Acta Polytechnica **66**(1):98–108, 2026
https://doi.org/10.14311/AP.2026.66.0098

Published by the Czech Technical University in Prague

OPTIMISED DEEP LEARNING FOR ORAL CANCER CLASSIFICATION

Chellasamy Sulochana*, Mahadevan Sumathi

Madurai Kamaraj University, Sri Meenakshi Government Arts College for Women(A), Department of Computer Science, 625002 Madurai, Tamilnadu, India

* corresponding author: sulochanaravisankar@gmail.com

Abstract. Oral cancer detection is essential, especially in areas with high occurrence rates, is essential for better early diagnosis and individualised treatment plans. SV-OnionNet is a deep learning framework presented in this article that aims to improve the classification accuracy of oral cancer diagnosis. While maintaining important structural details, the method lowers noise in medical pictures by integrating an adaptive Non-Linear Means (NLM) filter. Spatial features are improved by the Label-Guided Attention (LGA) module, which guarantees constant labelling and improves feature extraction. By enabling accurate pixel-level segmentation of lesions, Seg-UNet provides increased classification reliability. The Support Vector Machines (SVM) deep learning classification model used in the SV-OnionNet architecture preserves spatial relationships for improved feature learning, replacing traditional fully linked layers (LKN). The Competitive Search Optimization (CSO) algorithm fine-tunes model parameters, therefore optimising feature selection and classification. The evaluation on the Mouth and Oral Diseases dataset demonstrated exceptional accuracy, precision, recall, and specificity, with the proposed classification achieving a 99.94 % accuracy. These findings emphasise the effectiveness of SV-OnionNet in improving the diagnostic accuracy and reliability. The study highlights the potential of integrating deep learning techniques with optimisation strategies to advance oral cancer detection. Future research will focus on expanding datasets and exploring additional optimisation methods to further improve the classification performance.

Keywords: Oral cavity cancer, adaptive Non-Linear Means (NLM) filter, Seg-UNet model, SV-OnionNet, Competitive Search Optimizer (CSO).

1. Introduction

In order to reduce image noise, the conventional NLM method grants each pixel a search window of fixed size; however, as the noise variance increases, its effectiveness decreases. Finding the original pixel values in smooth regions of the image requires a bigger navigation window. Conversely, smaller navigation windows are necessary for more intricate areas in order to properly maintain textures and edges [1]. An adaptive NLM technique was created by Siegel et al. [1] that allows changing the size of the search window to be changed based on the grey level difference (GLD) of each pixel. We demonstrate that by adjusting the window size to each pixel's environment, this adaptive NLM technique can maintain crucial edge information using standard optical pictures. With the development of digital imaging technologies, medical practitioners can now identify cancer with X-rays, CT scans, and MRI scans [2]. Researchers are using these photos to create automated cancer diagnosis tools. In cancer imaging, deep learning and machine learning are frequently utilised. Machine learning techniques are used to train algorithms using characteristics of cancer cells, such as tissue texture, colour, and form [3].

An individual's life expectancy is seriously threatened by oral cancer, a common malignancy that affects the mouth, neck, and tongue. Approximately 90 % of instances of oral cancer are of the primary subtype, oral squamous cell carcinoma (OSCC) [4]. One OSCC development is the lack of particular clinical vital signs. Physicians have a hard time making accurate predictions about the condition. Specialists in otolaryngology and dentistry are responsible for diagnosing and treating this condition, which is regarded as the sixth most prevalent form of cancer globally [5]. In today's medical community, tackling cancer has become one of the foremost challenges. Despite recent advances in medical research, no definitive cure for cancer has yet been found. Recognising the symptoms of oral cancer can be particularly difficult, especially when the cancer occurs in the throat [6, 7].

A wider range of illnesses are referred to as oral cancer than mouth cancer [8]. Although it only affects the mouth cavity, oral cancer can spread to the tonsils, lips, and throat. On the one hand, the development of red or white patches and chronic ulcers are typically indicators of oral cancer [9]. On the other hand, oral cancer is also associated with symptoms such as swollen lymph nodes, difficulty swallowing, and changes in speech.

In this section, we review the latest developments and research in the detection of oral cancer using machine learning and deep learning methods [10, 11]. It details a strategy that employs imagery data from

oral areas. The authors applied a sophisticated contour box technique to accurately define regions of interest in oral images. Typically, the region is created as a contour box located in the centre of the oral canal. Using comparable measurement methods, multiple contour regions of different orientations and sizes can be distinguished from their associated images [12].

This chapter offers a thorough analysis of significant research in the field, emphasising the challenges and developments in the diagnosis of OSCC [13, 14]. As this study shows, in order to obtain high diagnostic accuracy, researchers have employed a variety of strategies. Pascal et al. [15] employed random forest (RF) and convolutional neural network (CNN) algorithms to identify keratin granules in pictures of oral diseases.

The study [16–18] examined the role of artificial intelligence in cancer detection and prevention. Dentists must diagnose oral disorders early in order to properly monitor and treat them, particularly oral cancer. The suggested approach was tested on 1 662 samples in order to accomplish this [19]. About 74 % of diseases can be successfully detected and diagnosed by AI with 99 % sensitivity, 80 % accuracy, and 0.99 % confidence, according to a study of the technology's efficacy [20].

2. Materials and methods

A sequential pipeline that combines pre-processing, feature extraction, segmentation, classification, and optimisation into a single workflow is the basis of the suggested framework. Adaptive Non-Linear Means (NLM) pre-processing is used to minimise noise in input medical images while maintaining structural features. Then, the Label-Guided Attention (LGA) module improves spatial feature representations while guaranteeing pixel-by-pixel label consistency. Following the refinement, these attributes are sent into the Seg-UNet model, which precisely segments lesions and draws the borders of malignant areas. To increase discrimination and retain spatial linkages, the SV-OnionNet design uses Support Vector Machines in place of traditional fully connected layers to classify the segmented outputs. Lastly, the network parameters are adjusted by the Competitive Search Optimiser (CSO) to reduce error and increase the classification accuracy. Through this gradual integration, the framework can go methodically from unprocessed input photos to accurate oral cancer diagnostic predictions.

2.1. Pre-processing – Non-Linear Mean filter

The UI's NLM adaptive filtering approach is a simple process. First, a noise image is acquired using an ultrasound imaging system. The first step is to apply basic NLM techniques to reduce noise. The traditional NLM noise removal method is based on the following equation:

$$I(i) = \sum_{j \in N_i} w_{ij} I(j), \quad \sum_j w(i,j) = 1, \quad 0 \le w(i,j) \le 1. \tag{1}$$

In this case, $I(i)$ represents the intensity level of pixels at position i, and N_i represents the search window value associated with that pixel. The weight value between pixels i and j is determined by (j), which represents the intensity of pixel j. This weight reflects the similarity between pixels i and j and is defined by the following equation:

$$w_{ij} = \frac{1}{z(i)} \exp\left(-\frac{\|p(i) - p(j)\|_2}{h^2}\right), \quad Z(i) = \sum_j \exp\left(-\frac{\|p(i) - p(j)\|_2}{h^2}\right). \tag{2}$$

The region inside the navigation window is indicated by $P(i)$ and $P(j)$, and both windows are of the same size. Two square arrays $P(i)$ and $P(j)$, each centred at pixels i and j, are compared for similarity using the normalising constant $Z(i)$. The filter parameter is h. There are three distances between (i) and (j), and the Gaussian kernel size is set at 11. Euclidean distance is used to set the weight distribution, which is intended to expand around pixels with comparable patterns.

Once a smooth result has been obtained by applying the median filter to the underlying NLM result, GLD(dX) is computed by calculating the absolute error value of each pixel. A 3×3 size is used in the experiment for the median filter. The mean value (μ) of GLD is used to set the first lower bound, and the weight (α) is subtracted from the standard deviation (σ) of GLD in order to obtain the second lower bound. The α value of the study is typically 0.5. The ideal search window (s_i^{opt}) was determined using the following equation:

$$s_i^{\text{opt}} = \begin{cases} c_1 & dX_i < T_1 \\ c_2 & T_1 \le dX_i < T_2 \\ c_1 & dX_i < T_1. \end{cases} \tag{3}$$

Users have the option to modify these constants according to the image's specific features and noise level. Finally, for each pixel, we apply the optimal search window determined by the standard NLM algorithm to the value of h to achieve maximum noise reduction in the noisy image.

2.2. Feature extraction – the Label-Guided Attention (LGA)

Let $Z \in R^{H \times W \times D}$ be the features extracted from the intermediate complex mass. Equation (4) uses the following formula to define the linear projection

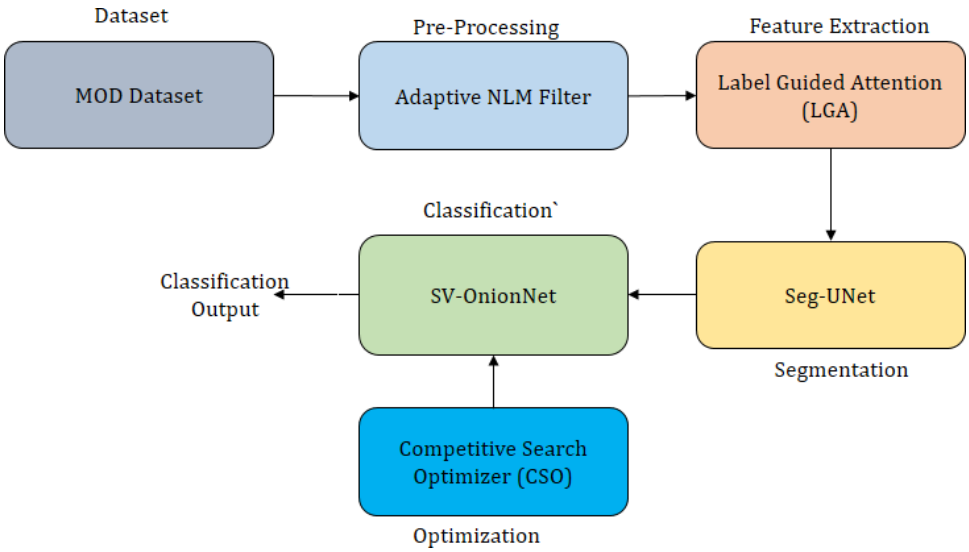


Figure 1. Overall proposed architecture.

of a feature over the median space along the weight matrix:

$$Z' = Z * W_{\text{proj}}, \tag{4}$$

where Z is the spatial feature map of the third convolution layer, $Z' \in R^{H\times W\times D'}$ is the linearly projected spatial feature and $W_{\text{proj}} \in R^{1\times 1\times D\times D'}$ is the learnable weight matrix. The distance between $Z' \in R^{H\times W\times D'}$ and its neighbour for each pixel (i, j) is calculated as follows:

$$d_{i,j} = \|Z'_{I,j} - Z'_{N_{i,j}}\|_2, \tag{5}$$

where $N_{i,j}$ is the set of neighbouring pixels, $\|\cdot\|_2$ is the L2 norm, $N = 8$, and $D \in R^{H\times W\times D}$ is the distance matrix. Using a Gaussian linear error unit (GeLU), the multi-layer perceptron (MLP) modifies the distance (d) in Equation (6) by applying two linear projections:

$$d' = W_1 \cdot \text{GeLU}(W \cdot d), \tag{6}$$

where $W_1 \in R^{1\times 1\times K'\times N}$ and $W_1 \in R^{1\times 1\times N\times K'}$ are the weight matrices and K' is the mean dimension. Using Equation (7), the feature affinity (FA) between neighbouring pixels (i, j) is computed:

$$\text{FA}_{i,j} = \exp(-\alpha.d'_{i.j}), \tag{7}$$

where the parameter that needs to be learned is denoted by α. Equation (8) illustrates how a softmax function applied to the adjacent dimensions normalises the FA:

$$\text{NFA}_{i.j} = \frac{\exp(-\alpha \cdot dd'_{i,j})}{\sum_{k\in N_i}(-\alpha \cdot d'_{i,k})}. \tag{8}$$

The normalised feature proximity of pixel (i, j) is $\text{NFA}_{i,j}$, and $i \in \{1, 2, 3, \ldots, N\}$. Following the determination of normalised AF, the estimated Z' features are used as the starting value, skip connections are applied, ReLU is enabled, and BN is used to reconstruct the features in accordance with:

$$Z''_{i,j} = \sum_k A_{i,j} \cdot Zd'_{i,k}, \tag{9}$$

$$F_{\text{out}} = \text{ReLU}(\text{BN}(Z'' + Z)), \tag{10}$$

where F_{out} is the feature map obtained by LGA, $Z''_{i,j}$ are the reconstructed features of pixel i and its surroundings j. Figure 2 displays the block structure of LGA, where the features taken from the third convolution block (Z) are normalised in relation to Z. The distance (d) between Z' and nearby pixels is then determined, and the Multi-Layer Perceptron (MLP) uses GeLU to apply bilinear projection, converting the distance d to d'. In addition, we compute the feature similarity (FA) between adjacent pixels (i, j) and use a softmax function to normalise it to an NFA. Last but not least, we create the output features (F_{out}) using the predicted features Z', skip connections, BN (Batch Normalization), and ReLU enable functions.

To calculate the distance, we begin with Conv3 and employ the Label Guided Attention (LGA) procedure. following linear projection of SF3 and its surrounding elements using Conv2D with a 1×1 kernel size. Additionally, in order to capture the local tissue pattern and change of the cancer site, we transformed this distance into a Multilayer Perceptron (MLP). Following Conv3, Equation (11) is used to compute the inference on pixel labels using LGA. According to Equation (12), the fine spatial features are first sent to a Conv4 block with a filter size of 256 before being sent to the second LGA block. The architecture of LGA can be seen in Figure 2.

$$\text{SF}_4 = \text{LGA}(\text{SF}_3), \tag{11}$$

$$\text{SF}_5 = \text{ReLU}(\text{BN}(\text{SF}_4 * W_5)). \tag{12}$$

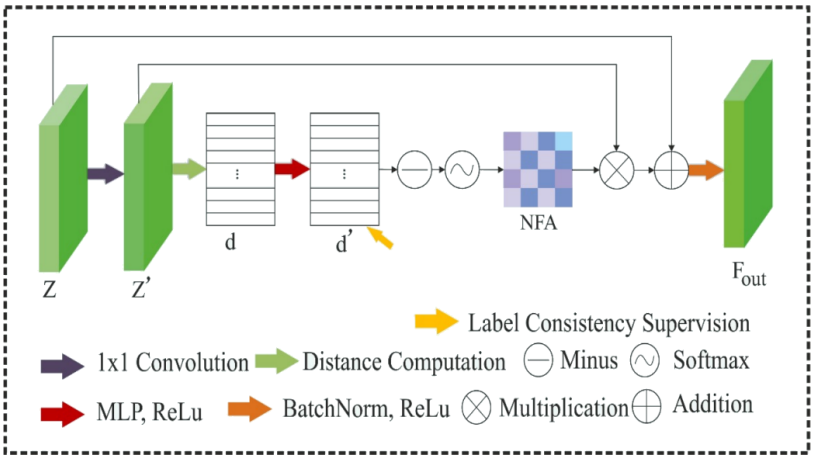


FIGURE 2. Architecture of LGA block.

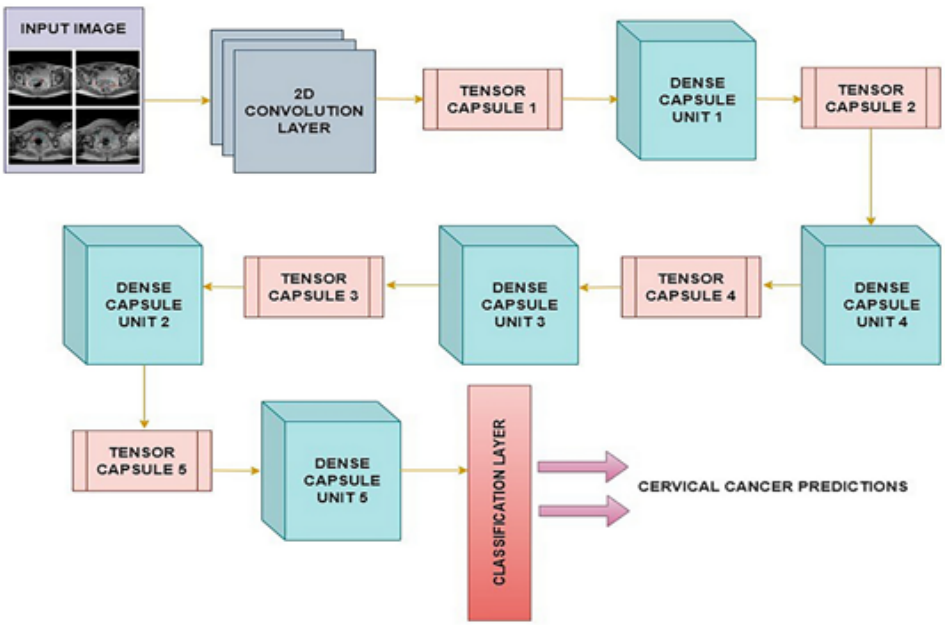


FIGURE 3. Seg-UNET architecture.

2.3. SEGMENTATION-UNET MODEL

During the Segmentation phase, images that have been enhanced with contrast are processed. In this stage, the Seg-UNet model is employed to identify and isolate cancerous regions within the oral cavity. This model is an integration of the SegNet and U-Net architectures, both of which are widely used for tasks involving visual segmentation. The SegNet model consists of an encoder that reduces the sample size and a decoder function that increases the sample size. Although the synthesis layers are added to each kernel block, the encoding and decoding mechanisms in this model are identical to the 13 synthesis layers in the VGG16 model. To match the 4D feature map domain in the synthetic stack, late links are introduced into the initial kernel architecture of the SegNet model, inspired by U-Net. The Seg-UNet architecture is shown in Figure 3.

The Seg-UNet model, an advanced derivative of the U-Net architecture, has been effectively employed for segmenting oral cancer lesions, particularly within histopathological images. By using deep encoders such as ResNet50, it adeptly captures the intricate structures of tissues, thereby improving the accuracy of detecting and classifying oral squamous cell carcinoma. In research conducted by Piyarathne and Liyanage [20] the model underwent training and evaluation using datasets that included both images and the ORCA dataset, highlighting its proficiency in addressing the segmentation challenge.

2.4. CLASSIFICATION SV-ONIONNET

In a typical CNN, neurons process input values without maintaining spatial connections with neighbouring neurons within the core. This reduces the robustness of input transitions. While max pooling helps to ensure robustness, it may not capture important spatial details among features. In this paper, we propose SV-OnionNet, a deep learning framework that can completely replace conventional deep learning models

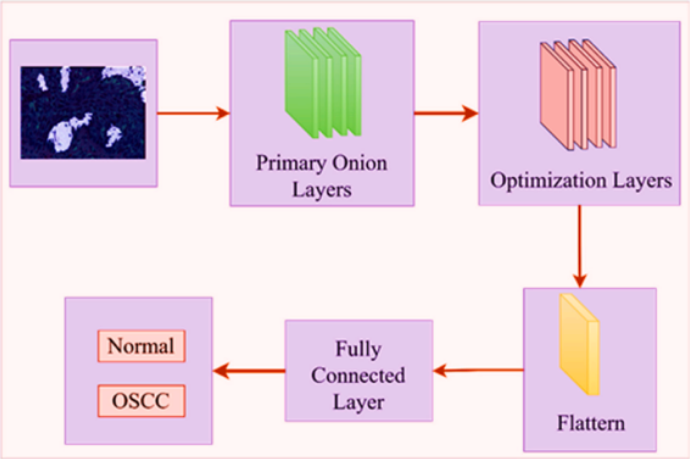


Figure 4. SV-OnionNet Framework.

by merging the existing SVM. By incorporating spatial relationships into its neural design, SV-OnionNet improbes the understanding of features and strengthens the robustness of oral cancer classification. The network design shown in Figure 4 highlights the key features of SVM and improves its discrimination capabilities and generalisation. SV-OnionNet aims to utilise the complex decision boundaries and spatial information processing capabilities of SVMs to improve the accuracy and reliability of clinical detection methods for oral cancer. The network consists of an input layer, main onion layer, enhancement layer, and final onion layer.

2.4.1. Input layer

Segmented images are used as input to train the proprietary deep learning network SV-OnionNet. It is developed to perform the classification using the extracted features.

2.4.2. Primary onion layer

The first layer is connected to the onion base consisting of 256 convolutional filters, each with a kernel size of 6, and applies convolution operations to extract deep feature maps from the input. This process is formally represented by:

$$m(i) = (h * f)(i) = \int h(j) f(i-j)\, dj. \qquad (13)$$

This layer uses zero padding to preserve the original dimensions of the input features. ReLU acts as a nonlinear activation function, as shown in Equation (14), where positive values are passed unchanged and negative values are set to 0:

$$o(i) = \max(0, i) = \begin{cases} i & i \geq 0 \\ 0 & i < 0. \end{cases} \qquad (14)$$

SV-OnionNet removes the max pool layer. In traditional CNNs, this layer prevents data protection, such as preserving unique features and spatial information of the input. This is the biggest limitation of standard CNN methods. Figure 4 presents the SV-OnionNet framework, which is a deep learning model designed for oral cancer classification.

2.5. Optimisation – the Competitive Search Optimizer (CSO)

The primary objective of optimisation in this study is to improve the classification of oral cancer images by maximising diagnostic accuracy while minimising error. Formally, the optimisation task is defined as:

$$\max_{X \in R^d} F(X) = \text{Accuracy}(X) - \lambda \cdot \text{Error}(X), \qquad (15)$$

where $X = (x_1, x_2, \ldots, x_d)$ is a d-dimensional candidate solution representing model parameters, $F(X)$ is the objective (fitness) function, Accuracy(X) measures the classification performance, Error(X) penalises misclassifications, and λ is a weighting factor to balance the two terms.

The Competitive Search Optimizer (CSO) is a metaheuristic algorithm inspired by the dynamics of competitions, where participants are ranked by performance and progressively refined across successive rounds. While the metaphor provides intuitive understanding, the mechanism can be rigorously modelled as an optimisation process.

In this framework, each participant corresponds to a candidate solution, the evaluation criteria correspond to a fitness function, and grouping participants into "excellent" and "general" categories reflects the balance between exploitation (local search) and exploration (global search).

Let the optimisation problem be defined over a d-dimensional search space, with a population of n candidate solutions:

$$X_i = (x_{i1}, x_{i2}, \ldots, x_{id}), \quad i = 1, 2, \ldots, n, \qquad (16)$$

where $X_i \in R_d$ is the position of the i^{th} solution. The quality of each candidate solution is evaluated by a fitness function:

$$F_i = F(X_i), \quad i = 1, 2, \ldots, n, \qquad (17)$$

where $F(\cdot)$ represents the objective function of the problem. Based on their fitness values, all solutions are ranked, and the population is divided into two groups:

- **Excellent group:** top-performing solutions (high fitness) that refine search locally.
- **General group:** remaining solutions that diversify the search globally.

The update rule of the CSO mimics competition rounds: high-ranked candidates exploit the search space by moving closer to the global best, while low-ranked candidates explore alternative regions through random perturbations. The general update equation is given by:

$$X_i^{t+1} = X_i^t + \alpha \cdot \left(X_{\text{best}}^t - X_i^t\right) + \beta \cdot R, \quad (18)$$

where X_i^t is the position of solution i at iteration t, X_{best}^t is the best solution at iteration t, R is a random perturbation vector encouraging exploration, and α, β are adaptive learning parameters controlling the trade-off between exploitation and exploration.

Through successive iterations, the CSO balances intensification and diversification, thereby improving the convergence to the global optimum.

2.5.1. Framework of the algorithm and mathematical modelling

The competition metaphor can be formally interpreted through the following rules:

- **Rule 1 (Ranking and grouping):** All candidate solutions are evaluated using the fitness function $F(X_i)$. The population is then ranked and divided into two groups: the excellent group (top-performing individuals) and the general group (remaining individuals):

$$\begin{aligned} &P = P_{\text{excellent}} \cup P_{\text{general}}, \\ &P_{\text{excellent}} \cap P_{\text{general}} = \emptyset. \end{aligned} \quad (19)$$

- **Rule 2 (Learning ability):** Each solution adapts its search behaviour according to its group. Excellent solutions focus on exploitation, improving their accuracy by learning from the global best. General solutions emphasise exploration, introducing randomness to escape local optima.
- **Rule 3 (Dynamic progression):** Over successive iterations, excellent solutions refine their search around high-fitness regions, while general solutions contribute diversity. This interaction prevents premature convergence while steadily guiding the population towards the global optimum.

Formally, the learning dynamics can be modelled as:

$$X_i^{t+1} = \begin{cases} X_i^t + \gamma \cdot (X_{\text{best}}^t - X_i^t) & X_i \in P_{\text{excellent}} \\ X_i^t + \delta \cdot R & X_i \in P_{\text{general}}, \end{cases} \quad (20)$$

where γ controls the exploitation rate for excellent solutions and δ regulates the exploratory step for general solutions.

This framework bridges the intuitive competition analogy with a rigorous mathematical model, ensuring both interpretability and effectiveness of CSO for complex optimisation problems.

2.6. Datasets

The oral and maxillofacial pathology (MOD) dataset consists of pictures taken at a dentist office in Okara, Punjab, Pakistan, as well as pictures from other dental websites. The collection contains a total of 517 photos. Oral ulcers (CaS), gingivitis, oral mesothelioma (OM), oral cancer (OC), oral lichen planus (OLP), oral candidiasis (OT), and cheilitis (CoS) are the seven disease groups into which these are separated. There are $256 \times 256 \times 3$ pixels in the image. We expanded the dataset using methods including rotation, cropping, and flipping it both vertically and horizontally because of its small size. A total of 5 170 photos were created for training and validation.

Using the OCI (Oral Cancer Images) dataset, the suggested approach was evaluated. The dataset was created by a number of ENT facilities in Ahmedabad, India, Shivam Barot and Prakrut Suthar. It is separated into two groups: the non-cancerous group, which has 47 images, and the malignant group, which has 87 images. Several diagnostic tools can be thoroughly validated with the help of a large number of instances. After that, ENT professionals carefully examined these pictures to guarantee that the labels were correct. We employ data augmentation techniques for each class independently, maintaining the photos within their respective classes, in order to expand the size of the dataset. To prevent skewed performance outcomes, we then used the five-fold cross-validation approach on the dataset. Figure 5 displays MOD example images.

2.7. Confusion matrix

We used the confusion matrix for each system to assess the correctness of the suggested models. Equations (21)–(25) demonstrate the following characteristics: accuracy, sensitivity, specificity, and area under the curve, respectively. The confusion matrix displays the proportion of texture images that were accurately identified as true negatives (TN) and true positives (TP). False negatives (FN) and false positives (FP) indicate the number of photographs falsely classified.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \times 100\,\%, \quad (21)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\,\%, \quad (22)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\,\%, \quad (23)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\,\%, \quad (24)$$

$$\text{AUC} = \frac{\text{Sensitivity}}{\text{Specificity}}. \quad (25)$$

Figure 5. Sample images of MOD datasets.

Hyperparameter	Value
Learning rate scheduler	Exponential decay
Layer dropout rate	0.25
Alpha (Leaky ReLU)	0.01
Max gradient norm	5
Batch normalization momentum	0.99
Epoch	30
Batch size	32

Table 1. Hyperparameter and values of the proposed model.

3. Results

A thorough analysis of the dataset, experimental configuration, and model assessment metrics for classification tasks is provided in this section. Specificity, recall, accuracy, precision, and F1-score were among the key metrics used to assess the classification results of several deep learning models in order to highlight the benefits of the proposed model. The cross-validation results are discussed to further demonstrate the model's stability and generalisability. In order to verify the significance of each component in obtaining the best outcomes, an ablation study was incorporated to examine the effects of various model elements. The advantages of the suggested model in terms of classification and segmentation performance were illustrated by comparisons with state-of-the-art methods from various studies.

3.1. Experimental configuration

Using Python 3.8 and powerful deep learning frameworks, such as PyTorch 1.10 and TensorFlow 2.6, the study developed and refined the models. Data were manipulated and analysed using Pandas 1.3.3 and NumPy 1.21.2, while visual representations were made using Matplotlib 3.4.3 and Seaborn 0.11.2. A powerful workstation with an NVIDIA RTX 3090 GPU, 24GB VRAM, 64GB DDR4 RAM, and an Intel Core i9-11900K processor – all designed to manage demanding computations and massive datasets – ran the computing environment. Table 1 contains a list of the values and parameters that were used in the procedure. The entire research process is improved by this configuration, which makes it possible to train, validate, and evaluate models effectively.

With the Receiver Operating Curve (ROC) curve, the effectiveness of a binary classification system can be visually evaluated. To show the relationship between the true positive rate (TPR) and the false positive rate (FPR), the ROC curve is used to change the classification threshold. Two fields, where this technique is frequently applied, are data analysis and machine learning. The true positive rate, or TPR, is defined as the proportion of correctly identified positive cases to actual positive cases. Comparatively, the false positive rate (FPR) is calculated by dividing the number of incorrectly projected positive cases by the number of actual negative cases. The ROC curve is a very important tool to evaluate how a classifier balances sensitivity and specificity and is also helpful in comparing the performance of different classifiers.

The ROC curve for the classification of oral cancer is shown in Figure 6. The study's ROC curve exhibits a strong classification performance, successfully striking balance between sensitivity and specificity for the detection of oral cancer. With an AUC of nearly 1.0, the suggested model outperforms prior studies in terms of accuracy and classification abilities.

A confusion matrix, which is a table used to evaluate a classification model's efficacy (e.g. a model used to diagnose oral cancer), is displayed in Figure 7. The number of false positives (FP), true negatives (TN), false negatives (FN), and true positives (TP) is separated in the matrix. The model was able to correctly identify 677 instances as non-cancerous (TN) and 658 cases as malignant (TP). It mislabelled 53

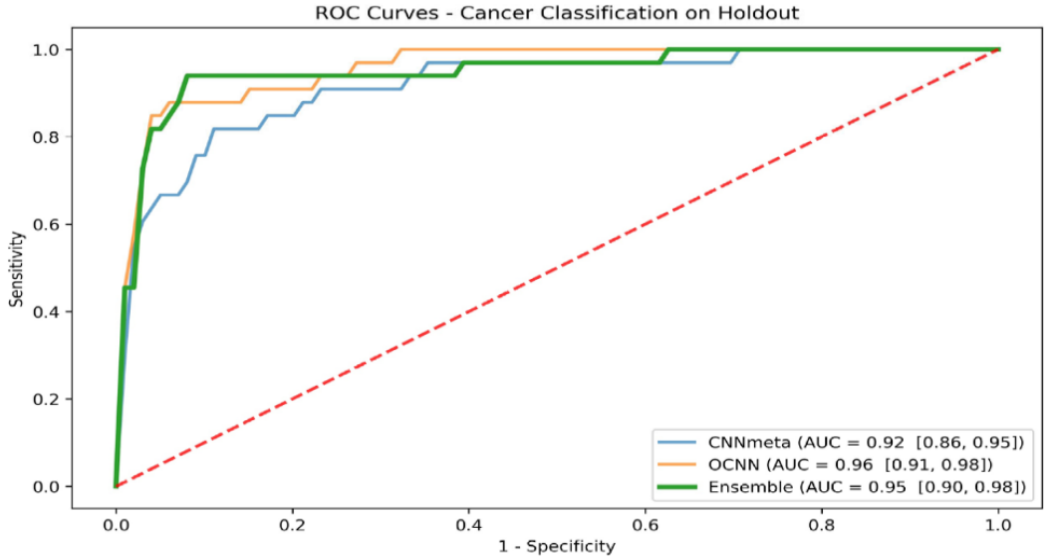


FIGURE 6. ROC curves for the oral cancer.

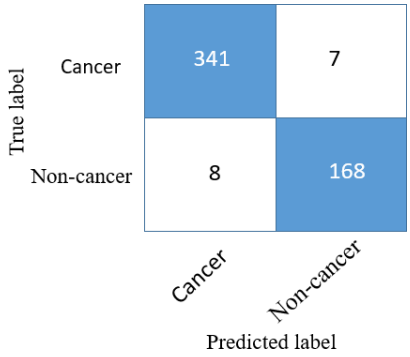


FIGURE 7. The confusion matrix for MOD dataset.

cancer cases as FN and 95 non-cancerous cases as FP. Determining performance metrics like accuracy, precision, sensitivity, and specificity requires knowing these values. They measure how well a model can differentiate between cases that are malignant and those that are not.

Over 30 epochs, the graphs show the evolution of training and validation accuracy as well as training and validation loss for the oral cancer dataset's classification. Figure 8 illustrates the model's training accuracy of 0.99 and validation accuracy of 0.92, as can be seen from the graph. After the first few epochs, the validation accuracy closely follows the training accuracy on the accuracy plot, indicating an effective generalisation without overfitting. The model exhibits rapid convergence. The loss curves demonstrated a steep drop in the first ten epochs before levelling out as the model got closer to the last epochs, demonstrating the model's capacity to learn efficiently while retaining high accuracy and stability during the training phase.

3.2. Segmentation results

The segmentation results for the oral cancer samples are shown by comparing the original images, the mask images, and the segmented outputs, each annotated with the Dice coefficient. The model demonstrates a highly accurate segmentation performance, with Dice coefficients ranging from 99.01 % to 99.70 %, which highlights its precision in identifying cancerous regions. These high coefficients, displayed alongside each segmented image, signify the effectiveness of the model in closely matching the predicted regions with the actual cancer areas, as illustrated in Figure 9.

3.3. Classification results

The proposed model has been proven to be effective, and the classification accuracy for the MOD dataset demonstrates its excellent ability to accurately identify different cancer types. With outstanding overall accuracy, for the MOD dataset, it continuously maintains high levels of precision, recall, and specificity across all categories. The model demonstrated a great capacity to generalise across a range of datasets with its exceptional classification accuracy, which ranged from 97.38 % for The Gorilla Troops Optimizer (GTO) to an astounding 99.94 % for The Competitive Search Optimizer (CSO). Across all classes, the precision metrics continuously surpassed 98.30 %, thereby reducing false positives (Table 2). Meanwhile, the model's capacity to reliably identify true positives was demonstrated by recall metrics that continuously remained over 98.12 %. With F1 scores ranging from 98.43 % to 98.70 %, the classification technique's efficacy is demonstrated, as well as maintaining a balance between precision and recall. With the Competitive Search Optimizer (CSO) achieving the greatest specificity of 99.02 % and the Gorilla Troops Optimizer (GTO) achieving at least 98.70 %, the model's ability to reduce false-positive detection is evident.

3.4. Fold cross validation

The cross-validation findings in Table 3 show high levels of accuracy, precision, recall, F1-score, and

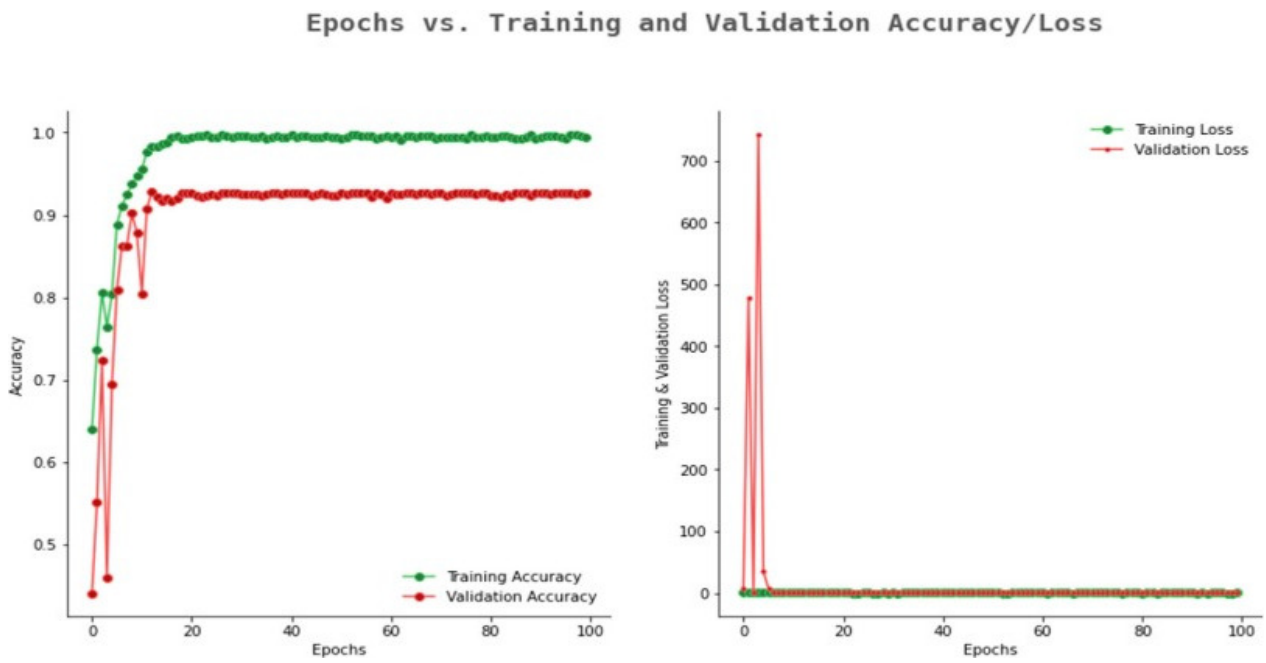
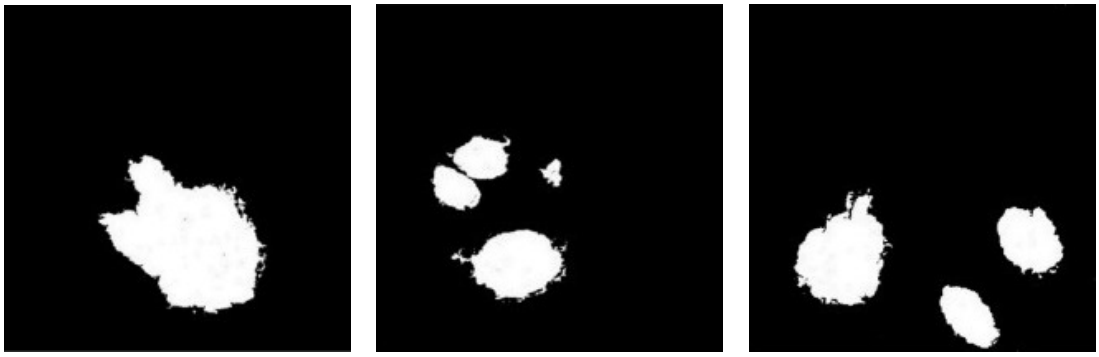
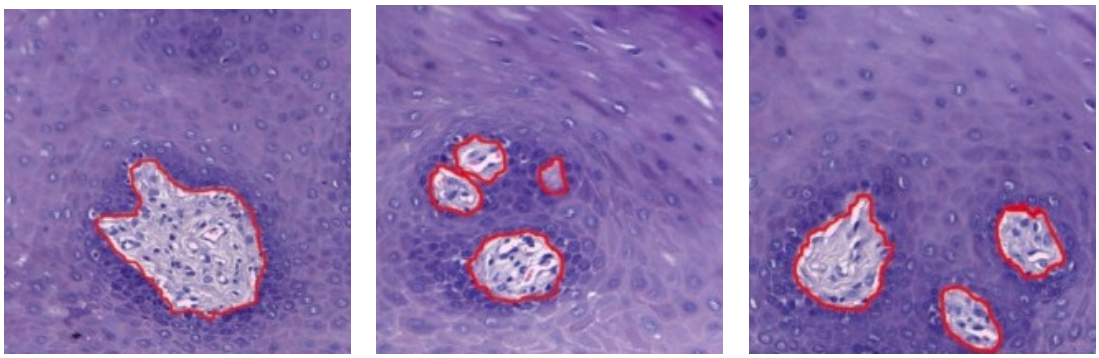


Figure 8. Accuracy and loss graph of Inception Net.



(a). Masks.



(b). Segmented images.

Figure 9. Mask and segmented images with dice coefficients for oral cancer dataset.

Methods	**Accuracy** [%]	**Precision** [%]	**Recall** [%]	**F1 Score** [%]	**Specificity** [%]
The Gorilla Troops Optimizer (GTO)	97.38	98.30	98.12	98.43	98.70
Sailfish Optimization	97.38	98.30	98.12	98.43	98.70
The Aquila Optimizer (AO)	98.27	98.05	98.56	98.67	98.59
Giraffe Kicking Optimization (GKO)	99.12	98.56	98.76	98.70	98.92
The Competitive Search Optimizer (CSO)	99.94	98.84	98.88	98.85	98.02

Table 2. Classification performance for MOD dataset.

	Accuracy [%]	**Precision** [%]	**Recall** [%]	**F1 Score** [%]	**Specificity** [%]
Fold 1	99.54	99.53	99.52	99.55	99.59
Fold 2	99.48	99.54	99.5	99.51	99.56
Fold 3	99.5	99.55	99.52	99.51	99.56
Fold 4	99.51	99.55	99.5	99.52	99.53
Fold 5	99.45	99.53	99.52	99.5	99.6
Average	99.5	99.54	99.51	99.52	99.57

TABLE 3. 5-fold cross-validation results for model performance metrics.

Ablation/Component	**Accuracy** [%]	**Precision** [%]	**Recall** [%]	**F1 Score** [%]	**Specificity** [%]
Baseline (Proposed)	99.5	99.52	99.5	99.51	99.55
Without NLM (Preprocessing)	98.7	98.7	98.65	98.67	98.75
Without LGA (Feature Extraction)	98.9	98.92	98.9	98.91	98.95
Without Seg-UNet (Segmenation)	99.1	99.15	99.13	99.14	99.18
Replaced SV-OnionNet Classifier	98.6	98.62	98.6	98.61	98.7
Without CSO	99.2	99.22	99.2	99.21	99.25

TABLE 4. Ablation study results for model components.

Ref. No.	**Accuracy** [%]	**Precision** [%]	**Recall** [%]	**F1 Score** [%]	**Specificity** [%]
[4]	99.13	84.68	82.68	83.67	82.68
[7]	85	82	82	82	91
[12]	90.36	89.51	89.59	89.35	-
[20]	97.5	98	97	97.5	97
[11]	93.9	93.3	93.3	93.3	94.4
Proposed Model	99.5	99.52	99.5	99.51	99.55

TABLE 5. Comparative analysis of classification performance metrics across studies.

specificity in each of the five folds, demonstrating a consistent performance. The model demonstrated its capacity to correctly categorise instances with few false positives, achieving an average accuracy of 99.5 % and a specificity of 99.57 %. Specifically, the average precision and recall are 99.54 % and 99.51 %, respectively.

3.5. Ablation study

When the NLM pre-processing step was removed, the accuracy of the model dropped significantly, to 98.7 %, indicating the importance of this step for optimal performance. Similarly, removing the attention mechanism from Seg-UNet resulted in a reduced accuracy of 98.9 %, highlighting its role in improving the precision of the segmentation, as shown in Table 4. The ablation study confirms the importance of the LGA mechanism introduced in Section 2.2. Removing it led to a notable accuracy drop (98.9 %), highlighting its critical role in the precision of segmentation. Additionally, replacing the SV-OnionNet classifier led to a further decrease in accuracy, to 98.6 %, as did exluding Competitive Search Optimisation components, which resulted in accuracy dropping to 99.2. These findings illustrate the importance of each component in the proposed model, which contribute to its overall high performance and stability.

The proposed model achieved the highest accuracy of 99.5 %, along with precision, recall, F1-score, and specificity values exceeding 99.5 %, indicating a superior classification performance. In contrast, reference [7] showed the lowest accuracy at 85 %, with all other metrics also being lower, which may reflect limitations in the model or data used. Reference [20] performed well with an accuracy of 97.5 %, but still lower than that of the proposed model, as demonstrated in Table 5. The disparities among studies highlight the effectiveness of the proposed model in delivering consistent and robust classification outcomes across multiple performance metrics, emphasising its suitability for high-stakes applications compared to other models.

4. Conclusion

The adaptive NLM technique is expected to be highly beneficial for diagnosing salivary gland conditions through ultrasound, as it excels in minimising noise while maintaining edge clarity. Among the proposed

components, the LGA mechanism was particularly impactful, as validated by the ablation study results. The Seg U-Net then leverages dense connectivity and attention mechanisms to capture intricate details, ensuring precise boundary delineation between healthy and cancerous tissues. SV-OnionNet was then applied to segment images into oral cancer and normal oral tissue. To quickly examine the model's capacity to predict oral cancer, the suggested approach was further assessed using performance metrics, such as sensitivity, error rate, specificity, and accuracy. The study made clear how important it is to have a reliable and precise tool for diagnosing oral cancer in order to detect it early and lower the risk of death. For the detection of oral cancer, the revised Competitive Search Optimization (CSO) tool significantly increased classification efficiency and accuracy. The patient's life is at stake if oral cancer is not treated promptly. A 99.5 % accuracy, 99.52 % precision, 99.5 % recall, 99.51 % F1-score, and 99.55 % specificity were obtained. Automated cancer diagnosis methods are the result of the rapid expansion of machine learning and deep learning applications in recent years. Future investigations into mouth and oral cancer diagnosis should consider employing a broader dataset. Moreover, the use of multiscale and nature-inspired algorithms could improve diagnostic accuracy.

References

[1] R. L. Siegel, K. D. Miller, N. S. Wagle, A. Jemal. Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians* **73**(1):17–48, 2023. https://doi.org/10.3322/caac.21763

[2] L. Malinverno, V. Barros, F. Ghisoni, et al. A historical perspective of biomedical explainable AI research. *Patterns* **4**(9):100830, 2023. https://doi.org/10.1016/j.patter.2023.100830

[3] Q. Huang, H. Ding, N. Razmjooy. Optimal deep learning neural network using ISSA for diagnosing the oral cancer. *Biomedical Signal Processing and Control* **84**:104749, 2023. https://doi.org/10.1016/j.bspc.2023.104749

[4] T. Flügge, R. Gaudin, A. Sabatakakis, et al. Detection of oral squamous cell carcinoma in clinical photographs using a vision transformer. *Scientific Reports* **13**(1):2296, 2023. https://doi.org/10.1038/s41598-023-29204-9

[5] M. Krichen. Convolutional neural networks: A survey. *Computers* **12**(8):151, 2023. https://doi.org/10.3390/computers12080151

[6] A. Chloupek, D. Jurkiewicz, J. Kania. The characteristics of Polish patients with salivary gland tumors: A ten-year single-center experience. *Clinical Oral Investigations* **28**(1):3, 2024. https://doi.org/10.1007/s00784-023-05396-2

[7] A. H. Alsaab, S. Zeghib. Analysis of X-ray and gamma ray shielding performance of prepared polymer micro-composites. *Journal of Radiation Research and Applied Sciences* **16**(4):100708, 2023. https://doi.org/10.1016/j.jrras.2023.100708

[8] H. Kim, Y. Lee. Comparative evaluation of filters for speckle noise reduction in a clinical liver ultrasound image. *Journal of Radiological Science and Technology* **46**(6):475–484, 2023. https://doi.org/10.17946/jrst.2023.46.6.475

[9] S. Ghufran, P. Soni, G. R. Duddukuri. *Bioprospecting of Tropical Medicinal Plants*, chap. The Global Concern for Cancer Emergence and Its Prevention: A Systematic Unveiling of the Present Scenario, p. 1429–1455. 2023. https://doi.org/10.1007/978-3-031-28780-0_60

[10] C. E. Niekrash, E. M. Ferneini, M. T. Goupil (eds.). *Dental science for the medical professional*. 2023. https://doi.org/10.1007/978-3-031-38567-4

[11] E. S. Mira, A. M. S. Sapri, R. F. Aljehanı, et al. Early diagnosis of oral cancer using image processing and artificial intelligence. *Fusion: Practice and Applications* **14**(1):293–308, 2024. https://doi.org/10.54216/fpa.140122

[12] H. Myriam, A. A. Abdelhamid, E.-S. M. El-Kenawy, et al. Advanced meta-heuristic algorithm based on particle swarm and Al-Biruni earth radius optimization methods for oral cancer detection. *IEEE Access* **11**:23681–23700, 2023. https://doi.org/10.1109/ACCESS.2023.3253430

[13] J. Rashid, B. S. Qaisar, M. Faheem, et al. Mouth and oral disease classification using InceptionResNetV2 method. *Multimedia Tools and Applications* **83**(11):33903–33921, 2024. https://doi.org/10.1007/s11042-023-16776-x

[14] X. Zhao, L. Wang, Y. Zhang, et al. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review* **57**(4):99, 2024. https://doi.org/10.1007/s10462-024-10721-6

[15] I. Pacal. MaxCerVixT: A novel lightweight vision transformer-based approach for precise cervical cancer detection. *Knowledge-Based Systems* **289**:111482, 2024. https://doi.org/10.1016/j.knosys.2024.111482

[16] A. Maman, I. Pacal, F. Bati. Can deep learning effectively diagnose cardiac amyloidosis with 99mTc-PYP scintigraphy? *Journal of Radioanalytical and Nuclear Chemistry* **334**(1):1033–1048, 2025. https://doi.org/10.1007/s10967-024-09879-8

[17] I. Pacal. A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. *International Journal of Machine Learning and Cybernetics* **15**(9):3579–3597, 2024. https://doi.org/10.1007/s13042-024-02110-w

[18] F. Hörst, M. Rempe, L. Heine, et al. CellViT: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis* **94**:103143, 2024. https://doi.org/10.1016/j.media.2024.103143

[19] I. Pacal, O. Celik, B. Bayram, A. Cunha. Enhancing EfficientNetv2 with global and efficient channel attention mechanisms for accurate MRI-based brain tumor classification. *Cluster Computing* **27**(8):11187–11212, 2024. https://doi.org/10.1007/s10586-024-04532-1

[20] N. S. Piyarathne, S. N. Liyanage, R. M. S. G. K. Rasnayaka, et al. A comprehensive dataset of annotated oral cavity images for diagnosis of oral cancer and oral potentially malignant disorders. *Oral Oncology* **156**:106946, 2024. https://doi.org/10.1016/j.oraloncology.2024.106946