Research Paper

Reinforcement Learning Guided Hierarchical Multimodal Feature Representation for Oral Cancer Diagnosis via Histopathological Image Analysis

Sherif Baig Ibrahim[1*], **Vivekanandan Kalimuthu**[2]

1. *Department of Computer Science and Engineering, Puducherry Technological University, Puducherry, India.*

2. *Department of Computer Science and Engineering, Puducherry Technological University, Puducherry, India*

Citation Ibrahim SB, Kalimuthu V. Reinforcement Learning Guided Hierarchical Multimodal Feature Representation for Oral Cancer Diagnosis via Histopathological Image Analysis. International Journal of Medical Toxicology and Forensic Medicine. 2026; 16:E52672.

https://doi.org/10.22037/ijmtfm.v16.52672

Article info:

Received: 01 July, 2026

Accepted: 23 July, 2026

Published: 28 August, 2026

Keywords:

Oral cancer, Histopathological images, Multimodal feature extraction, Deep learning, Deep Q-Network, Artificial intelligence

A B S T R A C T

Background: Oral cancer (OC) is a major public health concern due to its high mortality rate. Early detection is crucial to improving the survival chances of affected individuals. Conventionally, histopathological examination is used for the diagnosis of OC. However, existing approaches still encounter challenges in capturing complex histopathological patterns and achieving consistent performance across varying image conditions.

Methods: This study proposes a Reinforcement Learning-Based Multimodal Feature Extraction Approach for Early Oral Cancer Detection (RLMFE-OCD) model. The proposed method integrates multiple deep learning models, including GoogLeNet, ResNet-50, and SE-VGG16, to extract diverse and discriminative feature representations. A Deep Q-Network is employed for classification, while the Butterfly Optimization Algorithm is used for hyperparameter optimization.

Results: Experimental results on benchmark datasets establish that the RLMFE-OCD model achieves higher performance, with an accuracy of 94.44% compared with existing methods.

Conclusion: The RLMFE-OCD framework effectively supports early oral cancer detection through multimodal feature extraction and reinforcement learning. It shows potential as a reliable computer-aided diagnostic approach.

*** Corresponding Author:**

Sherif Baig Ibrahim, PhD

Department of Computer Science and Engineering, Puducherry Technological University, Puducherry, India.

E-mail: *sherifbaig1@gmail.com*

Copyright © 2026 The Author(s);

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-By-NC: https://creativecommons.org/licenses/by-nc/4.0/legalcode.en), which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Introduction

Oral cancer (OC) is a significant disorder that is often difficult to control in the later phases. Approximately 475,000 new cases of head and neck cancers are reported in every year. In addition to the OC, oral squamous cell carcinoma (OSCC) is a very common oral malignancy that occurs in the mouth cavity that happens when numerous genetic alterations accumulate in the cells, leading to epithelial damage. The majority of individuals are undiagnosed till the later phases of the disease, when medical treatments are restricted, and a poor prognosis [1]. Employing HI for cancer screening helps improve and predict anomalies. The applications of AI in OC diagnosis have been rapidly adopted, with notable successes in the early interpretation of medical imaging. DL methods, particularly CNNs, have become sophisticated for numerous image analysis tasks in recent years. Recent problems in the field of image diagnosis use these methods to tackle the classification problem. Because the DL method's efficacy is primarily based on obtaining an architecture that meets the difficult, numerous research studies are presently in progress to develop innovative and more intricate deep networks to improve the predictive results [2].

A brief review of related works on Oral Cancer Diagnosis. Appavu [3] proposed an innovative DL-driven method for the categorization and timely detection of OC employing medical imagery. Gab Allah et al. [4] proposed an edge-supervised DL structure for the automatic classification of OC using HI. Dhal et al. [5] proposed a lightweight, ensemble-driven DL framework for OC prediction. Moreover, the presented BiGRU block improves classification accuracy by leveraging sequential and spatial characteristics to better capture dependencies in image data. The authors [6] introduced an enhanced DL ensemble approach for OC classification. The recommended method amalgamates improved EfficientNet-B5, enhanced with SE, exploiting their synergistic benefits for accurate lesion classification. Shukla et al. [7] suggested an innovative approach that combines machine vision for cancer detection. Santhiya et al. [8] employed a DL method including classification, preprocessing, and feature extraction. Lastly, a modified CNN is used to classify the OC cases precisely. Song et al. [9] presented an innovative pipeline procedure that offers an efficient method for diagnosing OC images. The authors [10] proposed a method for diagnosing OC utilizing AI and image processing techniques. Then, the DenseNet169 method has been used to extract a valuable set of deep features.

Despite the significant advancements of DL method in the medical image investigations, most of the existing OC identification models rely on single CNN architectures, which limit their feature representation capability. Histopathological images usually contain overlapping and subtle patterns, which makes the classification more challenging. Therefore, this paper introduces a Reinforcement Learning-Based Multimodal Feature Extraction Approach for Early Oral Cancer Detection (RLMFE-OCD) method Using Medical Images. This study makes the following key contributions:

- The fusion of feature extraction techniques, namely GoogLeNet, ResNet-50, and SE-VGG16, enables the efficient extraction of important features from complex oral lesion images.

- Deep Q-networks (DQN) are employed to learn patterns and make reliable predictions effectively, which helps in classifying oral cancer.

- The butterfly optimization algorithm (BSA) is applied in the parameter tuning of the model.

The comparative analysis showed improvements in the RLMFE-OCD model across several measures.

Materials and Methods

Figure 1 shows the overall workflow of the proposed RLMFE-OCD method for timely OC employing histopathological images. Firstly, image preprocessing is conducted. Then, a multimodal feature extractor is performed using GoogLeNet, ResNet-50, and SE-VGG16. Consequently, classification is performed using a DQN model. The BOA is used to optimize model parameters. Finally, the model results are examined using different metrics.

Feature Representation Models

In this section, important features are extracted using a combination of GoogLeNet, ResNet-50, and SE-VGG16 models [11, 12].

GoogLeNet Model

This model is also known as Inceptionv1, a 22-layer deep network. It is a response to restrictions of preceding CNN structures, mainly in terms of computational efficacy and depth. GoogLeNet's main advanced feature is 9 inception modules, which form the structure. The outputs are concatenated and used as input for the following layer. At last, an attained feature mapping was compressed into a 1D vector and categorized by utilizing softmax.

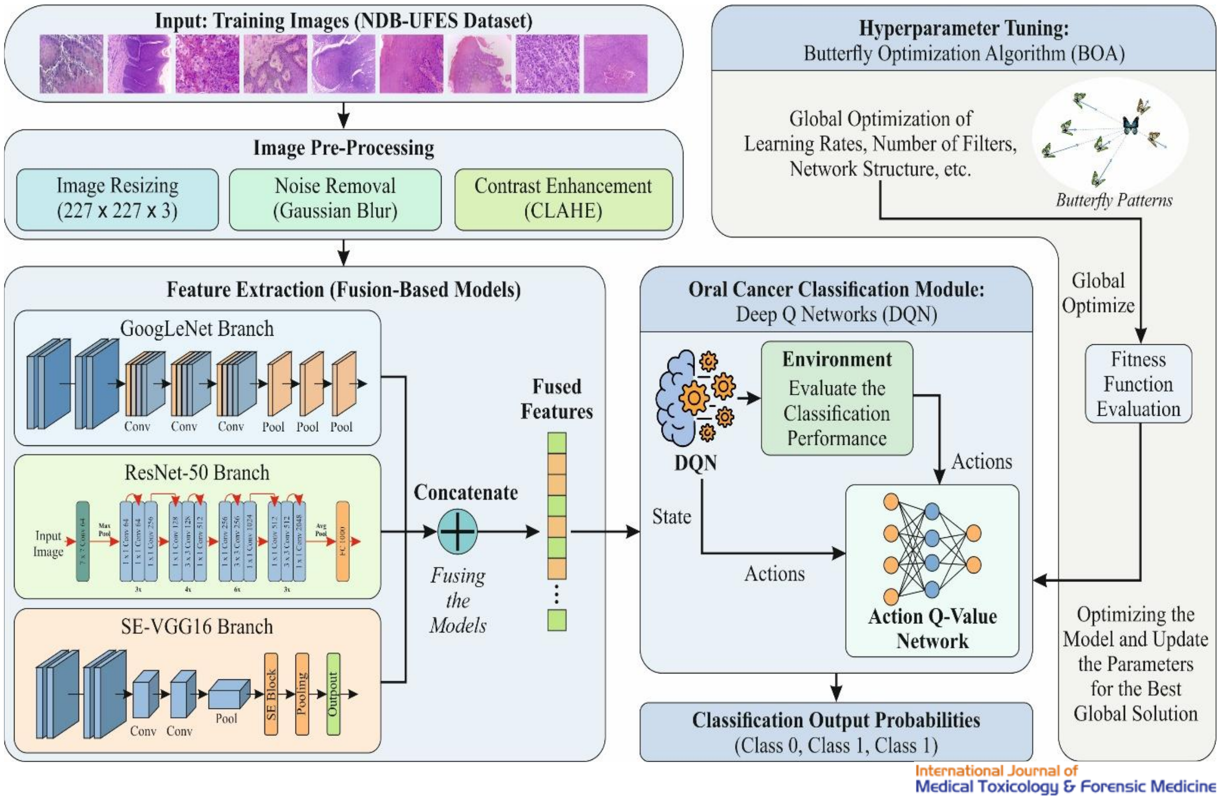


Figure 1. Process diagram of the proposed RLMFE-OCD approach.

ResNet-50 Technique

The ResNet50 includes 50 layers, with 4 residual blocks and 3 convolutional layers. The layers of convolutional are prepared with max pooling operation, ReLU, and batch normalization. Then, they are handled by a residual block, which is an essential element of ResNet50. The convolution block is equivalent to an identity block and contains a 1x1 convolutional layer to reduce the number of filters before the 3×3 convolution layer. Lastly, the structure concludes with a fully connected (FC) layer that plays a vital role in classifying the outcome. From the last FC layer, an output is passed through a softmax.

SE-VGG16 Architecture

The SE-VGG16 model was proposed for oral cancer detection. This structure is based on VGG-16, integrating the SENet module. VGGNet is a structure of CNN that contains convolutional layers and FC layers. The SE blocks improve feature representation by learning channel-wise relationships through FC layers, enabling the network to emphasize important features and improve classification accuracy. A SE block is a computation unit for mapping an input image $x, x \in R^{H' \times W' \times C'}$ to feature $U, U \in R^{H \times W \times C}$ by F_{tr}. In this calculation, F_{tr} acts as an operator, $V = [v_1, v_2, \dots, v_C]$, represented as v_c. Therefore, an output is written as $U = [u_1, u_2, \dots, u_C]$, whereas

$$u_c = v_c \times A = \sum_{s=1}^{C'} v_c^s * a^s \qquad (1)$$

where $*$ means the convolutional, $v_c = [v_C^1, v_C^2, \dots, v_C^{C'}]$, $A = [a^1, a^2, \dots, a^{C'}]$ and $u_C \in R^{H \times W}$. v_C^S denotes a 2D spatial kernel signifying the individual channels v_c. The SE block uses two FC layers with a ReLU activation in between. It first reduces the channel dimensions by a ratio r to lower computation, then restores them to the original size.

$$X_c = F_{scale}(u_c, s_c) = s_c u_c \qquad (2)$$

whereas $\tilde{X} = [x_1, x_2, \dots, x_C]$ and $F_{scale}(u_c, s_c)$ signifies channel multiplication among metrics s_c and feature map $U_C \in R^{H \times W}$.

DQN-based Classification Technique

In this part, the DQN classifies the images and identifies cancerous patterns. To integrate Q-learning with NN results in instability and divergence if completed simply by substituting the Q-table with an NN and implementing standard upgrades of Q-learning. DQN resolves this concern through algorithmic novelty that stabilizes learning [13]. To develop the Q-learning technique, by parametrizing the function of action-value utilizing a DNN $Q(s, b; \theta)$, here $\theta \in \mathbb{R}^d$ denotes network parameters. With a fixed state space, the system becomes tractable by taking states as input and producing Q-values for all discrete actions simultaneously. The problem arises in modifying the update rules of tabular Q-learning to NN.

$$Q(s,b) \leftarrow Q(s,b) + \alpha[r + \gamma \max_{b'} Q(s',b') - Q(s,b)] \quad (3)$$

To become the regression issue with a function of loss:

$$\mathcal{L}(\theta) = \mathbb{E}_{(s,b,r,s')\sim D}[(y - Q(s,b;\theta))^2] \quad (4)$$

Here, the mathematical representation of target $y = r + \gamma \max_{b'} Q(s',b';\theta)$ succeeds from the mathematical form of Bellman optimality.

DQN upholds dual separate networks: online network $Q(s,b;\theta)$ and targeted network $Q(s,b;\theta^-)$ The loss becomes:

$$\mathcal{L}(\theta) = \mathbb{E}_{(s,b,r,s')}\left[\left(r + \gamma \max_{b'} Q(s',b';\theta^-) - Q(s,b;\theta)\right)^2\right] \quad (5)$$

In the gradient:

$$\nabla_\theta \mathcal{L}(\theta) = -2E_{(s,b,r,s')}\left[\left(r + \gamma \max_{b'} Q(s',b';\theta^-) - Q(s,b;\theta)\right)\nabla_\theta Q(s,b;\theta)\right] \quad (6)$$

The target network is kept fixed for stability and periodically updated every Csteps (copying θ^-), making each interval equivalent to one fitted Q-iteration step. To optimize minibatch gradient descent. Specified batch $B = \{(s_i, b_i, r_i, s_i')\}_{i=1}^{B}$:

$$\nabla_\theta \mathcal{L}(\theta) \approx -\frac{2}{B}\sum_{i=1}^{X} \delta_i \nabla_\theta Q(s_i, b_i;\theta) \quad (7)$$

While $\delta_i = r_i + \gamma \max_{b'} Q(s_i', b';\theta^-) - Q(s_i, b_i;\theta)$ denotes temporal difference error.

Parameter Selection using BOA

Finally, the BOA model is used for fine-tuning the model. BOA is stimulated by butterflies' behaviors naturally, which mimic the mating and foraging behaviors for exploration [14]. This technique utilizes scent to pretend fitness value, directing entities to travel in the solution space over dual methods such as global and local searches. The scent was stated as an intensity function of stimulus:

$$f_{BOA} = c \times I^a \quad (8)$$

whereas c denotes a sensory factor. a is a power index. The BOA was mainly divided into 3 phases:

First stage

Initialize the population, objective function, parameters, and much more.

Second stage

Update butterfly positions iteratively and recalculate their fitness values.

A butterfly's insight into scent is passed on to a new butterfly that will travel near the source of scent. This phase is officially named global search:

$$P_i^{T+1} = P_i^T + (r_{BOA^2} \times g^* - P_i^T) \times f_{BOA-i} \quad (9)$$

Here,P_i^T is the position of the ith butterfly at iteration T, $g^*$is the current best solution, $f_{BOA-i}$is its scent value, and $r_{BOA} \in [0,1]$is a random number.

In another scenario, if a butterfly were incapable of perceiving the fragrance produced by another one, it would perform a casual movement. This phase is called local search:

$$P_i^{T+1} = P_i^T + \left(r_{BOA^2} \times P_j^T - P_k^T\right) \times f_{BOA-i} \quad (10)$$

Let, P_k^T and P_j^T represent the kth and jth butterflies,

In BOA, the switch probability($p = 0.8$)controls the transition between global and local search. At each iteration, a random value is generated and compared with p to select the appropriate search strategy.

$$r_{BOA} = rand(0,1) \quad (11)$$

Third stage

The algorithm stops when the maximum iterations are reached and outputs the best optimal solution found.

The BOA model offered a fitness function (FF) for achieving enhanced classifier outcomes. It explains the positive numbers to indicate the candidate outcomes. Here, the classifier rate of error minimization was defined as FF in Eq. (17).

$$fitness(x_i) = ClassifierErrorRate(x_i)$$

$$= \frac{count\ of\ misclassified\ instances}{Total\ count\ of\ instances} * 100 \quad (12)$$

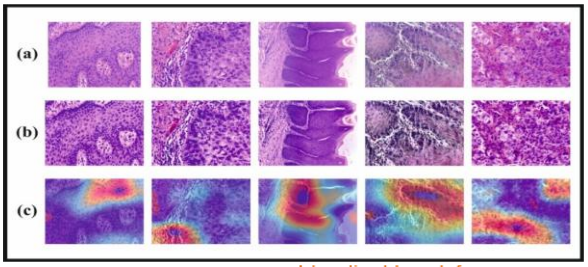
Results

In this section, the experimental investigation of the RLMFE-OCD model is presented using the NDB-UFES dataset [15]. Every case with this dataset is represented by the patients with at least one lesion in the OC that is more than dual HI are acquired; In these recovered slides, overall, 237 images can be acquired with the 10X and 40X objectives of the light microscope, Leica DM500. It has three classes: OSCC, Leukoplakia with dysplasia, and Leukoplakia without dysplasia, as shown in Table 1 below. Figure 2 denotes the sample images of the original, pre-processed, and feature extraction.

Table 1. Details of the dataset.

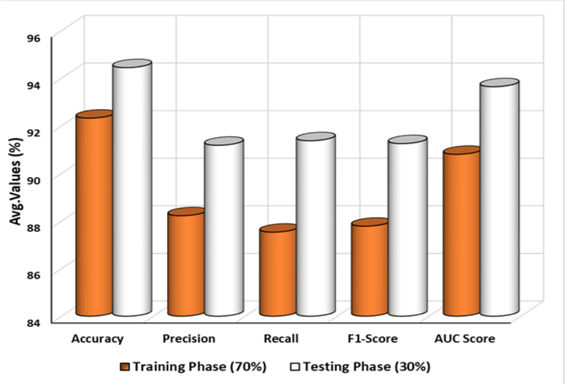
Classes	Labels class	Images
"OSCC"	0	91
"Leukoplakia with dysplasia"	1	89
"Leukoplakia without dysplasia"	2	57
Total Images		**237**

International Journal of
Medical Toxicology & Forensic Medicine



International Journal of
Medical Toxicology & Forensic Medicine

Figure 2. Sample Images of (a) Original, (b) Pre-processed, and (c) Feature Extraction.

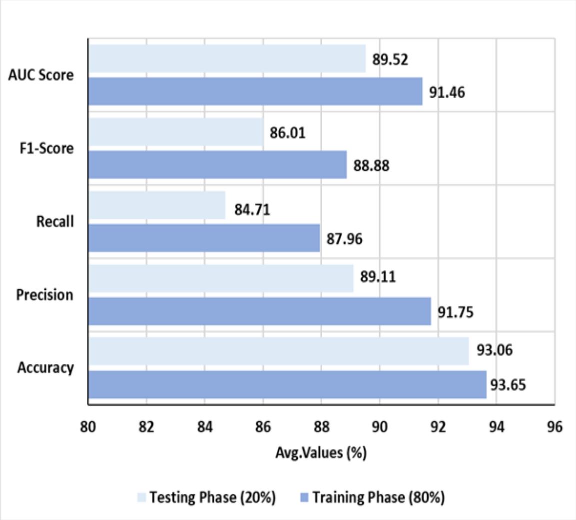


International Journal of
Medical Toxicology & Forensic Medicine

Figure 3. Average values of the RLMFE-OCD model on 70:30 of TRAIP/TESTP.

Figure 3 signify an average value of the RLMFE-OCD technique on 70:30 TRAIP/TESTP. The proposed method achieved average value with $accur_y$ of 92.32% and 94.44% for the TRAIP and TESTP.

Figure 4 portray an average value of the RLMFE-OCD technique on 80:20 of TRAIP/TESTP. The proposed method achieved average value with $accur_y$ of 93.65% and 93.06% for the TRAIP and TESTP.



International Journal of
Medical Toxicology & Forensic Medicine

Figure 4. Average values of the RLMFE-OCD method with 80 :20 of TRAIP/TESTP.

Discussion

To establish the enhanced solution for the RLMFE-OCD system, a brief comparative analysis is presented in Figure 5 [16, 17]. The solutions exhibited that the SqueezeNet, ResNet50, AlexNet, VGG19, ResNet34, Inception, and Regnet-X16GF models have shown lower outcomes. Moreover, the RLMFE-OCD approach signifies a maximal solution with $accur_y$ of 94.44%, $preci_n$ of 91.18%, $recal_l$ of 91.37%, and $F1_{score}$ of 91.25%.

Table 2 represents the ablation study of the proposed RLMFE-OCD methodology. The GoogLeNet + DQN, attains an $accu_y$ of 90.62%, $prec_n$ of 87.25%, $reca_l$ of 87.63%, and $F1_{score}$ of 87.44%. and the ResNet-50 + DQN providing an $accu_y$ of 91.37%, $prec_n$ of 88.41%, $reca_l$ of 88.72%, and $F1_{score}$ of 88.56%. Moreover, the SE-VGG16 + DQN, achieves a better $accu_y$ to 92.15%, $prec_n$ to 89.36%, $reca_l$ to 89.54%, and $F1_{score}$ to 89.45%. However, the GoogLeNet + ResNet-50 + DQN, further enhance the $accu_y$ to 93.08%, $prec_n$ to 89.92%, $reca_l$ to 90.18%, and $F1_{score}$ to 90.05%. Besides, the GoogLeNet + ResNet-50 + SE-VGG16 (without BOA) + DQN, get an improved $accu_y$ to 93.71%, $prec_n$ to 90.53%, $reca_l$ to 90.68%, and $F1_{score}$ to 90.68%. Lastly, the RLMFE-OCD (GoogLeNet + ResNet-50 +

Table 2. Ablation study of the proposed RLMFE-OCD methodology.

Model	$accu_y$	$prec_n$	$reca_l$	$F1_{score}$
GoogLeNet + DQN	90.62	87.25	87.63	87.44
ResNet-50 + DQN	91.37	88.41	88.72	88.56
SE-VGG16 + DQN	92.15	89.36	89.54	89.45
GoogLeNet + ResNet-50 + DQN	93.08	89.92	90.18	90.05
GoogLeNet + ResNet-50 + SE-VGG16 (without BOA) + DQN	93.71	90.53	90.68	90.60
RLMFE-OCD (GoogLeNet + ResNet-50 + SE-VGG16 + BOA + DQN)	94.44	94.44	91.37	91.25

International Journal of
Medical Toxicology & Forensic Medicine

SE-VGG16 + BOA + DQN) technique, got the highest outcome with an $accu_y$ of 94.44%, $prec_n$ of 94.44%, $reca_l$ of 91.37%, and $F1_{score}$ of 91.25%, shows the effectiveness of the components.

Conclusion

In this study, the RLMFE-OCD method is presented for timely detection of oral cancer (OC) using histopathological images. The framework integrated multiple DL architectures, including GoogLeNet, ResNet-50, and SE-VGG16, to capture diverse, high-level feature representations. The classification process is handled using a DQN method, which allows adaptive learning and improved decision-making in complex scenarios. Moreover, the BOA is used for parameter tuning, thereby enhancing model stability and performance. Experimental results show that the proposed RLMFE-OCD technique consistently surpasses existing models across multiple evaluation metrics. However, the model still has limitations regarding computational complexity owing to multiple feature extractors and reinforcement learning components. Future research can focus on reducing the computational complexity of the proposed model by introducing lightweight feature-extraction models. Additionally, the model could be extended to multi-class and multi-organ cancer detection scenarios to assess its generalization capability.

Ethical Approval

Not Applicable.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report there are no competing interests to declare.

References

[1] Panigrahi S, Swarnkar T. Machine learning techniques used for the histopathological image analysis of oral cancer: a review. Open Bioinform J. 2020;13(1):106-18. [DOI: 10.2174/1875036202013010106]

[2] Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. Sci Rep. 2019;9(1):6994. [DOI: 10.1038/s41598-019-43372-7]

[3] Appavu N. Oral cancer histopathological detection and diagnosis using hybrid AI deep learning methods. In: 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT). IEEE; 2025 Feb. p. 2061-7. [DOI: 10.1109/IDCIOT64235.2025.10915019]

[4] Gab Allah AM, Gaballa M, Elshennawy NM, Elkholy A. Edge-supervised convolutional neural network for histopathological classification of oral cancer images. Neural Comput Appl. 2025;1-20. [DOI: 10.1007/s00521-025-11457-2]

[5] Dhal P, Mishra D, Pradhan B. A deep ensemble-based framework for the prediction of oral cancer through histopathological images. Appl Soft Comput. 2025;113258. [DOI: 10.1016/j.asoc.2025.113258]

[6] Dharani R, Danesh K. Optimised deep learning ensemble for accurate oral cancer detection using CNNs and metaheuristic tuning. Intell Based Med. 2025;100258. [DOI: 10.1016/j.ibmed.2025.100258]

[7] Shukla R, Ajwani B, Sharma S, Das D. Identifying oral carcinoma from histopathological images using unsupervised nuclear segmentation. In: 2024 IEEE 9th International Conference for Convergence in Technology (I2CT). IEEE; 2024 Apr. p. 1-6. [DOI: 10.1109/I2CT61223.2024.10543340]

[8] Santhiya M, Sindhuja M, Jegatha R, Manikandan J. A practical automated framework for oral cancer detection by enhanced convolutional neural networks.

In: 2023 12th International Conference on Advanced Computing (ICoAC). IEEE; 2023 Aug. p. 1-7. [DOI: 10.1109/ICoAC59537.2023.10249983]

[9] Song S, Ren X, He J, Gao M, Wang JN, Wang B. An optimal hierarchical approach for oral cancer diagnosis using rough set theory and an amended version of the competitive search algorithm. Diagnostics. 2023;13(14):2454. [DOI: 10.3390/diagnostics13142454]

[10] Marzouk R, Alabdulkreem E, Dhahbi S, Nour MK, Duhayyim MA, Othman M, et al. Deep transfer learning driven oral cancer detection and classification model. Comput Mater Contin. 2022;73(2). [DOI: 10.32604/cmc.2022.029326]

[11] Marasović T, Papić V. Few-shot breast cancer diagnosis using a Siamese neural network framework and triplet-based loss. Algorithms. 2025;18(9):567. [DOI: 10.3390/a18090567]

[12] Li F, Weng C, Liang Y. SE-WiGR: a WiFi gesture recognition approach incorporating the squeeze-excitation mechanism and VGG16. Appl Sci. 2025;15(11):6346. [DOI: 10.3390/app15116346]

[13] Kurek P. Risk-aware deep Q-networks for volatility-driven execution in cryptocurrency markets [dissertation]. Warwick: Warwick Business School, University of Warwick; 2025. [Link]

[14] Ji Q, Li R, Jing S. Uncertainty-based design optimization framework based on an improved chicken swarm algorithm and Bayesian optimization neural network. Appl Sci. 2025;15(17):9671. [DOI: 10.3390/app15179671]

[15] Mendeley Data. Oral cancer histopathological dataset [Internet]. Mendeley Data; 2024. [Link]

[16] Liu P, Bagi K. A tailored deep learning approach for early detection of oral cancer using a 19-layer CNN on clinical lip and tongue images. Sci Rep. 2025;15(1):23851. [DOI: 10.1038/s41598-025-07957-9]

[17] Parola M, Galatolo FA, La Mantia G, Cimino MG, Campisi G, Di Fede O. Towards explainable oral cancer recognition: screening on imperfect images via informed deep learning and case-based reasoning. Comput Med Imaging Graph. 2024;117:102433. [DOI: 10.1016/j.compmedimag.2024.102433]