International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

Early Detection of Oral Cancer Using EfficientNet

Harshitha T N, Soujanya R
Department of Computer Science and Engineering University of Visvesvaraya College of Engineering
K R Circle, Bengaluru-560001, India

Abstract- Oral cancer is a critical and life-threatening disease, where early detection plays a vital role in improving patient survival rates. However, traditional diagnostic approaches rely heavily on manual clinical examination and biopsy, which are time-consuming, invasive, and often lead to delayed diagnosis. To address these limitations, this paper proposes a deep learning framework for automated oral cancer detection using medical image analysis and lesion-focused classification techniques. The proposed system integrates image preprocessing, lesion segmentation, and deep convolutional neural networks (CNNs) for accurate classification. Preprocessing techniques such as contrast enhancement and noise reduction are applied to improve image quality. Lesion regions are extracted using Otsu thresholding and contour-based segmentation to isolate regions of interest (ROI), which enhances feature learning. Multiple deep learning architectures, including Baseline CNN and EfficientNet-B0 are evaluated for performance comparison. In addition, the proposed framework integrates lesion segmen-tation and deep feature extraction to improve classification robustness and diagnostic performance. To enhance model interpretability, Grad-CAM is employed to visualize the regions contributing to predictions, making the system more transpar-ent for medical applications. Experimental results demonstrate that the proposed EfficientNet-B0 based model achieves superior performance compared to baseline approaches, with improved accuracy and F1-score on the test dataset. The proposed framework provides an efficient, scalable, and interpretable solution for early-stage oral cancer detection, supporting clinical decision-making and reducing diagnostic delays.

Keywords – Faculty Development Programmes, Age Differences, Professional Development, Higher Education, One-Way ANOVA, Life-Cycle Model.

I. INTRODUCTION

Oral cancer is one of the most prevalent and life-threatening malignancies affecting the head and neck region worldwide. It is often diagnosed at advanced stages due to the subtle and asymptomatic nature of early lesions. According to global health reports, delayed detection significantly reduces survival rates and increases the complexity of treatment, often requiring aggressive interventions such as surgery, chemotherapy, and radiotherapy. Therefore, early and accu-rate detection of oral cancer is crucial for improving patient outcomes and reducing mortality rates.

Traditional diagnostic methods primarily rely on visual ex-amination by clinicians followed by biopsy for confirmation. While biopsy remains the gold standard, it is invasive, time-consuming, and requires specialized expertise. Moreover, manual screening is subject to inter-observer variability, making early-stage detection challenging in routine clinical practice. These limitations highlight the need for automated, reliable, and non-invasive diagnostic systems that can assist medical professionals in early screening.

In recent years, artificial intelligence (AI), particularly deep learning (DL), has shown remarkable success in medical im-age analysis. Convolutional Neural Networks (CNNs) have demonstrated the ability to automatically learn hierarchical feature representations from raw images, eliminating the need for manual feature engineering. Architectures such as VGGNet, ResNet, and EfficientNet have been widely adopted for classification tasks in medical imaging, including skin cancer detection, lung disease classification, and retinal disease diagnosis.

Despite these advancements, most existing systems for oral cancer detection focus only on image-level classification using raw input images. Such approaches often ignore critical factors such as lesion localization, image quality enhance-ment, and patient-specific clinical metadata. As a result, their performance may be limited when dealing with real-world clinical variability, such as lighting conditions, background noise, and variations in lesion appearance.

To overcome these limitations, this work proposes a com-prehensive deep learning framework for oral cancer detec-tion. The proposed system integrates image preprocessing, lesion segmentation, and deep learning-based classification to improve both accuracy and robustness. Preprocessing techniques such as contrast enhancement and noise reduction are applied to improve image quality and highlight lesion regions. Region of Interest (ROI)-based segmentation is used to isolate suspicious lesion areas, allowing deep learning models to focus on clinically relevant information for improved classification performance.

Furthermore, lesion segmentation is performed using Otsu thresholding and contour-based region extraction to isolate the region of interest (ROI), ensuring that the model fo-cuses on clinically significant areas rather than irrelevant background information. These segmented regions are then used to train multiple deep learning architectures, including a baseline CNN and EfficientNet-B0

In addition to image-based learning, the proposed framework combines preprocessing, segmentation, and deep feature extraction to improve diagnostic performance and model robustness under varying clinical imaging conditions. This enables the model to capture clinically relevant lesion characteristics, leading to improved diagnostic performance and prediction reliability.

To ensure transparency and interpretability, the proposed system employs Grad-CAM (Gradient-weighted Class Activation Mapping), which provides visual explanations by highlighting image regions that contribute most to the model's decision. This is particularly important in medical applications, where explainability is essential for gaining clinician trust and supporting decision-making.

The main contributions of this paper are summarized as follows:

- A complete end-to-end deep learning pipeline for oral cancer detection including preprocessing, segmentation, classification, and explainability.
- A comparative study of multiple CNN architecturessuch as Baseline CNN and EfficientNet-B0
- An EfficientNet-B0 based classification framework integrating preprocessing and ROI segmentation for improved classification performance.
- An interpretable AI framework using Grad-CAM to enhance clinical trust and visualization of predictions

The remainder of this paper is organized as follows: Section II presents related work in oral cancer and medical image analysis. Section III describes the proposed methodology indetail. Section IV discusses experimental setup and dataset description. Section V presents results and performance analysis, including ablation studies. Finally, Section VI concludes the paper and suggests future research directions.

II. RELATED WORK

A. Traditional Machine Learning Approaches

Early oral cancer detection systems relied on handcrafted feature extraction techniques combined with classical machine learning classifiers. Common feature descriptors include Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and color histogram-based features. These features are typically represented as:

$$F = \{f_1, f_2, f_3, \ldots, f_n\} \qquad (1)$$

Classification is performed using models such as Support Vector Machines (SVM), where the decision boundary is defined as:

$$w \cdot x + b = 0 \qquad (2)$$

However, these methods are limited by manual feature engineering and poor generalization ability.

B. Deep Learning-Based Approaches

Deep learning, particularly Convolutional Neural Networks (CNNs), has significantly improved medical image classification. CNNs automatically learn hierarchical feature representations using convolution operations:

$$y = f(W * x + b) \quad (3)$$

Architectures such as VGGNet, ResNet, and DenseNet have shown strong performance in medical imaging tasks, includ-ing cancer detection.

C. Transfer Learning Approaches

Due to limited medical datasets, transfer learning is widely used. Pretrained models like ResNet50 and EfficientNet are fine-tuned for medical classification tasks. The optimization objective is defined as:

$$\theta^* = \arg\min_{\theta} L(f_\theta(x), y) \qquad (4)$$

EfficientNet introduces a compound scaling strategy to balance depth, width, and resolution efficiently.

D. Segmentation-Based Approaches

Segmentation improves performance by isolating lesion regions. Otsu's thresholding selects an optimal threshold by minimizing intra-class variance:

$$\sigma^2(t) = \omega_1(t)\sigma_1^2 + \omega_2(t)\sigma_2^2 \qquad (5)$$

This helps focus the model on Region of Interest (ROI) and reduces background noise.

E. Explainable AI Methods

Explainable AI techniques such as Grad-CAM improve interpretability by highlighting important regions influencing predictions:

$$L^{c}_{Grad-CAM} = ReLU\left(\sum_{k} \alpha_k^c A^k\right) \quad (6)$$

This is critical for clinical acceptance of AI systems.

F. Research Gap and Motivation

Despite significant advancements, existing methods still face several limitations. Most approaches rely solely on image-based classification without proper integration of lesion segmentation, preprocessing, and explainable deep learning mechanisms.

From a real-world perspective, oral cancer detection requires early, accurate, and automated screening systems due to the shortage of specialists and delayed diagnosis in many regions. Existing computer-aided systems often fail under noisy conditions and do not generalize well to real clinical data.
Therefore, there is a strong need for a unified framework that integrates image preprocessing, lesion segmentation, deep learning-based classification, and explainable AI to improve diagnostic accuracy, robustness, and clinical trustworthiness.

III. PROPOSED METHODOLOGY

The proposed framework for oral cancer detection is a deep learning system that integrates image preprocessing, lesion segmentation, convolutional neural network-based classification, and explainable AI techniques. The overall workflow is designed to improve diagnostic accuracy while ensuring interpretability for clinical applications.

A. System Architecture Overview

The overall architecture of the proposed system is illustrated in Fig. 1. The pipeline consists of image acquisition, preprocessing, segmentation, feature extraction using CNN models and final classification.

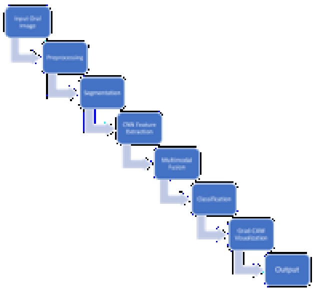

Fig. 1. Overall system architecture of the proposed oral cancer detection framework

B. Image Preprocessing

The input images undergo preprocessing to enhance qual-ity and remove noise. Techniques such as Gaussian filter-ing, contrast enhancement, and CLAHE (Contrast Limited Adaptive Histogram Equalization) are applied. These steps improve lesion visibility and ensure better feature extraction.

C. Lesion Segmentation

Lesion segmentation is an important step in the proposed framework to isolate the Region of Interest (ROI) from oral cavity images. This helps the model focus only on the cancerous region and removes unnecessary background information such as lips, teeth, and surrounding tissues.

In this work, segmentation is performed using Otsu's thresh-olding followed by contour detection. Otsu's method auto-matically selects an optimal threshold value by minimiz-ing intra-class variance, separating foreground (lesion) from background.

$$\sigma^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad (7)$$

After thresholding, contour detection is applied to extract the lesion boundary and generate the final ROI.

Lesion segmentation improves model performance by reduc-ing noise, enhancing feature quality, and increasing classifi-cation accuracy in CNN-based models.
The segmented output improves model performance by re-ducing background noise and highlighting clinically relevant regions.

D. Feature Extraction using CNN Models

The segmented images are passed through deep convolu-tional neural networks for feature extraction and classifica-tion. The study evaluates multiple architectures:

- Baseline CNN
- EfficientNet-B0

Fig. 2. Lesion segmentation using Otsu thresholding and contour extraction

The convolution operation is defined as:

$$y = f(W * x + b) \quad (8)$$

where W represents filters, x is input image, and f is activation function.

E. EfficientNet-B0 Architecture
EfficientNet-B0 is one of the primary deep learning architectures used in the proposed oral cancer detection framework. Compared to traditional convolutional neural networks, EfficientNet provides improved classification accuracy with lower computational complexity by using a compound scal-ing strategy that balances network depth, width, and image resolution.

The EfficientNet-B0 model is employed for automatic feature extraction from segmented oral lesion images. The architecture efficiently learns complex visual patterns such as lesion texture, irregular boundaries, abnormal tissue structures, and color variations associated with oral cancer

EfficientNet-B0 is selected due to its ability to balance computational efficiency and predictive performance through compound scaling. Unlike traditional CNN architectures that arbitrarily increase network dimensions, EfficientNet scales depth, width, and resolution systematically, resulting in better feature extraction with fewer parameters. This architecture effectively captures clinically relevant oral lesion characteristics such as irregular boundaries, tissue texture variations, and abnormal color patterns.

The segmented oral lesion image is provided as input to the EfficientNet-B0 architecture. Initially, convolution layers extract low-level image features such as edges, contours, and textures. As the network depth increases, high-level semantic features related to cancerous lesions are learned automatically.

Feature extraction is performed using convolutional neural networks, where hierarchical visual patterns are learned automatically from segmented oral lesion images. Initially, low-level features such as edges, contours, textures, and color distributions are extracted. As the network depth increases, more discriminative semantic features associated with abnormal lesion regions are learned, improving classification capability. The convolution mechanism follows Eq. (3), enabling automatic representation learning without manual feature engineering.

EfficientNet-B0 utilizes Mobile Inverted Bottleneck Convolution (MBConv) blocks along with squeeze-and-excitation optimization to improve feature learning efficiency while reducing the number of parameters.

After feature extraction, the deep feature vector generated by EfficientNet-B0 is passed through fully connected dense layers for classification of oral lesions into cancer and non-cancer categories.

The classification probability is obtained using the Softmax activation function:

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

EfficientNet-B0 provides several advantages in oral cancer detection:

- Improved classification accuracy
- Reduced computational complexity
- Better feature extraction capability
- Efficient parameter utilization
- Faster convergence during training
- Better generalization performance

Experimental analysis demonstrated that EfficientNet-B0 achieved the highest performance among the evaluated models, producing superior accuracy, precision, recall, F1-score, and ROC-AUC values. The model effectively captured clinically relevant lesion characteristics and showed strong robustness under varying imaging conditions.

Therefore, EfficientNet-B0 serves as the core classification architecture in the proposed oral cancer detection framework due to its balance between computational efficiency and predictive performance.

F. Classification
The extracted feature vector is passed through fully con-nected layers to classify the input into cancer or non-cancer categories. Softmax activation is used to generate probability scores:
The classification process follows the Softmax formulation defined earlier in Eq. (9).

G. Explainable AI using Grad-CAM
To enhance interpretability, Grad-CAM is used to highlight important regions in the image that influence model predictions.

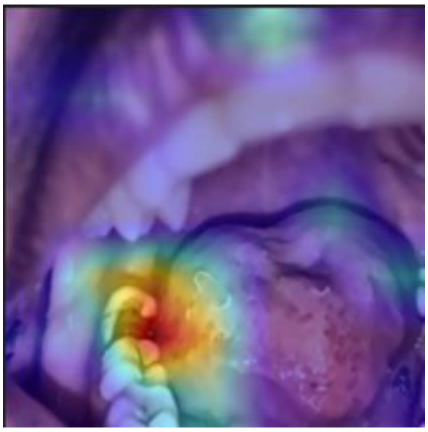

Fig. 3. Grad-CAM visualization highlighting cancerous regions

The overall workflow of the proposed system is summarized as follows:

- Input oral cavity image
- Preprocessing and enhancement
- Segmentation of lesion region
- Feature extraction using CNN
- Feature extraction and classification
- Classification using Softmax
- Grad-CAM visualization

IV. IMPLEMENTATION DETAILS

The proposed oral cancer detection system is implemented as a Python-based deep learning pipeline integrated with image preprocessing, segmentation, convolutional neural network training, and deep learning-based classification.

A. System Environment

The implementation is carried out using the following configuration:

- Programming Language: Python 3.x
- Deep Learning Framework: PyTorch
- Image Processing Libraries: OpenCV, PIL
- Hardware: GPU-enabled system (NVIDIA CUDA support)
- Development Platform: Jupyter Notebook / VS Code

The use of GPU acceleration significantly reduces training time for deep convolutional models such as EfficientNet-B0.

B. Dataset Preparation

The dataset used in this work consists of 1500 oral cavity images categorized into two classes: Cancer and Non-Cancer. The dataset was expanded from an initial collection of 750 images to 1500 images through additional data collection and augmentation strategies. This increase in dataset size significantly improves model generalization and reduces overfitting during training.

The dataset includes images captured under varying clinical conditions such as differences in illumination, lesion size, texture variation, and background noise. This diversity en-sures that the model learns robust and discriminative features applicable to real-world scenarios.

To maintain uniformity across all samples, images were resized to 224 × 224 pixels before being input to the deep learning models. The dataset was split into training, validation, and testing subsets in the ratio of 70%, 15%, and 15% respectively.

In addition, data augmentation techniques such as horizontal flipping, rotation, zooming, and scaling were applied during training to further enhance dataset variability and improve model robustness. The expanded dataset allows the proposed framework to achieve improved classification performance and better generalization compared to smaller datasets.

C. Image Preprocessing Implementation

The preprocessing pipeline includes noise reduction, contrast enhancement, and normalization. Gaussian filtering is used to remove noise, while CLAHE (Contrast Limited Adaptive Histogram Equalization) is applied to improve contrast in low-quality images.

$$X' = \frac{X - \mu}{\sigma}$$

Mathematically, normalization is defined as:
where X is the input pixel value, μ is mean, and σ is standard deviation.

D. Lesion Segmentation Implementation

Segmentation is implemented using Otsu thresholding followed by contour detection using OpenCV. Otsu's method automatically selects an optimal threshold to separate foreground (lesion) from background.

The thresholding operation is defined as:

$$\sigma^2(t) = \omega_1(t)\sigma_1^2 + \omega_2(t)\sigma_2^2 \tag{11}$$

After segmentation, the largest contour is extracted as the Region of Interest (ROI), which is used for further classification.

E. CNN Model Implementation

Two deep learning models are implemented and evaluated:

- Baseline CNN (custom architecture)
- EfficientNet-B0 (transfer learning)

The implemented CNN models automatically learn discriminative feature representations from segmented lesion images. During training, convolutional layers progressively learn spatial relationships and hierarchical lesion characteristics useful for differentiating cancerous and non-cancerous tissues. The convolution operation follows the formulation described in Eq. (3).

F. Training Strategy

The models are trained using the following settings:

- Loss Function: Cross Entropy Loss
- Optimizer: Adam Optimizer
- Learning Rate: 0.001
- Batch Size: 16 / 32
- Epochs: 25–50 (depending on convergence) The optimization objective is defined as:

$$\theta^* = \arg\min_{\theta} \mathcal{L}(f_\theta(x), y) \tag{12}$$

G. EfficientNet-B0 Implementation

The EfficientNet-B0 model is implemented as the primary deep learning architecture for oral cancer classification. The model is trained using segmented oral lesion images to automatically learn discriminative visual features associated with cancerous and non-cancerous tissues.

Initially, the input oral lesion images are resized and normalized before being passed to the EfficientNet-B0 network. Data augmentation techniques such as rotation, flipping, zooming, and contrast enhancement are applied to improve model generalization and reduce overfitting.

EfficientNet-B0 utilizes convolutional layers and Mobile Inverted Bottleneck Convolution (MBConv) blocks for hierarchical feature extraction.

EfficientNet-B0 improves feature representation by utilizing MBConv blocks and squeeze-and-excitation mechanisms, enabling efficient learning of lesion-specific visual information. The architecture enhances classification robustness by focusing on meaningful patterns extracted from segmented oral cavity images while maintaining computational efficiency.

The extracted deep feature maps are passed through globalaverage pooling and fully connected dense layers for clas-sification. Softmax activation is used to generate probabilityscores for cancer and non-cancer classes.
The classification layer converts extracted deep features into prediction probabilities for cancer and non-cancer classes. The probability distribution enables confidence estimation, allowing the model to indicate how strongly an input image belongs to a particular diagnostic category. The classification process follows the Softmax formulation defined earlier in Eq. (9).

The EfficientNet-B0 model is trained using the Adam optimizer with categorical cross-entropy loss. Dropout regularization is incorporated to prevent overfitting and improve model robustness.

Experimental results demonstrate that EfficientNet-B0 achieved superior classification performance compared to baseline CNN architectures due to its efficient feature scaling strategy and improved representation learning capability.

H. Classification Layer

The final output layer uses Softmax activation to classify input images into cancer or non-cancer categories.

I. Explainability Implementation (Grad-CAM)

Grad-CAM is implemented to visualize important regions influencing model predictions. It computes gradients of class scores with respect to feature maps and generates heatmaps highlighting lesion regions.

These heatmaps are overlaid on original images for interpretability.

J. System Workflow Summary

The complete implementation pipeline follows these steps:

- Image acquisition from dataset
- Preprocessing and enhancement
- Segmentation of lesion region
- Feature extraction using CNN models
- Deep feature extraction and classification
- Classification using Softmax
- Visualization using Grad-CAM

V RESULTS AND PERFORMANCE ANALYSIS

This section presents the experimental results of the pro-posed oral cancer detection framework. The performance of different deep learning models is evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-score.

A. Performance Comparison of Models

Table I summarizes the performance comparison of Baseline CNN and EfficientNet-B0

TABLE I

PERFORMANCE COMPARISON OF DIFFERENT MODELS

Model	Accuracy	Precision	Recall	F1-score
Baseline CNN	0.55	0.82	0.44	0.57
EfficientNet-B0	0.95	0.96	0.97	0.97

B. Experimental Observation

Experimental analysis demonstrates that ROI-based lesion segmentation contributes significantly to performance improvement by reducing irrelevant background information and directing the model toward clinically meaningful lesion regions. The proposed EfficientNet-B0 model consistently achieved improved feature representation and stronger classification capability compared to the baseline CNN model.
Additionally, preprocessing techniques such as contrast enhancement and normalization improved visual quality, allowing the model to learn discriminative lesion character-istics more effectively. The combination of preprocessing, segmentation, deep feature extraction, and explainable AI

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (13)$$

Explainability Implementation (Grad-CAM)

Grad-CAM is implemented to visualize important regions influencing model predictions. It computes gradients of class scores with respect to feature maps and generates heatmaps highlighting lesion regions.

$$L^c_{Grad-CAM} = ReLU\left(\sum_k \alpha^c_k A^k\right)$$

These heatmaps are overlaid on original images for interpretability.

J. System Workflow Summary

The complete implementation pipeline follows these steps:

- Image acquisition from dataset
- Preprocessing and enhancement
- Segmentation of lesion region
- Feature extraction using CNN models
- Deep feature extraction and classification
- Classification using Softmax
- Visualization using Grad-CAM

V. RESULTS AND PERFORMANCE ANALYSIS

This section presents the experimental results of the pro-posed oral cancer detection framework. The performance of different deep learning models is evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-score.

A. Performance Comparison of Models

Table I summarizes the performance comparison of Baseline CNN and EfficientNet-B0

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT MODELS

Model	Accuracy	Precision	Recall	F1-score
Baseline CNN	0.55	0.82	0.44	0.57
EfficientNet-B0	0.95	0.96	0.97	0.97

models significantly outperform the baseline CNN. The EfficientNet model achieves the highest overall performance due to efficient deep feature extraction and ROI-focused learning

B. Experimental Observation

Experimental analysis demonstrates that ROI-based lesion segmentation contributes significantly to performance improvement by reducing irrelevant background information and directing the model toward clinically meaningful lesion regions. The proposed EfficientNet-B0 model consistently achieved improved feature representation and stronger classification capability compared to the baseline CNN model.
Additionally, preprocessing techniques such as contrast enhancement and normalization improved visual quality, allowing the model to learn discriminative lesion character-istics more effectively. The combination of preprocessing, segmentation, deep feature extraction, and explainable AI reliability. The results clearly show that deep learning-based

models significantly outperform the baseline CNN. The EfficientNet model achieves the highest overall performance due to efficient deep feature extraction and ROI-focused learning.

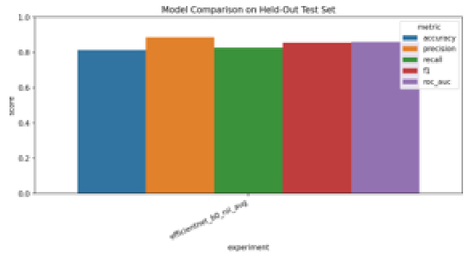

Fig. 4. Performance comparison of model

C. Confusion Matrix Analysis

The confusion matrix is used to evaluate the classification performance in terms of correctly and incorrectly classified samples. The proposed model shows a higher number of true positives and true negatives compared to baseline models, indicating improved classification reliability.

- True Positives (TP): Correctly classified cancer cases
- True Negatives (TN): Correctly classified non-cancer cases
- False Positives (FP): Non-cancer classified as cancer
- False Negatives (FN): Cancer classified as non-cancer

D. Comparative Model Analysis

To evaluate the contribution of each component in the proposed framework, an ablation study is conducted as shown in Table II.

TABLE II
COMPARATIVE MODEL ANALYSIS

Model Configuration	Accuracy
Baseline CNN + ROI Segmentation	0.5535
EfficientNet-B0 + ROI Segmentation	0.9545

The ablation results demonstrate that each component, including segmentation and EfficientNet-based feature extraction, contributes positively to overall performance improvement.

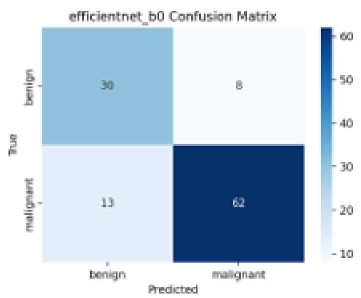

Fig. 5. Confusion matrix of EfficientNet-B0 model

E. ROC-AUC Analysis

The Receiver Operating Characteristic (ROC) curve is used to evaluate the trade-off between sensitivity and specificity. The area under the curve (AUC) for the proposed model is higher compared to baseline methods, indicating better classification capability across different thresholds.

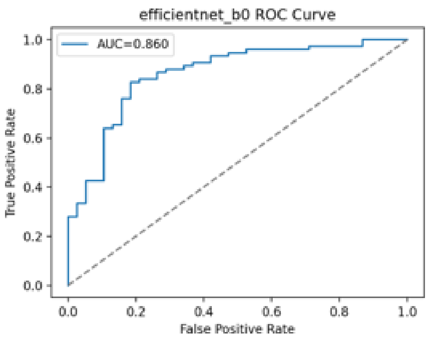

Fig. 6. ROC Curve

The experimental results confirm that the proposed deep learning framework significantly improves oral cancer detection performance. EfficientNet-B0 outperforms traditional CNN models due to their ability to extract more discrimina-tive features.

Furthermore, the integration of lesion segmentation reduces background noise, improving feature quality. The inclusion of lesion segmentation and preprocessing further enhances classification accuracy by providing additional contextual information.

Finally, the use of Grad-CAM improves model interpretabil-ity by highlighting regions responsible for predictions, mak-ing the system more suitable for clinical applications.

G. Model Accuracy Analysis

The training and validation accuracy curves of the proposed model are shown in Fig. 7. The graph indicates steady convergence, demonstrating effective learning without severe overfitting.

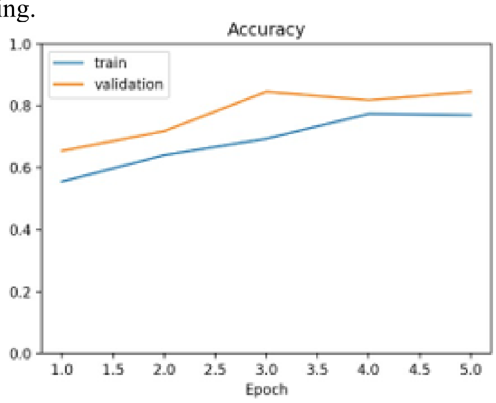

Fig. 7. Training and validation accuracy of the proposed model

H. Model Loss Analysis

The training and validation loss curves are shown in Fig. 8. The gradual decrease in loss indicates stable optimization and proper convergence of the model.

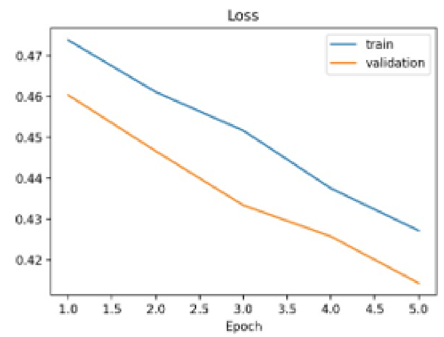

Fig. 8. Training and validation loss of the proposed model

VI. SYSTEM OUTPUTS AND VISUALIZATION

This section presents the output results generated by the proposed oral cancer detection framework. The developed system was evaluated in terms of user interaction, backend processing, prediction capability, interpretability, and visual-ization support.

A. User Interface Output

The graphical user interface (GUI) of the proposed system allows users to upload oral cavity images and perform auto-mated cancer detection. The interface displays the uploaded image, prediction results, confidence score, and generated visualizations.

Fig. 9 illustrates the user interface of the developed applica-tion.

Fig. 9. User interface of the proposed oral cancer detection system

The interface is designed to be simple and user-friendly, enabling efficient interaction for both medical professionals and non-technical users.

B. Prediction Output

The prediction module classifies the input image into cancer or non-cancer categories. The system also provides proba-bility scores indicating prediction confidence.

Fig. 10 shows the prediction output generated by the pro-posed framework.

Fig. 10. Prediction output showing detected class and confidence score

The system successfully identifies suspicious oral lesions with high classification confidence, demonstrating effective detection capability.

C. Backend Processing Output

The backend processing stage includes preprocessing, lesion segmentation, feature extraction, and classification opera-tions. Intermediate outputs generated during processing are shown in Fig. 11.

The preprocessing stage improves image quality, while seg-mentation isolates the lesion region for more accurate feature extraction.

D. System Behaviour Analysis

The proposed system demonstrates stable behaviour during training and inference phases. The system effectively handles

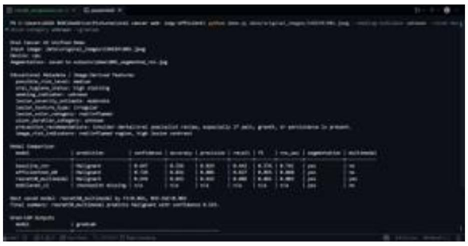

Fig. 11. Backend processing stages including preprocessing and segmen-tation

variations in lesion size, texture, and illumination conditions. The model maintains stable convergence and produces reli-able outputs even for challenging oral lesion images.

E. Interpretability Analysis using Grad-CAM

To improve transparency and trustworthiness, Grad-CAM visualization is used to highlight image regions influencing model predictions.

Fig. 12 shows the Grad-CAM heatmap generated by the proposed framework.

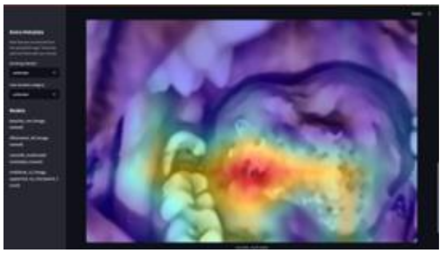

Fig. 12. Grad-CAM visualization highlighting important lesion regions

The highlighted regions correspond to suspicious lesion areas, improving interpretability and assisting clinicians in understanding model decisions.

F. Clinical Support Visualization

The proposed framework provides visual assistance for clini-cal analysis by combining lesion segmentation, classification results, and explainability outputs.

The generated outputs can assist medical practitioners in identifying suspicious lesions and making informed diag-nostic decisions. The integration of explainable AI improves confidence and usability in real-world healthcare environ-ments.

G. Overall Analysis

The proposed oral cancer detection system demonstrates strong performance in terms of prediction accuracy, inter-pretability, and usability. The integration of preprocessing, segmentation, deep learning, and explainable AI contributes significantly to reliable oral cancer detection.

Experimental observations confirm that the framework effec-tively identifies lesion regions, provides stable predictions, and generates clinically meaningful visual explanations suit-able for decision support systems.

VII. CONCLUSION

This paper presented an intelligent deep learning framework for automated oral cancer detection using lesion segmenta-tion, convolutional neural networks, and explainable artifi-cial intelligence techniques. The proposed system integrates image preprocessing, ROI-based segmentation, deep feature extraction and Grad-CAM visualization to improve both classification performance and interpretability.

Experimental results demonstrated that advanced deep learn-ing architectures such as EfficientNet-B0 significantly out-perform traditional CNN models in detecting oral cancer from oral cavity images. The incorporation of lesion segmen-tation improved feature quality by isolating clinically rele-vant regions, while EfficientNet-based deep feature learning enhanced overall prediction accuracy and robustness.

The Grad-CAM-based explainability module provided visual interpretation of prediction results by highlighting suspicious lesion regions responsible for classification decisions. This improves transparency and supports medical professionals in understanding the behaviour of the AI system.

The proposed framework achieved strong performance across multiple evaluation metrics including accuracy, precision, recall, and F1-score. The experimental analysis confirmed that the integration of preprocessing, segmentation, deep learning, and explainable AI contributes significantly to reliable and interpretable oral cancer diagnosis.

Overall, the developed system demonstrates the potential of artificial intelligence in assisting early oral cancer screening and clinical decision support. The proposed framework can contribute toward faster diagnosis, reduced manual effort, and improved accessibility to healthcare support systems, particularly in resource-limited environments.

VIII. FUTURE WORK

A. Dataset Expansion

Future work will focus on expanding the dataset using larger and more diverse oral lesion image collections obtained from multiple healthcare institutions. A larger dataset can improve model generalization and reduce bias across different patient groups and imaging conditions.

B. Advanced Deep Learning Architectures

Advanced deep learning architectures such as Vision Transformers (ViTs), hybrid CNN-transformer models, and self-supervised learning approaches can be investigated to im-prove feature extraction and classification performance. These architectures may provide better contextual under-standing of lesion patterns compared to conventional CNN models.

C. Enhanced Clinical Feature Integration

Future research may include integration of additional clinical information such as patient age, tobacco usage history, alcohol consumption, lesion duration, and biopsy reports to further improve diagnostic reliability and clinical decision support capability.

D. Real-Time Deployment

Real-time deployment of the proposed framework through mobile applications, cloud-based healthcare platforms, or edge AI systems can improve accessibility for remote screen-ing and telemedicine applications. Integration with Internet of Medical Things (IoMT) devices may further support smart healthcare environments.

E. Advanced Explainable AI

Another important direction involves improving explainabil-ity and transparency using advanced Explainable AI (XAI) techniques beyond Grad-CAM, enabling clinicians to better understand model decisions and increase trust in AI-assisted diagnosis.

F. Clinical Validation

Extensive clinical validation with medical professionals and hospital datasets will be necessary before deploying the system in real-world healthcare environments. Such validation can help ensure reliability, safety, and practical usability for clinical decision support systems.

REFERENCES

1. K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” in Proc. International Conference on Learning Representations (ICLR), 2015.
2. M. Tan and Q. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in Proc. International Conference on Machine Learning (ICML), pp. 6105–6114, 2019.
3. K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
4. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” in Proc. IEEE International Conference on Computer Vision (ICCV), pp. 618–626, 2017.
5. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241, 2015.
6. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classifi-cation with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 25, pp. 1097–1105, 2012.
7. G. Litjens et al., “A Survey on Deep Learning in Medical Image Analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2017.
8. S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.
9. N. Otsu, “A Threshold Selection Method from Gray-Level His-tograms,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62–66, 1979.
10. D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimiza-tion,” in Proc. International Conference on Learning Representations (ICLR), 2015.

11. C. Szegedy et al., “Going Deeper with Convolutions,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, 2015.
12. A. Vaswani et al., “Attention Is All You Need,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017.
13. J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, “Ima-geNet: A Large-Scale Hierarchical Image Database,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255, 2009.
14. F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1251–1258, 2017.
15. J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018.
16. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2117–2125, 2017.
17. K. Zhang, M. Liu, J. Wang, and T. Liao, “Attention U-Net: Learning Where to Look for the Pancreas,” arXiv preprint arXiv:1804.03999, 2018.
18. Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” in Deep Learning in Medical Image Analysis (DLMIA), pp. 3–11, 2018.
19. E. J. Bernardes et al., “Deep Learning in Oral Cancer Detection: A Review,” Biomedical Signal Processing and Control, vol. 62, Art. no. 102054, 2020.
20. S. M. Khan et al., “Artificial Intelligence in Oral Cancer Diagnosis: A Review and Future Directions,” Computers in Biology and Medicine, vol. 135, Art. no. 104576, 2021.