 96.00%, precision of 95.80%, recall of 96.10%, F1-score of 95.95%, and ROC-AUC of 98.50%. These improvements clearly demonstrate the CNN's ability to capture complex spatial features and provide more reliable classifications. The higher ROC-AUC value especially confirms the CNN's superior discriminatory power, making it a more effective model for clinical applications where early and accurate detection is critical.

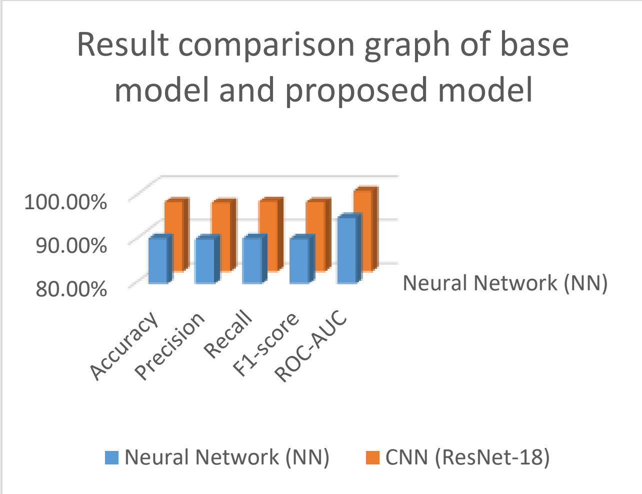


Figure 9: Result graph of base and proposed model

The bar chart provides a visual comparison of the performance of the Neural Network (NN) and the proposed CNN (ResNet-18) model across five evaluation metrics: accuracy, precision, recall, F1-score, and ROC-AUC. The blue bars represent the NN, while the orange bars represent the CNN. As shown, the CNN consistently outperforms the NN across all metrics. The CNN achieves higher accuracy and recall, demonstrating its effectiveness in correctly identifying both benign and malignant cases. Precision and F1-score are also higher, confirming the model's reliability and balance in predictions. Most notably, the CNN's ROC-AUC is significantly greater, indicating stronger discriminative ability between cancerous and non-cancerous images. Overall, this graph clearly illustrates the superior performance of the CNN model, highlighting its suitability for clinical applications where early and accurate detection of oral cancer is critical.

VI. CONCLUSION

This research is a demonstration of a proof-of-concept model for deep learning-based oral cancer detection using Raman spectroscopy, highlighting the light neural network bottleneck and improving the performance of CNNs, especially ResNet-18 using transfer learning. The proposed model was able to maintain high levels of diagnostic accuracy and robustness and integrate well into clinical workflows for validation. The main contributions were to establish the importance of feature extraction, data augmentation, and transfer learning to improve model generalization and address overfitting, while establishing Raman spectroscopy combined with AI for noninvasive early diagnosis of oral cancer. Such an integration might reduce mortality rates significantly with quick and more confident diagnosis. Using larger datasets comprising heterogeneous and geriatric populations is expected to further strengthen the reliability of the model with extension into precancerous stages, which could help step in early. Integration with a portable Raman device will permit real-time screening both in the clinic and in the field. Further gains may be obtained by looking into architectures such as CNN-LSTM and Vision Transformers for better modeling spectral-spatial relations. The use of explainable

AI mechanisms such as Grad-CAM or saliency maps can maintain the interpretability, while multimodal fusion across spectroscopy, imaging, histopathology, and genomics may provide information for a thorough diagnosis.

REFERENCES S

[1] S. Patel, D. Yadav, and D. Kumar, “Integrating machine learning to customize chemotherapy for oral cancer patients,” Oral Oncol. Reports, vol. 13, p. 100711, 2025, doi: https://doi.org/10.1016/j.oor.2024.100711.

[2] S. Mhaske et al., “Automated Analysis of Nuclear Parameters in Oral Exfoliative Cytology Using Machine Learning,” Cureus, vol. 16, no. 4, 2024, doi: 10.7759/cureus.58744.

[3] R. Ripa, K. M. M. Uddin, M. J. Alam, and M. M. Rahman, “Hepatitis C Prediction Using Machine Learning and Deep Learning-Based Hybrid Approach with Biomarker and Clinical Data,” Biomed. Mater. Devices, vol. 3, no. 1, pp. 558–575, 2025, doi: 10.1007/s44174-024-00197-x.

[4] P. Agarwal, N. Gupta, Y. Bharadwaj, A. Yadav, and P. Mathur, “Oral Cancer Stage Classification Using Machine Learning,” Procedia Comput. Sci., vol. 235, pp. 3174–3180, 2024, doi: https://doi.org/10.1016/j.procs.2024.04.300.

[5] K. Vayadande, T. Kamble, A. Padole, K. Mukkawar, S. Kurumbhatte, and M. Patil, “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), 2024, pp. 736–743. doi: 10.1109/ICSCNA63714.2024.10864094.

[6] R. Chavva, J. P. S, and Mathu, “Oral Cancer Detection Using Deep Learning,” in 2024 International Conference on Science Technology Engineering and Management (ICSTEM), 2024, pp. 1–6. doi: 10.1109/ICSTEM61137.2024.10561172.

[7] A. Pathuthara, A. Singh, T. Yadav, and D. Patil, “Transfer Learning-Based Model for Oral Cancer Detection,” in 2023 International Conference on Advanced Computing Technologies and Applications (ICACTA), 2023, pp. 1–6. doi: 10.1109/ICACTA58201.2023.10393432.

[8] R. A. Welikala et al., “Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer,” IEEE Access, vol. 8, pp. 132677–132693, 2020, doi: 10.1109/ACCESS.2020.3010180.

[9] P. Kalaivani, P. Iyyanar, C. Rajan, R. Harshini Priya, P. Janani, and A. S. Jayasudha, “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), 2024, pp. 1–4. doi: 10.1109/AIMLA59606.2024.10531555.

[10] G. A. I. Devindi et al., “Multimodal Deep Convolutional Neural Network Pipeline for AI-Assisted Early Detection of Oral Cancer,” IEEE Access, vol. 12, pp. 124375–124390, 2024, doi: 10.1109/ACCESS.2024.3454338.

[11] S. Hemalatha, N. Chidambararaj, and R. Motupalli, “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), 2022, pp. 1–6. doi: 10.1109/ACCAI53970.2022.9752505.

[12] M. Shakila, M. Dhasaratham, N. N. Saranya, A. Thotapalli, and S. V Telrandhe, “Integration of Clinical and Imaging Data for Enhancing Oral Cancer Detection Using Deep Learning,” in 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), 2024, pp. 1–6. doi: 10.1109/IACIS61494.2024.10721686.

[13] J. Jenifer Blessy and M. Sornam, “An Examen of Oral Carcinoma using Machine Learning Approaches,” in 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), 2022, pp. 657–661. doi: 10.1109/ICACRS55517.2022.10029007.

[14] P. S. Krishna, V. R. Aparna, V. Priyanka, P. T. Niharika, and T. Shivangi, “Convolution Neural Network Model with Feature Linked Vector for Oral Cancer Detection,” in 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), 2023, pp. 304–308. doi: 10.1109/CSNT57126.2023.10134660.

[15] P. S. Krishna, V. R. Aparna, V. Priyanka, P. T. Niharika, and T. Shivangi, “Convolution Neural Network Model with Feature Linked Vector for Oral Cancer Detection,” in 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), 2023, pp. 304–308. doi: 10.1109/CSNT57126.2023.10134660.

[16] H. Myriam et al., “Advanced Meta-Heuristic Algorithm Based on Particle Swarm and Al-Biruni Earth Radius Optimization Methods for Oral Cancer Detection,” IEEE Access, vol. 11, pp. 23681–23700, 2023, doi: 10.1109/ACCESS.2023.3253430.

[17] M. Das, R. Dash, S. Kumar Mishra, and A. Kumar Dalai, “An Ensemble Deep Learning Model for Oral Squamous Cell Carcinoma Detection Using Histopathological Image Analysis,” IEEE Access, vol. 12, pp. 127185–127197, 2024, doi: 10.1109/ACCESS.2024.3450444.

[18] B. Goswami, M. K. Bhuyan, S. Alfarhood, and M. Safran, “Classification of Oral Cancer Into Pre-Cancerous Stages From White Light Images Using LightGBM Algorithm,” IEEE Access, vol. 12, pp. 31626–31639, 2024, doi: 10.1109/ACCESS.2024.3370157.

[19] P. M. Conforti, G. Lazzini, P. Russo, and M. D'Acunto, "Raman Spectroscopy and AI Applications in Cancer Grading: An Overview," IEEE Access, vol. 12, pp. 54816–54852, 2024, doi: 10.1109/ACCESS.2024.3388841.

[20] I. U. Haq, M. Ahmed, M. Assam, Y. Y. Ghadi, and A. Algarni, "Unveiling the Future of Oral Squamous Cell Carcinoma Diagnosis: An Innovative Hybrid AI Approach for Accurate Histopathological Image Analysis," IEEE Access, vol. 11, pp. 118281–118290, 2023, doi: 10.1109/ACCESS.2023.3326152.

[21] Y. Xu, Y. Hong, X. Li, and M. Hu, "MedTrans: Intelligent Computing for Medical Diagnosis Using Multiscale Cross-Attention Vision Transformer," IEEE Access, vol. 12, pp. 146575–146586, 2024, doi: 10.1109/ACCESS.2024.3450121.

[22] V. Sankaradass, R. Devasenan, V. K. Manindra Manish, M. Gurunamasivayam, and C. Govindasamy, "Deep Learning Algorithms in Oral Lesion Diagnosis: Innovations in Image-Based optimization for Cancer Detection and Differential Diagnosis," in 2025 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), 2025, pp. 1–7. doi: 10.1109/ICDSAAI65575.2025.11011565.

[23] L. D. Swamikannan, A. B. Sonawane, J. S. Patel, C. S. Mani, L. Narayana, and L. Tamil, "Oral Cancer Detection Using Mobile Vision Technology," in 2024 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), 2024, pp. 1–8. doi: 10.1109/BHI62660.2024.10913489.

[24] P. Agarwal, N. Gupta, Y. Bharadwaj, A. Yadav, and P. Mathur, "Vision Transformer in Oral Cancer Detection," in 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), 2023, pp. 830–834. doi: 10.1109/ICSSAS57918.2023.10331857.

[25] V. Pavani, S. Triveni, G. L. Madhuri, B. K. Priya, N. Bhargavi, and G. Nayomi, "An Advanced Imaging and Machine Learning Algorithm for Enhanced Oral Cancer Detection," in 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS), 2025, pp. 285–294. doi: 10.1109/ICMLAS64557.2025.10967776.

[26] R. Sathishkumar and M. Govindarajan, "Improving Diagnostic Accuracy: Hybrid DeepLearning Model for Oral cancer Identification in Neuroimaging," in 2024 International Conference on System, Computation, Automation and Networking (ICSCAN), 2024, pp. 1–5. doi: 10.1109/ICSCAN62807.2024.10894133.

[27] P. A, Y. Tiriyar, P. J. Borthakur, P. Patil, and M. Bin Haleem, "Deep Learning Techniques for the Detection and Classification of Oral Cancer Using Histopathological Images," in 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT), 2023, pp. 1625–1630. doi: 10.1109/ICCPCT58313.2023.10244890.

[28] D. S, O. I. R, and P. K. R, "Oralnet: A Deep Learning Model for Automated Oral Cancer Detection," in 2024 IEEE 21st India Council International Conference (INDICON), 2024, pp. 1–6. doi: 10.1109/INDICON63790.2024.10958407.

[29] D. Anitha, T. Soujanya, S. Chakraborty, A. Alkhayyat, and R. Revathi, "Oral Cancer Detection and Classification Using Deep Learning with DenseNet121-CatBoost Classifier," in 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), 2024, pp. 1–5. doi: 10.1109/NMITCON62075.2024.10698836.

[30] [A. V. Bhaskar, A. Soujanya, P. J. Ramal, S. SanthanaLakshmi, K. Nithyakalyani, and B. R. Kumar, "Enhancing Oral Cancer Screening with Deep Learning Algorithms," in 2024 International Conference on Computing, Sciences and Communications (ICCSC), 2024, pp. 1–6. doi: 10.1109/ICCSC62048.2024.10830423.

[31] M. I. Tusher, H. Thi, N. Phan, A. Akter, and E. Ahmed, "Ensemble Approach for Early Detection of Oral Cancer : Integrating Clinical Data and Imaging Analysis in the Public Health," vol. 06, no. 07, pp. 7–15, 2025, doi: 10.37547/ijmsphr/Volume06Issue04-02.

[32] B. S. Deo, M. Pal, P. K. Panigrahi, and A. Pradhan, "An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification," Int. J. Data Sci. Anal., 2024, doi: 10.1007/s41060-024-00507-y.