Deep Representation Learning-Based Oral Potentially Malignant Disorders Detection Model

Samah Alzanin
Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Kharj, Saudi Arabia
s.alzanin@psau.edu.sa (corresponding author)

Mohammed Alonazi
Department of Information Systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia
mn.alonazi@psau.edu.sa

Received: 2 February 2026 | Revised: 25 February 2026, 6 March 2026, and 17 March 2026 | Accepted: 27 March 2026

 DOI: https://doi.org/10.48084/etasr.17920

ABSTRACT

Clinical Decision Support Systems (CDSSs) play a crucial role in modern healthcare by enabling health professionals to efficiently analyze patient data and make accurate, evidence-based clinical decisions. In the context of CDSSs, the analysis of Oral Potentially Malignant Disorders (OPMDs) has seen advances through digital technologies, as Computer-Aided Diagnosis (CAD) techniques that incorporate Artificial Intelligence (AI) and image processing play a vital role in early detection. Deep Learning (DL) in OPMD diagnosis has the capacity to handle intricate patterns, employ differences in image quality, and continuously improve with more data. Incorporation of DL into oral healthcare not only improves diagnostic accuracy but also has the potential to streamline screening, minimize human errors, and provide earlier intervention, eventually improving patient outcomes and supporting the overall treatment of oral health conditions. This study presents an Improved Oral Potentially Malignant Disorder Diagnosis using a Stacked Sparse Autoencoder (IOPMDD-SSAE) model to support clinical decisions in the recognition and classification of OPMD. Oral images of patients can be uploaded to a CDSS, where IOPMDD-SSAE can analyze them, offering an accurate and automatic investigation. IOPMDD-SSAE uses a Wiener Filtering (WF) technique to remove noise. The complex and intrinsic features of the oral images are captured by an SE-ResNet model. Finally, OPMD detection and classification take place using an SSAE. A comparison of IOPMDD-SSAE with existing methods demonstrated a superior accuracy of 98.08% on the Oral cancer (Lips and Tongue) images dataset.

Keywords-oral cancer; Internet of Things (IoT); Wiener filtering; computer-aided diagnosis; clinical decision support

I. INTRODUCTION

Oral cancer is a general term that refers to cancer of the salivary glands, lips, hard palate, soft palate, mouth floor, gingiva, and tonsils. Many people suffer from Oral Squamous Cell Carcinoma (OSCC) [1]. The occurrence of oral cancer increases with age and displays differences in physical distribution. Oral Potentially Malignant Disorders (OPMDs) are a heterogeneous set of clinically definite states linked to a high risk of progression to OSCC [2]. These clinical states require an operation to evaluate the level of dysplasia, if necessary. Once the expert pathologist has made the diagnosis, it must be linked with the medical features, and a complete analysis is required for the lesion at an exact place [3]. This complete analysis provides the physician who handles the patient with the finest data available to judge. Histology is significant, and the absence of medical records hinders the progress of a structural report [4]. Furthermore, it is significant to recognize that not only the place of the original lesion requires careful evaluation, but the complete oral mucosa can also change and show greater exposure to carcinoma growth.

Machine Learning (ML) has been developed as an assistant for medical analysis [5]. AI has gained wide attention to attain cognitive functions of people, such as diversity, learning, and problem-solving. Convolutional Neural Networks (CNNs), which are capable of identifying and classifying a few tumorous lesions, drive Computer-Aided Diagnosis (CAD) systems.

The credibility and widespread adoption of Clinical Decision Support Systems (CDSSs) are influenced not only by their diagnostic accuracy, but also by their ability to support direct and meaningful interaction with clinicians. To effectively integrate a CDSS into routine clinical workflows, it must incorporate several essential design features. Foremost among them is transparency, which enables clinicians to understand the reasoning underlying system recommendations, as opaque "black-box" models are generally unacceptable in clinical practice. In addition, clinical knowledge and information should be presented in a clear, relevant, and actionable manner that aligns with the user's level of expertise. Ultimately, a CDSS should be designed to support and enhance clinical decision-making, helping healthcare professionals rather than trying to replace human judgment or expertise.

For accurate detection and classification of oral disorders, this study presents the Improved Oral Potentially Malignant Disorder Diagnosis using a Stacked Sparse Autoencoder (IOPMDD-SSAE) model. This study is based on a DL architecture that integrates Wiener filtering (WF)-based preprocessing, SE-ResNet feature extraction, and SSAE-based classification for the detection of potentially malignant oral disorders. This hybrid framework enables enhanced representation of lesion characteristics and improves the diagnostic performance within a medical decision support architecture. The key functional contributions include:

- WF is employed to eliminate noise and improve oral image quality, enhancing diagnostic performance.
- A SE-ResNet-based deep feature extraction model captures complex and intrinsic lesion characteristics, enabling effective representation of pathological patterns.
- A Stacked Sparse Autoencoder (SSAE) is utilized to perform precise detection and classification of OPMDs.
- The integration of SE-ResNet with SSAE allows enhanced representation of subtle oral lesion patterns while conserving significant structural details in medical images. This hybrid model improves detection performance by exploiting feature learning and sparse representation-based classification.
- Extensive experimental evaluations demonstrate that the IOPMDD-SSAE model achieves superior performance compared to existing state-of-the-art models across several evaluation metrics.

II. RELATED WORKS

In [6], a DL model depended on the analysis of images captured through smartphones. In addition, a resampling model was developed to improve the quality of images from handheld smartphone cameras. In [7], an optimum Inception-DCNN for OPMD Detection (OIDCNN-OPMDD) approach was presented, where the feature extraction and identification procedure were executed using an IDCNN. In [8], DL models utilized effective feature extraction approaches, including appearance and pattern-based features, and the Bee Pulse Couple NN (BeePCNN) technique was employed to pick the finest features. Finally, an advanced FGPSOCNN decreased the computational cost of the CNN. In [9], patient surveys and oral cavity images were incorporated with ResNet50, ResNet101, VGG19, and ResNet152 models to classify oral cavity images. In [10], a lightweight DCNN approach, based on the pre-trained EfficientNetB0, was used to classify oral lesions as malignant, benign, or possibly malignant using real medical images. In [11], SFO was utilized with a Fusion-based Classification (CADOC-SFOFC) approach, with the fusion-based feature extraction procedure utilizing VGGNet16 and ResNet. In [12], an oral cancer analysis method was based on few-shot learning.

In [13], oral cancer detection was performed using a CNN and an improved DBN method, with their parameters improved using a combination of PSO and Al-Biruni Earth Radius (BER) optimizer models, termed PSOBER. In [14], Electrical Impedance Spectroscopy (EIS) was utilized with a CNN-based DL method. In [15], smartphone-based web-application platforms were developed with DenseNet201 and a modified Capsule Network (FixCaps) model. In [16], the risk of malignant transformation was predicted using a multimodal attention-based deep Multiple Instance Learning (COCOH) method that combined histopathology and immunohistochemistry whole-slide images. In [17], a DL framework with a Cross-Channel Self-Attention Fusion Unit (CCSAFU) model was developed for more precise and reliable detection. In [18], a DL model was used with mixup data augmentation and negative sample utilization to mitigate false positives and improve detection robustness.

Many of these studies are restricted due to high imbalance, small datasets, and inconsistent image quality, affecting the robustness of the model. Few-shot learning and EIS-CNN approaches partially address data scarcity but may not generalize across the population. In addition, although some models are lightweight, there is a research gap in achieving real-time applicability and reliable multimodal integration. Previous works on OSCC detection are mainly based on traditional preprocessing and DL models, which do not effectively resolve image noise and subtle lesion differences, affecting diagnostic reliability. Besides, several methods are mainly based on traditional CNN models without incorporating attention mechanisms to highlight discriminative features in oral images. Moreover, the limited integration of advanced representation learning and sparse-based classification methods limits the ability of the model to distinguish normal and cancer images proficiently. Therefore, there is a need for a hybrid framework that includes robust denoising, attention-enhanced deep feature extraction, and efficient classification for improved oral cancer detection.

III. THE PROPOSED METHOD

This study introduces a clinically driven intelligent system for automated OPMD diagnosis, designed to support decision-making in real-world healthcare settings. The proposed IOPMDD-SSAE framework integrates noise-resilient preprocessing using WF with advanced feature learning through an SE-ResNet network. Leveraging an SSAE for classification, the system provides an accurate and automatic analysis of oral images uploaded by patients or clinicians. Figure 1 illustrates the entire process of IOPMDD-SSAE.

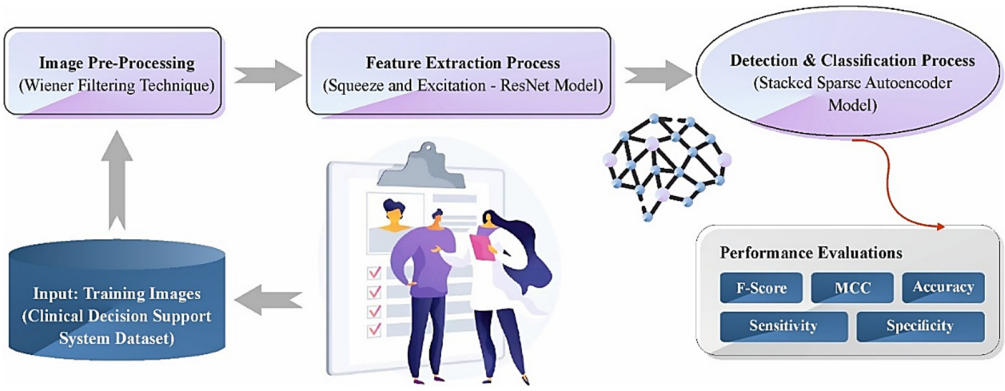

Fig. 1. Overall process of the IOPMDD-SSAE system.

A. WF-Based Preprocessing

At the primary level, the IOPMDD-SSAE technique uses WF for noise elimination [19], selected for its ability to improve image clarity without over-smoothing. This technique employs statistical data for the image and noise to make an optimum linear filter. With the help of the power spectrum of the new image and noise, WF aims to decrease the Mean Squared Error (MSE) between the true image and its filtered form, minimizing noise while maintaining significant details. WF finds applications in diverse image processing activities such as satellite images, medical imaging, and photography, where the elimination of noise is crucial to improve overall visual quality and to support the following interpretation or exploration.

B. Feature Extraction Using SE-ResNet

The complex and intrinsic features of oral images can be captured by the SE-ResNet module [20], which improves relevant patterns and classification performance at a lower computational cost. SE-ResNet is based on a convolutional process that extracts useful features by combining spatial and channel-wise data in locally accessible areas. The fundamental module is a mixture of the SE block and the ResNet residual block. According to CNN theory, the convolution operator applies a transformation: $T: X \to O$, $X \in \mathbb{R}^{C' \times H' \times W'}$, $O \in \mathbb{R}^{C \times H \times W}$. Let L be the final convolution layer in the SE-ResNet module. Assume X_0 the input and $X = [x^1, x^2, x^{C'}]$ the input of L. Let $K = [k_1, k_2, \dots, k_C]$ be the filter kernels, where k_c denotes the c^{th} filter parameters. The L output is defined as $O = [o_1, o_2, \dots, o_C]$ where:

$$o_c = k_c * X = \sum_i^{C'} k_c^i * x^i \quad (1)$$

where $*$ represents the convolution, $k_c = [k_c^1, k_c^2, k_c^{C'}]$, with k_c^i specifying the $2D$ kernel, and k_c denotes a single channel that performs on the equivalent network of X. The output is produced by the biased outline of the entire input network, and channel dependencies are indirectly fixed in k_c, but they are tangled with the spatial correlation. The SE block adjusts per-channel feature responses by clearly demonstrating inter-dependencies among channels. The filter response contains dual phases, such as squeeze and excitation. Officially, a statistic $S = [s_1, s_2, \dots, s_C] \in \mathbb{R}^C$ is produced by reducing O over spatial sizes $H \times W$, where:

$$s_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} o_c(i, j) \quad (2)$$

To utilize the data combined from the squeeze process, a second phase aims to remove individual channel dependencies completely. This phase employs a Fully Connected (FC) NN with two Hidden Layers (HL) to absorb the non-linear interface and non-mutually special connection among them. The FC NN output is defined as:

$$\tilde{S} = \sigma\big(W_2 d(W_1 S)\big) \quad (3)$$

where δ denotes the ReLU activation function, σ specifies the sigmoid function, $W_1 \in \mathbb{R}^{\overline{C} \times C}$, $W_2 \in \mathbb{R}^{C \times \overline{C}}$, $\overline{C} = C/r$, and r refers to the reduction ratio. Thus, L can be rewritten as $\tilde{O} = [\tilde{o}_1, \tilde{o}_2, \dots, \tilde{o}_C]$, while

$$\tilde{o}_c = \tilde{s_c} \cdot o_c \quad (4)$$

Here $\tilde{s}_c \in \tilde{S}$ and $\tilde{o}_c$ states to the individual channel multiplication among the scalar $\tilde{s}_c$ and mapping feature o_c. Supposing the input and output sizes are similar, the last output of SE-ResNet is written as:

$$\tilde{X} = X_0 + \tilde{O} \quad (5)$$

After training a set of images, the objective function calculates the space between the forecast and objective outcomes, giving a loss value by Backpropagation (BP):

$$\frac{\partial \tilde{X}}{\partial X_0} = \frac{\partial (X_0 + \tilde{O})}{\partial X_0} = 1 + \frac{\partial \tilde{O}}{\partial X_0} \quad (6)$$

The SR certifies that the gradient is constantly bigger or equivalent to 1 from the BP, evading the CNN gradient vanishing issue. The important change is that SE-ResNet uses a global average pooling process from the squeeze stage.

C. Classification Using the SSAE Model

Finally, the detection and classification of OPMD takes place using the SSAE [21] model. This model efficiently learns intrinsic and discriminative data transformations by using sparse encoding and mitigates repetition, thereby improving robustness against overfitting. AE is an unsupervised NN that incorporates a decoder and an encoder. The decoder recognizes connections among features, reducing data dimensionality. Then, the encoder rebuilds novel data from the condensed features. The evaluation of condensed features includes computing the reconstruction error between the output and

input. Assuming x the unlabeled input data with N instances, the decoder and encoder formulations are:

$$h = f(x) = \sigma_e(W_e x + b_e) \quad (7)$$

$$y = \hat{x} = g(h) = \sigma_d(W_d h + b_d) \quad (8)$$

where σ_e and σ_d refer to non-linear activation functions, such as ReLU, tanh, or sigmoid, h denotes the output of the HL, W_d and W_e denote the weights and b_d and b_e represent the biases, and y denotes the output. The AE error, denoted as loss, is calculated utilizing the MSE function:

$$L(W, b) = \frac{1}{N}\Sigma(y - x)^2 \quad (9)$$

The AE aims to generate a y that estimates the input x, diminishing the constructed error of $L(W, b)$. SAE inserts a sparsity restriction on AE, attaining the outcome of sparsity and feature size decrease. To achieve this result, ρ is selected to a value near 0 (usually 0.05).

$$\rho = \hat{\rho}_j = \frac{1}{N}\sum_{i=1}^{N} f_j \cdot (x_i) \quad (10)$$

where $f_j(\cdot)$ denotes the input. Next, the KL divergence is employed to find $\hat{\rho}_j$ reverse from ρ, as:

$$KL(\rho \| \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho)\log \frac{1-\rho}{1-\hat{\rho}_j} \quad (11)$$

The SAE reconstructed error can be formulated as:

$$L_{sparse}(W, b) = J(W, b) + \beta \sum_{j=1}^{k} KL(\rho \| \hat{\rho}_j) \quad (12)$$

where k represents the number of neurons, and β denotes the penalty parameter. A common utilized technique is to stack numerous SAE encoder designs, i.e., SSAE, whereas the HL outcome of the preceding SAE is the subsequent SAE input. In an unsupervised method, the SAE is trained layer-wise, utilizing a greedy algorithm for minimizing the loss between the output and input of every SAE, which aids in decreasing the loss rate of the feature element.

IV. PERFORMANCE ANALYSIS

The proposed IOPMDD-SSAE model was experimentally evaluated on a benchmark image dataset [22]. The method was run on Python 3.6.5 with an i5-8600k CPU, 4 GB GPU, 16 GB RAM, 250 GB SSD, and 1TB HDD, using a 0.01 learning rate, ReLU, 50 epochs, 0.5 dropout, and batch size 5. Overfitting was reduced using WF-based noise removal and SE-ResNet deep feature extraction with SSAE-based classification, while strict data splitting prevented leakage and class weighting addressed imbalance. Table I describes the dataset, which comprises 131 images with two classes.

TABLE I. DATASET DESCRIPTION

Classes	Instance count
Cancer	87
Non-Cancer	44
Total Instances	**131**

Figure 2 presents the confusion matrices attained by the IOPMDD-SSAE model at 80:20 and 70:30 training/test (TRAPH/TESPH) splits.

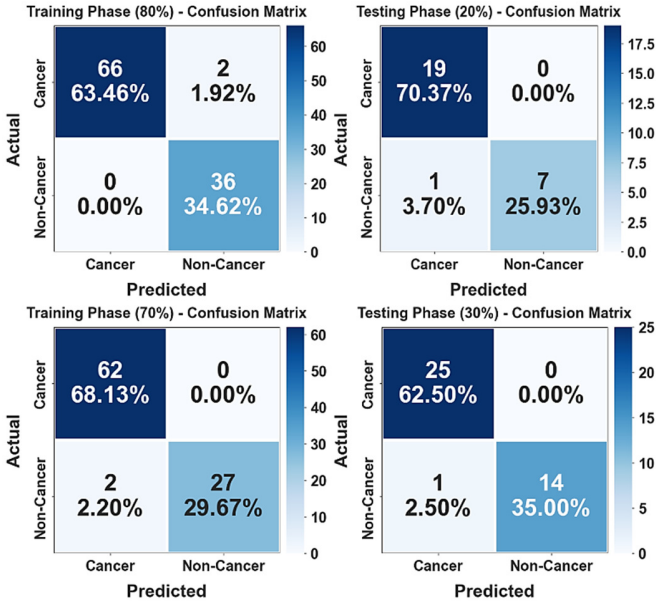

Fig. 2. Confusion matrices of (a, c) TRAPH for 80% and 70% and (b, d) TESPH for 20% and 30%.

Table II shows the detection outputs of the IOPMDD-SSAE. According to 80% TRAPH, the proposed technique achieved an average $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 98.08%, 98.53%, 98.53%, 97.90%, and 95.89%, respectively. Meanwhile, based on 20% TESPH, the IOPMDD-SSAE method attained average $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 96.30%, 93.75%, 93.75%, 95.38%, and 91.17%, respectively. Finally, based on 70% TRAPH, the IOPMDD-SSAE method gained an average $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 97.80%, 96.55%, 96.55%, 97.42%, and 94.97%, whereas on 30% of TESPH, the IOPMDD-SSAE method achieved an average $accu_y$, $sens_y$, $spec_y$, F_{score}, and MCC of 97.50%, 96.67%, 96.67%, 97.30%, and 94.73%, respectively.

TABLE II. DETECTION RESULT OF THE IOPMDD-SSAE TECHNIQUE WITH DISTINCT CLASSES

Classes	$Accu_y$	$Sens_y$	$Spec_y$	F_{Score}	MCC
TRAPH (80%)					
Cancer	98.08	97.06	100.00	98.51	95.89
Non-Cancer	98.08	100.00	97.06	97.30	95.89
Average	98.08	98.53	98.53	97.90	95.89
TESPH (20%)					
Cancer	96.30	100.00	87.50	97.44	91.17
Non-Cancer	96.30	87.50	100.00	93.33	91.17
Average	96.30	93.75	93.75	95.38	91.17
TRAPH (70%)					
Cancer	97.80	100.00	93.10	98.41	94.97
Non-Cancer	97.80	93.10	100.00	96.43	94.97
Average	97.80	96.55	96.55	97.42	94.97
TESPH (30%)					
Cancer	97.50	100.00	93.33	98.04	94.73
Non-Cancer	97.50	93.33	100.00	96.55	94.73
Average	97.50	96.67	96.67	97.30	94.73

Table III presents a comprehensive comparative analysis of the IOPMDD-SSAE method. The results highlight that IOPMDD-SSAE achieved improved performance.

TABLE III. COMPARATIVE RESULTS OF IOPMDD-SSAE WITH EXISTING MODELS

Methods	$Accu_y$	$Sens_y$	$Spec_y$	F_{Score}
IOPMDD-SSAE	98.08	98.53	98.53	97.90
OIDCNN-OPMDD	97.50	97.83	97.83	97.46
DBN Method	86.36	84.12	91.15	85.74
CNN Technique	94.14	93.93	96.89	95.39
Inceptionv4	85.14	86.68	89.42	87.24
DenseNet161	90.06	88.21	85.59	86.22

The IOPMDD-SSAE model gained a higher $accu_y$ of 98.08%, whereas the OIDCNN-OPMDD, DBN, CNN, Inceptionv4, and DenseNet-161 models obtained lower $accu_y$ values of 97.50%, 86.36%, 94.14%, 85.14%, and 90.06%, respectively. IOPMDD-SSAE achieved an increased F_{score} of 97.90%, whereas the OIDCNN-OPMDD, DBN, CNN, Inceptionv4, and DenseNet-161 models had lower F_{score} values of 97.46%, 85.74%, 95.39%, 87.24%, and 86.22%. IOPMDD-SSAE attained an improved $sens_y$ of 98.53%, whereas the OIDCNN-OPMDD, DBN, CNN, Inceptionv4, and DenseNet-161 models had lower $sens_y$ values of 97.83%, 84.12%, 93.93%, 86.68%, and 88.21%, respectively. Finally, IOPMDD-SSAE achieved a higher $spec_y$ of 98.53%, whereas the OIDCNN-OPMDD, DBN, CNN, Inceptionv4, and DenseNet-161 models achieved lower $spec_y$ values of 97.83%, 91.15%, 96.89%, 89.42%, and 85.59%, respectively. These experimental results demonstrate that the IOPMDD-SSAE method achieved increased performance.

Table IV describes an ablation study assessment of the IOPMDD-SSAE method. The SSAE with SE-ResNet feature extraction attained an $Accu_y$ of 98.08%, a precision of 98.53%, a recall of 98.53%, and an F_{score} of 97.90%, respectively. Furthermore, by removing the SE module and using only SSAE, lower performance was obtained with an $accu_y$ of 97.33%, $sens_y$ of 97.09%, $spec_y$ of 97.44%, and F_{score} of 97.11%, thus emphasizing the contribution of the SE component.

TABLE IV. ABLATION STUDY OF IOPMDD-SSAE

Methods	$Accu_y$	$Sens_y$	$Spec_y$	F_{Score}
IOPMDD-SSAE (SSAE with SE-ResNet feature extraction)	98.08	98.53	98.53	97.90
SSAE+ResNet (Without SE model)	97.33	97.09	97.44	97.11

V. CONCLUSION

This study presented the design of the IOPMDD-SSAE technique for clinical decision support, aiming to enable the recognition and classification of OPMD. In the proposed technique, oral images of patients can be uploaded to a CDSS, where IOPMDD-SSAE can analyze them, offering an accurate and automatic investigation. IOPMDD-SSAE uses WF for noise elimination, while the complex and intrinsic features of the oral images are captured by the SE-ResNet model. Finally, detection and classification of OPMD is performed using an SSAE model.

A comparison analysis demonstrated a superior accuracy of 98.08% over existing methods on the Oral cancer (Lips and Tongue) images dataset [22]. The limitations of this study include small sample sizes, class imbalance, and image discrepancy, affecting model accuracy and generalization, while research gaps remain in standardizing data collection and improving robustness for real-world clinical deployment. Future work will focus on incorporating larger datasets and more extensive validation strategies to further improve the generalization capability of the proposed framework.

DATA AVAILABILITY STATEMENT

The data used in this study are available in [22].

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing interests that could have appeared to influence the work reported in this manuscript.

ACKNOWLEDGMENT

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through project number (PSAU/2025/01/37523).

REFERENCES

[1] 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.

[2] S. M. Sagari, V. P. Malagi, and B. Chandrahas, "A Novel Feature Extraction Approach Using Deformable Adaptive Instance-Based U-Net Architecture for Segmentation and Classification of Oral Mucosal Lesion," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25228–25234, Aug. 2025, https://doi.org/10.48084/etasr.11273.

[3] G. Tanriver, M. S. Tekkesin, and O. Ergen, "Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders," *Cancers*, vol. 13, no. 11, June 2021, https://doi.org/10.3390/cancers13112766.

[4] M. Parola, F. A. Galatolo, G. La Mantia, M. G. C. A. Cimino, G. Campisi, and O. Di Fede, "Towards explainable oral cancer recognition: Screening on imperfect images via Informed Deep Learning and Case-Based Reasoning," *Computerized Medical Imaging and Graphics*, vol. 117, Oct. 2024, Art. no. 102433, https://doi.org/10.1016/j.compmedimag.2024.102433.

[5] X. Liang, Q. Chen, Y. Cao, B. Liu, S. Chen, and X. Liu, "Automated Scene Classification in Endoscopy Videos Using Convolutional Neural Networks," in *2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, June 2024, pp. 157–161, https://doi.org/10.1109/CHASE60773.2024.00026.

[6] E. S. Mira *et al.*, "Early Diagnosis of Oral Cancer Using Image Processing and Artificial Intelligence," *Fusion: Practice and Applications*, vol. 14, no. 1, pp. 293–308, 2024, https://doi.org/10.54216/FPA.140122.

[7] R. Alabdan, A. Alruban, A. M. Hilal, and A. Motwakel, "Artificial-Intelligence-Based Decision Making for Oral Potentially Malignant Disorder Diagnosis in Internet of Medical Things Environment," *Healthcare*, vol. 11, no. 1, Dec. 2022, https://doi.org/10.3390/healthcare11010113.

[8] R. Dharani, S. Revathy, and K. Danesh, "Fuzzy Genetic Particle Swarm Optimization Convolution Neural Network Based On Oral Cancer Identification System," *Journal of Applied Engineering and Technological Science (JAETS)*, vol. 5, no. 1, pp. 150–169, Dec. 2023, https://doi.org/10.37385/jaets.v5i1.2874.

[9] M. Shariff, P. Singh SM, S. DP, V. MH, and A. S. Poornima, "Convolutional neural network for detection of oral cavity leading to oral cancer from photographic images," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 865–877, 2024.

[10] F. Jubair, O. Al-karadsheh, D. Malamos, S. Al Mahdi, Y. Saad, and Y. Hassona, "A novel lightweight deep convolutional neural network for early detection of oral cancer," *Oral Diseases*, vol. 28, no. 4, pp. 1123–1130, 2022, https://doi.org/10.1111/odi.13825.

[11] M. Al Duhayyim *et al.*, "Sailfish Optimization with Deep Learning Based Oral Cancer Classification Model," *Computer Systems Science and Engineering*, vol. 45, no. 1, pp. 753–767, 2023, https://doi.org/10.32604/csse.2023.030556.

[12] Z. Guo, S. Ao, and B. Ao, "Few-shot learning based oral cancer diagnosis using a dual feature extractor prototypical network," *Journal of Biomedical Informatics*, vol. 150, Feb. 2024, Art. no. 104584, https://doi.org/10.1016/j.jbi.2024.104584.

[13] 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, https://doi.org/10.1109/ACCESS.2023.3253430.

[14] Z. Lin *et al.*, "Deep learning-based electrical impedance spectroscopy analysis for malignant and potentially malignant oral disorder detection," *Scientific Reports*, vol. 15, no. 1, June 2025, Art. no. 19458, https://doi.org/10.1038/s41598-025-05116-8.

[15] K. M. Desai *et al.*, "Screening of oral potentially malignant disorders and oral cancer using deep learning models," *Scientific Reports*, vol. 15, no. 1, May 2025, Art. no. 17949, https://doi.org/10.1038/s41598-025-02802-5.

[16] J. Adeoye and Y. X. Su, "COCOH: A Multimodal Deep Learning Framework for Cancer Risk Assessment of Oral Potentially Malignant Disorders." medRxiv, Jan. 05, 2026, https://doi.org/10.64898/2026.01.04.26343399.

[17] X. Liang *et al.*, "Enhancing polyp detection in endoscopy with cross-channel self-attention fusion," *Smart Health*, vol. 36, June 2025, Art. no. 100578, https://doi.org/10.1016/j.smhl.2025.100578.

[18] Y. He, Q. Chen, Z. Xiong, X. Liang, Y. Cao, and B. Liu, "One-stage Framework for Thyroid Nodule Detection with Mixup and Negative Sample Utilization," in *2025 IEEE International Conference on Image Processing (ICIP)*, Sept. 2025, pp. 2205–2210, https://doi.org/10.1109/ICIP55913.2025.11084660.

[19] R. Hardie, "A Fast Image Super-Resolution Algorithm Using an Adaptive Wiener Filter," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2953–2964, Sept. 2007, https://doi.org/10.1109/TIP.2007.909416.

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 7132–7141, https://doi.org/10.1109/CVPR.2018.00745.

[21] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy Layer-Wise Training of Deep Networks," in *Advances in Neural Information Processing Systems*, 2006, vol. 19.

[22] "Oral Cancer (Lips and Tongue) images." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/shivam17299/oral-cancer-lips-and-tongue-images.