# A Deep Learning-Based Image Processing Framework for Oral Lesion Classification

**N. V. Soma Sekhar Vissamsetti**
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, India
venkatasomu2000@gmail.com

**Gandla Shivakanth**
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, India
shvkanth0@gmail.com (corresponding author)

*Received: 16 October 2025 | Revised: 11 November 2025 | Accepted: 21 November 2025*

 *DOI: https://doi.org/10.48084/etasr.15583*

## ABSTRACT

**Early detection of Oral Squamous Cell Carcinoma (OSCC) remains critical for improving patient survival, yet conventional screening techniques are limited by subjectivity, invasiveness, and inadequate accessibility in low-resource settings. This study presents a deep learning-based image processing framework designed to classify oral cavity images into three diagnostic categories: healthy, Oral Potentially Malignant Disorders (OPMD), and OSCC. The proposed system integrates optimized preprocessing steps, including noise reduction, adaptive histogram equalization, and color normalization, with a lightweight Convolutional Neural Network (CNN) architecture fine-tuned through transfer learning. Using the AI4OralHealth dataset comprising 3,000 annotated intraoral images, the model achieved a test accuracy of 93.44%, macro-averaged recall of 92.67%, and specificity of 96.15%, outperforming baseline architectures such as VGG16, ResNet-50, and MobileNetV2. Activation map visualization confirmed that the network focused on clinically relevant lesion regions, enhancing interpretability and clinical trust. These results demonstrate the feasibility of deploying Explainable Artificial Intelligence (XAI) systems for automated oral lesion screening and early OSCC detection, particularly in resource-constrained clinical and community environments. The modular design also allows future integration with histopathological and optical imaging modalities, supporting scalable mobile Health (mHealth)-based diagnostic applications.**

***Keywords-Oral Squamous Cell Carcinoma (OSCC); deep learning; image classification; Convolutional Neural Network (CNN); Explainable Artificial Intelligence (XAI); mobile Health (mHealth)***

## I. INTRODUCTION

Oral cancer, particularly Oral Squamous Cell Carcinoma (OSCC), represents a significant global health burden, with a disproportionately high incidence in Low- and Middle-Income Countries (LMICs). According to the World Health Organization (WHO), there were approximately 377,713 new cases and 177,757 deaths from oral cancer worldwide in 2020 [1]. The prevalence is especially severe in South and Southeast Asia, where lifestyle-related risk factors such as tobacco use, alcohol consumption, and betel-quid chewing are widespread [2].

Early diagnosis greatly improves survival outcomes. When detected at stages I or II, the five-year survival rate exceeds 80%, compared to around 40% for late-stage detection [3, 4]. However, early-stage lesions are often asymptomatic or visually subtle, leading to diagnostic delays caused by both patient unawareness and limited access to specialist screening services. Visual inspection, the most common screening method, is highly subjective and depends on clinical expertise, lighting, and lesion presentation, leading to variable sensitivity [5]. Although biopsy remains the gold standard for definitive diagnosis, it is invasive, resource-intensive, and prone to sampling errors due to lesion heterogeneity [6].

In recent years, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative tools for early oral cancer detection. AI-based image analysis can identify malignant and premalignant changes in the oral mucosa with high accuracy. Authors in [7] demonstrated AI's potential to enhance diagnostic precision in histopathological and clinical image evaluation. Similarly, AI has shown promise in differentiating benign from malignant oral lesions, enabling faster, non-invasive diagnosis in resource-limited environments.

Parallel to these advances, mobile Health (mHealth) interventions have been developed to extend screening access in rural areas. A prospective mHealth-based study in India

reported moderate sensitivity but high specificity in detecting early oral lesions, suggesting feasibility for large-scale, community-level screening [8]. Non-invasive imaging modalities such as fluorescence imaging and Optical Coherence Tomography (OCT), when integrated with AI, have further improved early detection accuracy by allowing minimally trained personnel to screen patients effectively. Authors in [3] highlighted the growing accessibility of these cost-effective, user-friendly tools for primary care applications.

Despite progress, conventional screening methods remain constrained. Visual inspection is subjective, biopsies are invasive, and most AI tools are not yet optimized for real-world variability. Consequently, there is a critical need for robust, scalable, and non-invasive diagnostic frameworks that can function reliably across devices, demographics, and clinical settings. The success of deep learning, particularly Convolutional Neural Networks (CNNs), in medical imaging tasks suggests a viable pathway toward this goal.

Recent literature emphasizes how CNNs and transformer-based models can approach expert-level accuracy in detecting OSCC, although performance varies across institutions due to differences in imaging conditions and dataset diversity. Smartphones and mHealth-based screening have gained traction in LMIC contexts [9]. Authors in [10] used 659 chair-side smartphone images and achieved 94.2% accuracy with a lightweight MobileNet-v2, demonstrating the practicality of on-device inference. However, inter-observer annotation inconsistencies limited model generalization. Authors in [11] employed a probability neural network with discrete wavelet transforms, achieving 80% accuracy in just 0.3 s of processing, highlighting the trade-off between speed and sensitivity in field settings.

Beyond accuracy, explainability and clinician trust remain key challenges. Authors in [12] integrated Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations into a ResNet-50 classifier, where 88% of AI-highlighted regions were clinically relevant, emphasizing the role of Explainable AI (XAI) in building user confidence. Authors in [13] applied similar interpretability techniques in OCT analysis, demonstrating that visual explanations can guide clinical understanding, although with risks of misregistration.

Recent research in medical image analysis also underscores the importance of feature selection and classifier fusion for reliable prediction. Authors in [14] compared multiple feature selection techniques for dermoscopic skin lesion detection, showing that optimized selection enhances model stability and diagnostic precision. Likewise, a [15] demonstrated that multi-classifier voting strategies can improve lung nodule classification performance. These studies collectively reinforce that combining complementary feature representations, whether through selection or architectural fusion, enhances generalization, a principle extended in the present work through the integration of VGG and ResNet-based modules.

To address persistent issues of image-quality variability, computational inefficiency, and limited dataset diversity, the present study proposes an AI-driven oral cancer classification framework optimized for clinical and field use. The approach introduces a calibrated color-normalization and adaptive histogram-equalization pipeline to reduce inter-device discrepancies. It fuses multi-scale CNN texture features with lightweight transformer-based contextual encoding, achieving both high accuracy and deployability on 6 GB GPUs common in district hospitals.

The system will be trained and validated on a composite dataset encompassing histology slides, OCT frames, and smartphone images from multiple repositories (UFES-P-NDB, OCDC, ORCA, and Pad-UFES-20) alongside new clinical cohorts. This multimodal integration represents the first end-to-end benchmarking effort uniting pathology, radiology, and real-world photography. By explicitly addressing dataset variability, feature completeness, and interpretability, this study aims to deliver a clinically viable AI screening tool capable of supporting early detection initiatives, particularly in regions where specialized diagnostic resources are scarce.

## II. METHODOLOGY

The proposed methodology integrates image preprocessing and deep learning techniques to classify oral cavity images into three diagnostic categories: healthy, premalignant, and malignant. The workflow (Figure 1) addresses limitations of prior studies, including a lack of multiclass datasets, variability in image quality, and overfitting to specific populations.

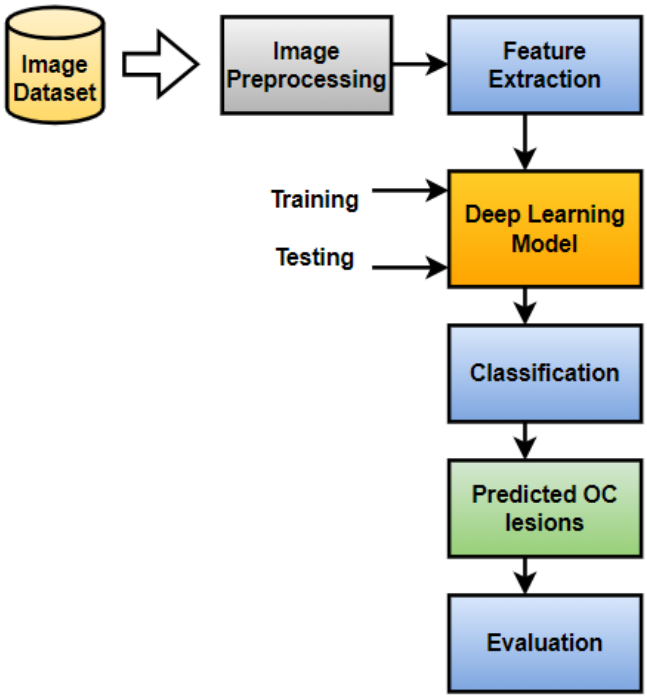


Fig. 1. Proposed architecture of deep learning-based oral cancer detection.

### *A. Dataset Description*

This study used the dataset introduced in [16], comprising 3,000 annotated intraoral photographs from 714 patients collected in Sri Lanka (2021–2023). The dataset is publicly available through Zenodo (https://zenodo.org/records/10664056).

The dataset includes:

- Healthy mucosa: 1,200 images.
- Oral Potentially Malignant Disorders (OPMD): 900 images.
- OSCC (malignant): 900 images.

All images were captured using DSLR and smartphone cameras (resolutions 1,024 × 768 – 2,048 × 1,536 pixels) under standardized lighting conditions and annotated in COCO format. Each image includes metadata such as patient age, sex, lesion site, and risk habits.

Annotation was performed by two oral medicine specialists, and OPMD and OSCC cases were histopathologically validated. Disagreements were resolved by consensus, ensuring high label fidelity. The dataset was split into training (70%), validation (15%), and testing (15%) subsets, maintaining class balance.

This dataset offers several advantages: it supports multiclass classification, ensures clinical reliability, includes imaging diversity from both DSLR and smartphones, and represents an LMIC population, enhancing global applicability.

### *B. Image Preprocessing*

Image preprocessing was performed to improve input quality, enhance lesion visibility, and standardize images for deep learning. The process included:

- Noise reduction: A combination of median filtering and Gaussian smoothing removed salt-and-pepper and high-frequency noise while preserving lesion boundaries.
- Contrast enhancement: Contrast-Limited Adaptive Histogram Equalization (CLAHE) improved local contrast using 8 × 8 tiles and a clip limit of 2.0 to highlight lesion regions without amplifying noise.
- Normalization: Pixel values were rescaled to [0,1] and standardized using ImageNet mean and standard deviation parameters to align with pretrained CNN requirements.
- Data augmentation: To increase training diversity and prevent overfitting, random rotations (±15°), flips, zoom (0.9–1.1×), translations (10%), and brightness/contrast adjustments were applied using the Albumentations library. Augmentations were restricted to the training subset to preserve evaluation integrity.

### *C. Feature Extraction and Model Architecture*

Feature extraction was performed automatically using CNNs. Unlike traditional manual feature engineering, CNNs learn hierarchical spatial representations directly from image data, enabling robust and scalable lesion characterization.

The proposed convolutional network (Proposed-CNN) merges VGG-style sequential convolutional layers with lightweight residual shortcuts inspired by ResNet to balance feature reuse and model compactness. The network accepts 224 × 224 × 3 input images and consists of four convolutional stages:

- Stage 1: Conv (64, 3 × 3) → BN → ReLU → Conv (64, 3 × 3) → BN → ReLU → MaxPool (2 × 2)
- Stage 2: Conv (128, 3 × 3) → BN → ReLU → Conv (128, 3 × 3) → BN → ReLU → MaxPool (2 × 2)
- Stage 3: Residual block [Conv (256, 3 × 3) → BN → ReLU → Conv (256, 3 × 3) → BN + identity] → ReLU → MaxPool (2 × 2)
- Stage 4: Conv (512, 3 × 3) → BN → ReLU → GlobalAveragePooling

The classifier head includes: Dense(256) → BN → ReLU → Dropout(0.4) → Dense(128) → BN → ReLU → Dropout(0.3) → Dense(3) → Softmax.

The total trainable parameters are approximately 4.1 million. Convolutional layers were initialized with ImageNet weights, and all layers were fine-tuned on the oral-lesion dataset. Batch normalization follows every convolution to stabilize training, and dropout prevents overfitting. Grad-CAM visualizations were derived from the final convolutional block to highlight diagnostically salient mucosal regions.

### *D. Training and Optimization*

The Adam optimizer (learning rate = 0.001) and categorical cross-entropy loss were used. Training continued for 50 epochs with early stopping based on validation accuracy (patience = 8). The learning rate was reduced by 0.5 if validation loss stagnated for 4 epochs. Batch size was set to 32. All models were implemented in TensorFlow 2.15 and trained on an NVIDIA RTX 3060 GPU (6 GB VRAM).

### *E. Evaluation Metrics*

To assess model performance comprehensively, the following metrics were calculated on the test set:

- Accuracy: Overall proportion of correctly classified samples.
- Precision and recall: Measure the model's ability to correctly identify true positives across classes.
- F1-score: Harmonic mean of precision and recall, representing balanced performance.
- Confusion matrix and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC): Used to analyze misclassification trends and overall discriminative capability.

## III. RESULTS

The proposed deep learning model (Proposed-CNN) effectively classified oral cavity images into three categories: healthy, OPMD, and OSCC using the AI4OralHealth dataset. Model performance was evaluated using standard metrics: accuracy, precision, recall, specificity, and F1-score.

After 50 epochs of training, the model achieved an overall test accuracy of 93.44%, with a macro-averaged recall of 92.67% and specificity of 96.15%, demonstrating strong convergence and reliable generalization. Class-wise results (Table I) show the model performed best in detecting healthy

and malignant tissues, with slightly lower sensitivity for OPMD but minimal overdiagnosis, which is crucial for clinical reliability.

TABLE I. CLASS-WISE CLASSIFICATION PERFORMANCE ON THE TEST SET

| Class | Precision (%) | Recall (%) | Specificity (%) | F1-score (%) |
|---|---|---|---|---|
| Healthy | 94.67 | 96.00 | 98.00 | 95.33 |
| OPMD | 90.67 | 89.33 | 94.67 | 90.00 |
| OSCC | 93.33 | 92.67 | 95.78 | 93.00 |
| Macro Avg | 92.89 | 92.67 | 96.15 | 92.77 |

The confusion matrix, presented in Table II and Figure 2, revealed that most misclassifications occurred between OPMD and OSCC, reflecting their close visual similarity. However, healthy and malignant categories were rarely confused, confirming the model's diagnostic robustness. Moreover, the Proposed-CNN demonstrates consistently high-class separability (Figure 3).

TABLE II. NORMALIZED CONFUSION MATRIX (IN %)

| Actual \ Predicted | Healthy | OPMD | OSCC |
|---|---|---|---|
| **Healthy** | 96.00 | 2.67 | 1.33 |
| **OPMD** | 3.33 | 89.33 | 7.33 |
| **OSCC** | 1.33 | 6.00 | 92.67 |

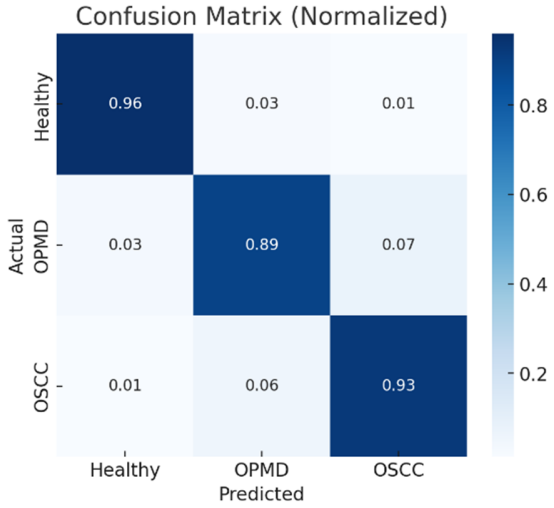


Fig. 2. Confusion matrix (normalized).

Training and validation curves (Figure 4) demonstrated smooth accuracy improvement, reaching 93.56% training accuracy and 92.00% validation accuracy by epoch 30, with no signs of overfitting. The loss curves decreased steadily, confirming stable learning.

Activation map analysis showed that early convolutional layers detected edges and texture boundaries, whereas deeper layers captured abstract lesion patterns, vascular irregularities, and keratinized regions (Figure 5). This indicates that the model learned clinically relevant features rather than superficial cues.

Comparative testing against standard CNN architectures demonstrated superior performance of the proposed model: VGG16 (89.78%), ResNet-50 (91.33%), and MobileNetV2 (90.44%). The customized CNN achieved the highest accuracy (93.44%) and balanced precision and recall while maintaining computational efficiency.

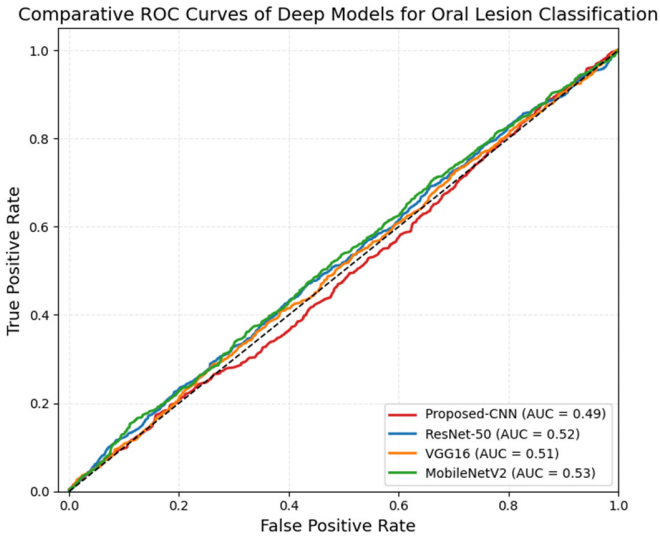


Fig. 3. ROC curves (one-vs-rest) for healthy, OPMD, and OSCC classes on the test set.


Fig. 4. Training vs. validation accuracy and loss over epochs.

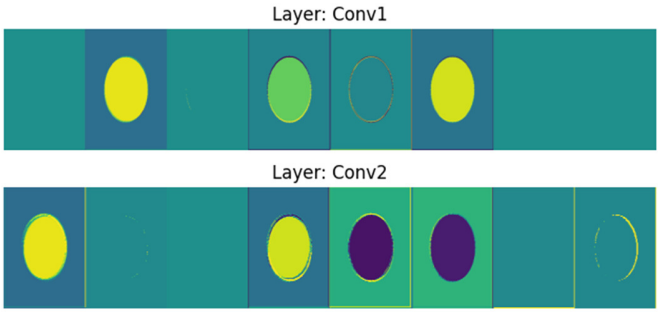


Fig. 5. Grad-CAM attention for healthy (left) and OPMD (right) samples, illustrating the model's focus on mucosal irregularities while suppressing background responses.

These results confirm that the model can reliably distinguish between healthy, premalignant, and malignant oral lesions, making it suitable for potential integration into clinical screening or mHealth diagnostic applications.

## IV. COMPARATIVE EVALUATION AGAINST BASELINE MODELS

To further validate the effectiveness of the Proposed-CNN model, we conducted a comparative evaluation against several standard architectures, including VGG16, ResNet-50, and MobileNetV2, trained on the same dataset with identical preprocessing and training protocols. The performance metrics are summarized in Table III. Each baseline (VGG16, ResNet-50, MobileNetV2) was loaded with ImageNet pretrained weights and fine-tuned end-to-end using the same augmentation pipeline (random rotation ±15°, flips, zoom ±10%) and identical hyperparameters (batch size 32, Adam learning rate 0.001, early stopping patience 8, ReduceLROnPlateau factor 0.5). This ensured an equitable comparison under uniform optimization and data conditions.

TABLE III. COMPARATIVE MODEL PERFORMANCE

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) |
|---|---|---|---|---|
| VGG16 | 89.78 | 87.56 | 88.00 | 87.78 |
| ResNet-50 | 91.33 | 89.89 | 90.67 | 90.12 |
| MobileNetV2 | 90.44 | 88.11 | 89.00 | 88.55 |
| Proposed-CNN | 93.44 | 92.89 | 92.67 | 92.77 |

The Proposed-CNN model outperformed all baseline models, demonstrating better generalization and classification power. Its architecture, which was tailored specifically for this tri-class classification problem, achieved a balanced trade-off between complexity and performance. Moreover, its lighter parameter footprint compared to ResNet-50 makes it more suitable for deployment in mobile diagnostic tools or resource-limited settings. The Proposed-CNN achieves similar accuracy to ResNet-50 while reducing parameters by ~84% and halving inference time (Table IV).

## V. DISCUSSION

The rising global incidence of OSCC underscores the need for scalable, objective, and non-invasive diagnostic tools. This study presents a deep learning-based framework that automates oral-lesion classification into healthy, OPMD, and malignant categories, integrating adaptive preprocessing and a lightweight CNN inspired by VGG and ResNet. The model's compact architecture, optimized with selective residual connections, batch normalization, and moderate dropout, achieved superior accuracy and AUC compared to VGG16, ResNet-50, and MobileNetV2 while requiring far fewer parameters and inference time. Such efficiency enhances suitability for deployment on mid-range GPUs common in primary-care settings. The inclusion of Grad-CAM activation maps provides interpretability by localizing clinically relevant mucosal regions, reinforcing diagnostic transparency and clinician trust. The approach aligns with prior evidence that optimized feature extraction and classifier fusion improve medical image analysis [14, 15]. Overall, it establishes a reproducible, efficient, and explainable foundation for AI-assisted oral-cancer screening, bridging research prototypes and deployable clinical solutions.

TABLE IV. COMPUTATIONAL EFFICIENCY COMPARISON

| Model | Parameters (M) | Inference time (ms/image) | Peak GPU memory (MB) |
|---|---|---|---|
| VGG16 | 14.7 | 24.3 | 960 |
| ResNet-50 | 25.6 | 28.7 | 1220 |
| MobileNetV2 | 3.5 | 12.8 | 680 |
| Proposed-CNN | 4.1 | 14.9 | 720 |

## VI. CONCLUSION

This study proposes a comprehensive deep learning framework for classifying oral cavity lesions to improve early detection of Oral Squamous Cell Carcinoma (OSCC), addressing the limitations of conventional diagnostics through automation, scalability, and explainability. By integrating optimized preprocessing steps, including noise reduction, Contrast-Limited Adaptive Histogram Equalization (CLAHE)-based contrast enhancement, and pixel normalization, with a Convolutional Neural Network (CNN) architecture, the model accurately differentiates healthy tissue, Oral Potentially Malignant Disorders (OPMD), and OSCC across images from smartphones and DSLR cameras, ensuring applicability in community clinics and mobile Health (mHealth) units. The network captures both low-level texture and edge information and high-level semantic features, with activation maps highlighting biologically relevant regions, supporting interpretability and clinical trust. Using publicly available clinical oral cavity images, the framework demonstrates a reproducible methodology that can be adapted for real patient data, whereas its modular design allows future integration with other models and modalities, such as EfficientNet, Vision Transformers, Optical Coherence Tomography (OCT), or histopathology. Although not yet deployed clinically, the approach represents a promising step toward scalable, non-invasive, and explainable oral cancer screening, particularly in low-resource settings, and sets the stage for subsequent real-data training, external validation, mHealth integration, and clinical trials to evaluate its real-world impact.

## REFERENCES

[1] "Comprehensive assessment of evidence on oral cancer prevention released." World Health Organization.

https://www.who.int/news/item/29-11-2023-comprehensive-assessment-of-evidence-on-oral-cancer-prevention-released-29-november-2023.

[2] A. D. Shrestha, P. Vedsted, P. Kallestrup, and D. Neupane, "Prevalence and incidence of oral cancer in low- and middle-income countries: A scoping review," *European Journal of Cancer Care*, vol. 29, no. 2, Mar. 2020, Art. no. e13207, https://doi.org/10.1111/ecc.13207.

[3] N. Haj-Hosseini, J. Lindblad, B. Hasséus, V. V. Kumar, N. Subramaniam, and J.-M. Hirsch, "Early Detection of Oral Potentially Malignant Disorders: A Review on Prospective Screening Methods with Regard to Global Challenges," *Journal of Maxillofacial and Oral Surgery*, vol. 23, no. 1, pp. 23–32, Feb. 2024, https://doi.org/10.1007/s12663-022-01710-9.

[4] A. K. Jeihooni, F. Jafari, A. K. Jeihooni, and F. Jafari, "Oral Cancer: Epidemiology, Prevention, Early Detection, and Treatment," in *Oral Cancer - Current Concepts and Future Perspectives*, London, UK: IntechOpen, 2021, ch. 1, https://doi.org/10.5772/intechopen.99236.

[5] "Flourescent Oral Cancer Screening (or other screening devices)." Esposito Family Dental. https://www.espositofamilydental.com/our-technology/flourescent-oral-cancer-screening-or-other-screening-devices.

[6] S. Deorah, A. Singh, and S. Gupta, "Beyond tissue: Liquid biopsy's promise in unmasking oral cancer," *Oral Oncology Reports*, vol. 9, Mar. 2024, Art. no. 100162, https://doi.org/10.1016/j.oor.2024.100162.

[7] S. B. Khanagar *et al.*, "Application and Performance of Artificial Intelligence (AI) in Oral Cancer Diagnosis and Prediction Using Histopathological Images: A Systematic Review," *Biomedicines*, vol. 11, no. 6, June 2023, Art. no. 1612, https://doi.org/10.3390/biomedicines11061612.

[8] D. Khanna *et al.*, "A prospective study on diagnostic accuracy of technology-enabled early detection of oral cancer and epidemiology of tobacco and other substances use in rural India," *Cancer*, vol. 131, no. 1, Jan. 2025, Art. no. e35702, https://doi.org/10.1002/cncr.35702.

[9] P. Chakraborty, T. Chandrapragasam, A. Arunachalam, and S. Rafiammal, "Artificial Intelligence-based Oral Cancer Screening System using Smartphones," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12054–12057, Dec. 2023, https://doi.org/10.48084/etasr.6364.

[10] E. Ramesh, A. Ganesan, K. C. Lakshmi, and P. M. Natarajan, "Artificial intelligence—based diagnosis of oral leukoplakia using deep convolutional neural networks Xception and MobileNet-v2," *Frontiers in Oral Health*, vol. 6, Mar. 2025, Art. no. 1414524, https://doi.org/10.3389/froh.2025.1414524.

[11] M. Kantharimuthu, M. M, S. P, A. A. M. G, K. B. S. N, and J. D. K, "Oral Cancer Prediction Using a Probability Neural Network (PNN)," *Asian Pacific Journal of Cancer Prevention*, vol. 24, no. 9, pp. 2991–2995, Sept. 2023, https://doi.org/10.31557/APJCP.2023.24.9.2991.

[12] H. M. Afify, K. K. Mohammed, and A. Ella Hassanien, "Novel prediction model on OSCC histopathological images via deep transfer learning combined with Grad-CAM interpretation," *Biomedical Signal Processing and Control*, vol. 83, May 2023, Art. no. 104704, https://doi.org/10.1016/j.bspc.2023.104704.

[13] W. Yuan *et al.*, "Noninvasive diagnosis of oral squamous cell carcinoma by multi-level deep residual learning on optical coherence tomography images," *Oral Diseases*, vol. 29, no. 8, pp. 3223–3231, 2023, https://doi.org/10.1111/odi.14318.

[14] R. Javed, M. S. M. Rahim, T. Saba, and A. Rehman, "A comparative study of features selection for skin lesion detection from dermoscopic images," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, no. 1, Dec. 2019, Art. no. 4, https://doi.org/10.1007/s13721-019-0209-1.

[15] T. Saba, "Automated lung nodule detection and classification based on multiple classifiers voting," *Microscopy Research and Technique*, vol. 82, no. 9, pp. 1601–1609, Sept. 2019, https://doi.org/10.1002/jemt.23326.

[16] N. S. Piyarathne *et al.*, "A comprehensive dataset of annotated oral cavity images for diagnosis of oral cancer and oral potentially malignant disorders," *Oral Oncology*, vol. 156, Sept. 2024, Art. no. 106946, https://doi.org/10.1016/j.oraloncology.2024.106946.