Research Paper

Deep Convolutional Multi-Feature Fusion Approach for Early Laryngeal Carcinoma Detection using Narrow-Band Medical Imaging

Annamalai Gopinath[1*], Bharathi Alagiri[2], Karthikeyan Kasi[3], Kayalvili Sundaramoorthy[4], Biju Theruvil Sayed[5], Vivek Marotrao Korde[6]

1. *Department of Electronics with Artificial Intelligence, DRBCCC Hindu College, Pattabiram, Chennai-600 072, Tamil Nadu, India*
2. *Department of Software Engineering, Accenture, No. 40, Chakrapani street, Mayor sa ganesan kudil, chennai – 600033, India*
3. *Department of Software Engineering, Tata Consultancy Services Pvt Ltd, No:12/6 Model Hutment Road Extn, CIT Nagar, Chennai – 35, India*
4. *Department of Artificial Intelligence and Data Science, Kongu Engineering College, Erode -638060, Tamil Nadu, India*
5. *Department of Computer Science, Dhofar University, PO Box 2509, PCode 211, Salalah, Oman*
6. *Department of Mechanical Engineering, Yeshwantrao Chavan College of Engineering (YCCE), Nagpur Hingna Road, Wanadongri, Nagpur, Maharashtra 441110, India*

Citation Gopinath A, Alagiri B, Kasi K, Kayalvili S, Theruvil Sayed B, Korde VM. Deep Convolutional Multi-Feature Fusion Approach for Early Laryngeal Carcinoma Detection using Narrow-Band Medical Imaging. International Journal of Medical Toxicology and Forensic Medicine. 2026; 16:E52406.

https://doi.org/10.22037/ijmtfm.v16.52406

Article info:
Received: 8 June, 2026
Accepted: 30 June, 2026
Published: 01 August, 2026

Keywords:
Laryngeal Carcinoma, Deep Feature Fusion, Medical Image Analysis, Artificial Intelligence, Computer-aided diagnosis

ABSTRACT

Background: Laryngeal carcinoma (LC) is one of the major cancers affecting the head and neck region, where delayed diagnosis can reduce survival rate and affect vocal function. Although medical image-based automated detection systems are useful and accurate, LC detection remains challenging due to variations in inter-class differences, unclear lesion boundaries, noise, and the distinct appearance of malignant tissues. The current state of the art in deep learning models relies on single-network feature extraction approaches, which restricts the ability to derive spatial and semantic representations from complex laryngeal images.

Methods: To address these issues, this paper presents an Artificial Intelligence-based Deep Convolutional Feature Fusion Model (AI-DCFFM) for Robust LC Detection using Medical Imaging. The presented method extracts deep features using MobileNetV2, DenseNet, and CapsNet architectures. Further, a parallel maximum covariance (PMC) based fusion strategy is introduced to combine the complementary feature representations obtained from the multiple DL models.

Results: For classification, a deep belief network (DBN) classifier is used to differentiate healthy and cancerous tissues. Additionally, the zebra optimization algorithm (ZOA) has been employed for optimal tuning of DBN hyperparameters to enhance the detection performance. The proposed model can be assessed by utilizing the benchmark laryngeal image dataset containing four tissue classes.

Conclusion: Experimental outcomes indicate that the proposed technique accomplishes boosted classification performance with an accuracy of 99.35%, outperforming existing methods under different evaluation metrics. The experimentation outcomes highlight the efficiency of the proposed method for reliable computer-aided LC diagnosis.

*** Corresponding Author:**
Annamalai Gopinath, PhD
Department of Electronics with Artificial Intelligence, DRBCCC Hindu College, Pattabiram, Chennai-600 072, Tamil Nadu, India.
E-mail: *gopinath@drbccchinducollege.edu.in*

Introduction

Laryngeal carcinoma (LC) poses a significant threat to public health in the modern era, impacting individuals globally. Imaging techniques like magnetic resonance imaging (MRI), ultrasound (US), and computed tomography (CT) are vital for staging, helping assess tumor size, cartilage involvement, local spread, and nearby lymph node metastasis [1]. Recently, leveraging endoscopic imaging systems has been the conventional approach for effectively examining individuals suspected of having LC at an initial stage in clinical settings. Current breakthroughs in Artificial Intelligence (AI) models, computer hardware, and extensive medical imaging datasets are empowering healthcare researchers and computer scientists to collaboratively enhance the dependability of cancer risk classification. These improvements signify a substantial step beyond traditional techniques, including Cox analyses, logistic regression, and multi-factor analysis. The development of Machine Learning (ML) has drawn rising attention, sustained by a rapid increase in evidence for its extensive applications under several cancers [2]. Deep learning (DL) algorithms have also developed highly popular presently due to their capability for achieving the highest prediction accuracy [3]. Despite the recent developments in AI and DL methods for LC detection, numerous challenges remain. Existing methods mainly depend on single DL architectures, which may not efficiently capture complementary feature representations from complex laryngeal tissue images.

Therefore, there is a need for an effective deep feature fusion framework that can improve the robustness and accuracy of LC detection using medical imaging data. Motivated by these challenges, this paper presents an AI-based Deep Convolutional Feature Fusion Model (AI-DCFFM) for Robust LC Detection using Medical Imaging. The major contributions of the proposed work are as follows:

- Carry out deep feature extraction using MobileNetV2, DenseNet, and CapsNet approaches to obtain discriminative feature representations from medical images.

- Introduce a parallel maximum covariance (PMC) based feature fusion strategy to combine complementary deep features for improved tissue characterization and classification performance.

Utilize Zebra Optimization Algorithm (ZOA) with a deep belief network (DBN) classifier for effective recognition of healthy and cancerous laryngeal tissues.

Related Studies on LC Detection

Nie et al. [4] presented an enhanced YOLOv8 architecture termed MSEC-YOLO, particularly developed for detecting and classifying LC from endoscopic imaging. A new multi-scale improved convolution model is presented to enhance the system's feature extraction ability. Al Khulayf et al. [5] presented the fusion of efficient TL methods alongside Pelican Optimizer for precise LC classification. The authors [6] established a timely LC recognition system leveraging the hybridization of deep and manual features. This process leads to more informative features crucial for the prompt recognition of LC, contributing to a reduction in mortality rates. The authors [7] presented an automated LC Detection through the dandelion optimizer algorithm with the ensemble learning methodology on biomedical throat region images. Joseph and Vidyarthi [8] proposed a two-fold model for the timely detection of Laryngeal Squamous Cell Carcinoma (LSCC) through a lightweight Deep CNN and statistical features. The visual examination presents the major difficulty because it strongly depends on the clinical experience and knowledgeable doctors. Naga Padmavathi et al. [9] have developed a novel Optimization-Based Cascaded DNN (OCDNN), which differentiates between healthy and unhealthy tissues. Furthermore, offers cascaded patch-based CNN (CP-CNN) image segmentation. Luo et al. [10] presented end-to-end networks (DCA-DAFFNets) alongside the deep adaptive feature fusion (DAFF) and deformable convolution guided attention blocks (DCABs), to enhance the grading capability. Bu et al. [11] presented the multi-modal internal-based feature representation learning (MIFRL) approach by detecting multi-attributes and creating endoscopic procedure documentation.

Materials and Methods

In this work, an automated LC detection model named AI-DCFFM is developed using medical images. Figure 1 demonstrates the workflow of the presented AI-DCFFM approach.

Dataset Descriptions

In this study, a publicly available laryngeal image dataset from the Kaggle platform was utilized [12]. The dataset was divided into 80% training data and 20% testing data for evaluation. The experiments were conducted using Python and TensorFlow/Keras with GPU support. The dataset contains 1320 image patches of healthy and early-stage cancerous laryngeal tissues,

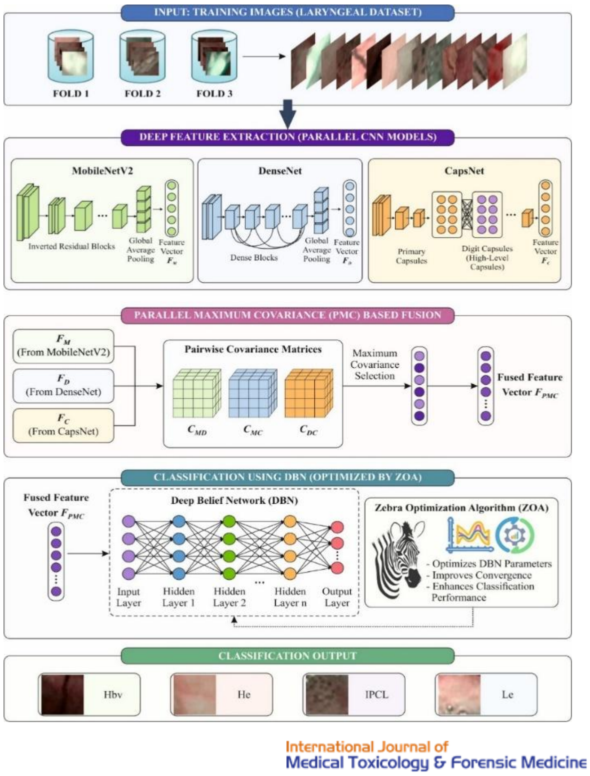

Figure 1. Complete Process of the proposed AI-DCFFM Model.

where each image patch has a size of 100×100 pixels, collected from narrow-band laryngoscopic images. The dataset consists of four tissue classes, such as Hypertrophic Blood Vessels (HBV), Healthy Tissue (HE), Abnormal IPCL-like Vessel (IPCL), and Leukoplakia (LE), with 330 images in each class. The balanced distribution of images helps in effective training and evaluation of the classification model. Figure 2 indicates a sample set of images.

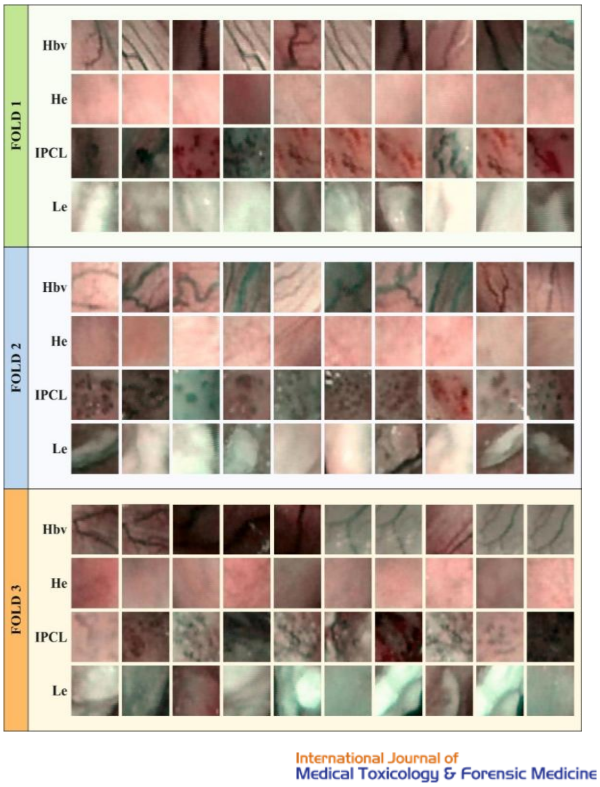

Figure 2. Sample Images of the Laryngeal dataset.

Deep Convolutional Neural Network Architectures

The initial point-wise convolution in the bottleneck component of MobileNet-V2 leads to developing the feature mapping channel. The second point-wise convolution in the bottleneck component of MobileNetV2 helps in improving feature extraction efficiency with reduced computational complexity. MobileNet-V2 includes stacking bottleneck modules. All bottleneck components are the inverse residual framework that acknowledges a depth-wise separable convolution, for example, a point-wise convolution accompanied by a depth-wise separable convolution.

$$Y = X * f + b \quad (1)$$

The parameter complexities of the depth-wise separable convolution in MobileNet-V2 are $(1/N + 1/K^2)$ times that of normal convolutions. DenseNet is a DL technique employed for identifying images, with its core aspects being the densely connected framework. Each layer receives inputs from each preceding layer, which improves feature recycling and gradient flow, consequently improving the reutilization of aspects. The dense connection is specified.

$$H_l = H_{(l-1)} \oplus H_{(l-2)} \ldots \oplus H_0 \quad (2)$$

Now $[x_0, x_1, \ldots, x_{(l-1)}]$ refers to the concatenation of outputs from each preceding layer, and H_l signifies the transformation function of $[\![the\ l]\!]\_th$ layer. Compared to conventional CNN models, CapsNet captures spatial relationships between image features more effectively. The input and output of CapsNet are represented as vectors. CNN separates the iterative dynamic weights and routing methods by selecting the capsule's value.

$$\hat{u}_{(j/i)} = w_{ij}\, u_i \qquad (3)$$

The prediction vector $\hat{u}_{(j/i)}$ It is utilized to determine the similarity between the input vector and higher-level capsule representation during the routing process.

Feature fusion offers a richer feature area for recognition applications. The incorporation of multiple deep CNN feature mapping outcomes is performed in the proposed feature space. Assume the deeper CNN attributes removed from the 2 deeper CNN methods represent φ^{m1} and φ^{m2}. The features extracted utilizing the pretrained MobileNetV2, DenseNet, and

CapsNet models are represented as feature vectors with different dimensions. The extracted feature spaces are balanced through average padding before the fusion process to obtain compatible feature representations. The maximal covariance of φ_1 and φ_2 can be defined as follows.

$$\breve{f} = Cov[z\_(1^{\wedge\prime})\; z\_2\,] \qquad (4)$$

$$F\varphi_1\varphi_2 = \frac{1}{n-1}\left(\varphi^{(m1)}\varphi^{(m2)T}\right) \qquad (5)$$

In the proposed method, the feature vectors extracted from MobileNetV2, DenseNet, and CapsNet are fused through the PMC-based fusion strategy to attain a more discriminative feature representation. The covariance-based fusion process assists in preserving complementary information from different DL approaches while reducing redundant representations [13]. The fused feature vector is then passed to the DBN classifier for LC detection.

LC Classification using DBN

For the LC detection model, the DBN classifier is used. The DBN was selected because of its ability to learn hierarchical probabilistic feature representations efficiently with limited dataset sizes, thereby minimizing overfitting risks compared to deeper fully connected architectures. The fused feature vectors obtained from the PMC fusion model are provided as input to the DBN classifier for LC detection. The DBN offers multi-level hierarchical features utilizing numerous Restricted Boltzmann Machines (RBM) [14].

$$E(v,h) = -\sum\_(i=1)^{\wedge}v ▒ v\_i\;\; b\_i - \sum\_(j=1)^{\wedge}\mathcal{H} ▒ h\_j\;\; c\_j - \sum\_(i=1j)^{\wedge}v ▒ \sum\_(=1)^{\wedge}\mathcal{H} ▒ v\_i\;\; h\_j\, W\_ij \qquad (6)$$

RBM calculates the energy for all configurations of the HLs and input (visible) layer utilizing formulations. Now, v_i and h_j represent binary states of the HL and input layer, b_i and h_j represent bias values of the HL and input layer, $\mathcal{V}$ specifies the input layer counts, $\mathcal{H}$ signifies the HL counts, and W_ij symbolizes weights connecting the input layer and HL. This model assigns probability. $p(v)$ to the input vector v utilizing the expression below. Figure 3 demonstrates the architecture of DBN.

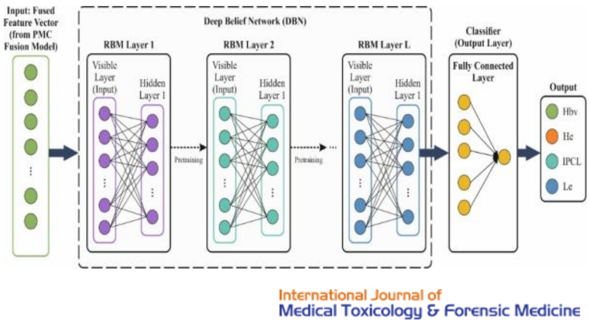

Figure 3. Architecture of the proposed DBN.

Model Optimization

Finally, the ZOA is applied for optimal fine-tuning of the DBN parameters to improve predictive performance. ZOA is a new bio-inspired metaheuristic model mainly inspired by the zebra behaviour [15]. It simulates their searching habits and predator-avoidance strategy. The basic principle of ZOA is the replication of social dynamics monitored in wild zebra herds. The ZOA model upgrades its members by integrating the dual natural behaviours of zebra such as foraging and defence against predators. In every iteration, the population of ZOA is upgraded in a dual-stage manner depending on these behaviours. In DL approaches, hyperparameters play a significant role in controlling the classification performance. Proper tuning of the hyperparameters assists in enhancing the generality of the model and reducing overfitting issues. Therefore, in this work, ZOA is used to optimally tune the DBN classifier parameters for improving LC detection performance.

Fitness=Max (Accuracy) (7)

In the proposed model, ZOA is used to determine the optimal hyperparameter values of the DBN classifier by maximizing the classification accuracy. The optimization process iteratively updates the search agents to obtain the best fitness value with reduced classification error.

Results and Discussions

This section offers a detailed result analysis of the model, including training configuration, performance analysis, comparative evaluation, and ablation study. Figure 4 exhibits the confusion matrices produced by MobileNetV2, DenseNet, CapsNet and the AI-DCFFM technique for the classifier solutions. The diagonal components in every matrix indicate the correctly categorized instances, although the non-diagonal elements denote misclassified samples. Fig. 4a illustrate the MobileNetV2 model performed moderately less effectively than the alternate approaches. Fig. 4d presents that the AI-DCFFM technique can be realized the maximum number of predictions with some misclassification errors throughout every class. Such outcomes confirm the efficiency of the proposed AI-DCFFM system for

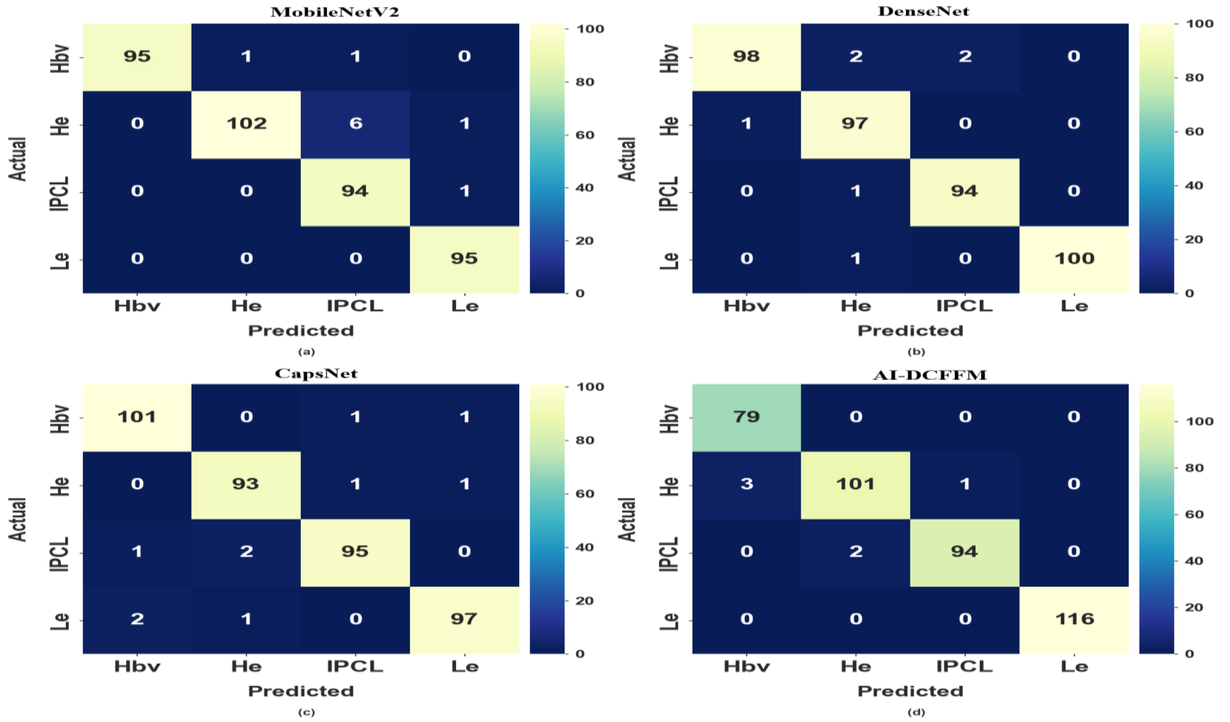

Figure 4. Confusion matrices with (a) MobileNetV2, (b) DenseNet, (c) CapsNet and (d) AI-DCFFM methods.

Table 1. Laser Photobiomodulation Dosimetric and Calibration Parameters.

Methods	$Accur_y$	$Preci_n$	$Recal_l$	$F1_{score}$	Computational Complexity (sec)
Compared with ML Model					
Linear Discriminant Analysis	90.90	89.81	98.79	92.06	7.15
Support Vector Machines	94.46	93.27	97.75	97.52	5.64
KNN Algorithm	94.85	93.17	87.04	87.13	5.90
Decision Tree	89.73	87.74	94.90	90.41	5.64
Naïve Bayes	94.97	90.90	97.63	87.56	6.27
Compared with DL Model					
Densenet201 Model	91.07	85.64	96.15	88.20	6.01
Alexnet Method	90.97	91.72	95.37	94.55	6.76
Inception v3 Model	89.43	92.72	88.97	93.18	5.70
Mobile Neural Architecture Search Network	88.72	92.97	95.08	98.22	6.98
Mobilenet v3 Algorithm	85.58	96.08	91.84	91.73	5.46
Compared with Proposed Model					
SqueezeNet+DenseNet+ResNet +VGG	97.32	93.29	91.75	97.31	5.50
Novel Feature Fusion Network	96.49	81.33	94.40	95.22	3.57
ResNet-152+MobileNetV3	84.19	98.42	82.82	88.56	4.87
CNN—BiT	87.07	85.06	86.81	82.30	6.13
MobileNetV2+ResNeXt-50	86.17	92.61	83.27	84.00	5.21
AI-DCFFM	**99.35**	**98.70**	**98.70**	**98.70**	**2.36**

enriching classifier performance.

Table 1 shows the comparative results of the AI-DCFFM approaches with other methodologies using evaluation metrics [16, 17]. The stimulation outcome indicated that the AI-DCFFM method achieves better performance.

The DL models have attained lower performance, while the ML models have slightly higher performance.

In addition, the SqueezeNet+DenseNet+ResNet +VGG, Novel Feature Fusion Network, ResNet-152+MobileNetV3, CNN-BiT, and MobileNetV2+ResNeXt-50 approaches have performed moderate outcomes. At the same time, the proposed AI-DCFFM method has achieved a greater outcome with *acc* 〖*ur*〗_*y* of 99.35%, *pre* 〖*ci*〗_*n* of 98.70%, *rec* 〖*al*〗_*l* of 98.70%, and *F_score* of 98.70% long with a lower computational complexity of 2.36. Therefore, the AI-DCFFM model demonstrates better overall effectiveness compared to state-of-the-art approaches.

Conclusion

In this paper, we have designed and developed the AI-DCFFM algorithm for reliable computer-aided LC diagnosis in a healthcare environment. Primarily, the image pre-processing is implemented using an AMF to remove noise while maintaining crucial structural details. Following, a PMC fusion technique is used to incorporate complementary features removed from MobileNetV2, DenseNet, and CapsNet frameworks to successfully differentiate between healthy and cancerous tissues. For the LC detection model, the DBN classifier is used. Finally, the ZOA is applied for optimal fine-tuning of the DBN parameters to improve predictive performance. The empirical result analysis of the AI-DCFFM system occurs against the benchmark laryngeal dataset, and the comparative outcomes establish the enhanced performance over the recent techniques under different measures.

Future research should focus on enhancing the robustness and generalizability of the presented architecture by validating large-scale, multi-institutional clinical datasets. External clinical validation would help assure the model's reliability over various imaging conditions and patient populations. Additionally, it might discover the improvement of computationally efficient or lightweight frameworks to decrease processing time and allow real-time clinical deployment. Expanding the architecture to comprise a broader range of laryngeal disease classes and integrating multi-modal clinical data could further improve its diagnostic performance and practical applicability in real-world healthcare environments.

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] Sahoo PK, Mishra S, Panigrahi R, Bhoi AK, Barsocchi P. An improvised deep-learning-based mask R-CNN model for laryngeal cancer detection using CT images. Sensors. 2022;22(22):8834. [DOI: 10.3390/s22228834]

[2] Wellenstein DJ, Woodburn J, Marres HA, van den Broek GB. Detection of laryngeal carcinoma during endoscopy using artificial intelligence. Head Neck. 2023;45(9):2217-26. [DOI: 10.1002/hed.27441]

[3] Lee JC. Application of Artificial Intelligence (AI) in Laryngeal Cancer. J Clin Otolaryngol Head Neck Surg. 2023;34(2):27-35. [DOI: 10.35420/jcohns.2023.34.2.27]

[4] Nie X, Zhang X, Wang D, Liu Y, Xing L, Liu W. Laryngeal cancer diagnosis based on an improved YOLOv8 algorithm. Mach Learn Sci Technol. 2025;6(1):015011. [DOI: 10.1088/2632-2153/ada2d9]

[5] Al Khulayf AMF, Alamgeer M, Obayya M, Alqahtani M, Ebad SA, Aljehane NO, et al. Mathematical fusion of efficient transfer learning with a fine-tuning model for laryngeal cancer detection and classification in the throat region using biomedical images. Alex Eng J. 2025;129:658-671. [DOI: 10.1016/j.aej.2025.06.044]

[6] Joseph JS, Vidyarthi A, Singh VP. An improved approach for initial stage detection of laryngeal cancer using effective hybrid features and an ensemble learning method. Multimedia Tools Appl. 2024;83(6):17897-919. [DOI: 10.1007/s11042-023-16077-3]

[7] Alzakari SA, Maashi M, Alahmari S, Arasi MA, Alharbi AA, Sayed A. Towards laryngeal cancer diagnosis using Dandelion Optimizer Algorithm with ensemble learning on biomedical throat region images. Sci Rep. 2024;14(1):19713. [DOI: 10.1007/s11042-023-16077-3]

[8] Joseph JS, Vidyarthi A. Dual deep learning and feature-based models for classification of laryngeal squamous cell carcinoma using narrow-

band imaging. Trait Signal. 2024;41(1):237. [DOI: 10.18280/ts.410119]

[9] Naga Padmavathi P, Sree Devi M, Maheswara Rao VVR, Gopala Gupta AS. Construction of Cascaded Deep Neural Network with Optimization-Based Feature Selection in CT Images for Detecting Laryngeal Cancer. In: Innovative Product Design and Intelligent Manufacturing Systems. Singapore: Springer Nature Singapore; 2023. p. 459-72. [DOI: 10.18280/ts.410119]

[10] Luo J, Huang P, He P, Wei B, Guo X, Xiao H, et al. DCA-DAFFNet: An end-to-end network with deformable fusion attention and deep adaptive feature fusion for laryngeal tumor grading from histopathology images. IEEE Trans Instrum Meas. 2023;72:1-15. [DOI: 10.1109/TIM.2023.3328088]

[11] Bu Z, Zheng Y, Zheng Y, Yang L, Liao K, Liu T, et al. MIFRL: A Model for Automatic Diagnosis of Laryngeal Diseases Based on Multi-Modal Information Fusion. In: 2023 8th International Conference on Communication, Image and Signal Processing (CCISP). IEEE; 2023. p. 148-52. [DOI: 10.1109/CCISP59915.2023.10355778]

[12] Hajian M. Laryngeal Dataset [Internet]. Kaggle; 2023 [cited 2026 Aug 1]. [Link]

[13] Mohammad F, Al-Razgan M. Deep feature fusion and optimization-based approach for stomach disease classification. Sensors. 2022;22(7):2801. [DOI: 10.3390/s22072801]

[14] Andhare MS, Vijayan T, Karthik B, Urooj S. A Novel Optimized Hybrid Deep Learning Framework for Mental Stress Detection Using Electroencephalography. Brain Sci. 2025;15(8):835. [DOI: 10.3390/brainsci15080835]

[15] Ramadhani A, Robandi I, Djalal MR, Prakasa MA. Improving the Energy Management System with the Interval Type-2 Fuzzy Inference System–Zebra Optimization Algorithm (IT2FIS-ZOA) for Predicting the Load Consumption of Healthcare Facilities in the National Holiday Season. Eng Technol Appl Sci Res. 2025;15(4):25878-88. [DOI: 10.48084/etasr.12043]

[16] Xu ZH, Fan DG, Huang JQ, Wang JW, Wang Y, Li YZ. Computer-aided diagnosis of laryngeal cancer based on deep learning with laryngoscopic images. Diagnostics. 2023;13(24):3669. [DOI: 10.3390/diagnostics13243669]

[17] Ramamurthy K, George TT, Shah Y, Sasidhar P. A novel multi-feature fusion method for classification of gastrointestinal diseases using endoscopy images. Diagnostics. 2022;12(10):2316. [DOI: 10.3390/diagnostics12102316]