IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 3, June, 2025, pp. 1789-1800
Journal homepage: http://iieta.org/journals/ts

A Novel Enhanced U-Net Feature Extraction Using OCT Images for Diabetic Macular Edema Classification

Minarva Devi K[1], Murugeswari S[2*]

[1] Department of Electronics and Communication Engineering, Sri Raaja Raajan College of Engineering and Technology, Karaikudi 630301, India
[2] Department of Bio Medical Engineering and Technology, Syed Ammal Engineering College, Ramanathapuram 623502, India

Corresponding Author Email: murugeskabilan@syedengg.ac.in

https://doi.org/10.18280/ts.420347

Received: 11 January 2025
Revised: 15 May 2025
Accepted: 16 June 2025
Available online: 30 June 2025

Keywords:
optical coherence tomography, self-calibrated convolutions, enhanced U-Net, Henry's gas swarm optimization

ABSTRACT

In this work, an advanced diabetic macular edema (DME) disease classification is proposed by using an optical coherence tomography (OCT) image dataset. This proposed work performs both the feature extraction and classification processes. The proposed feature extraction executed a novel hybrid Enhanced U-Net architecture that includes both the Swin Transformers (SwinT) and Self-Calibrated Convolutions (SC-Convs) to enhance the performance. Initially, Enhanced U-Net consists of a convolutional layer, SwinT to capture hierarchical features and SC-Convs to recalibrate a dynamic feature. The proposed Enhanced U-Net process both a two-stage encoder-decoder architecture using an OCT image through multiple stages to enhance local and global feature representation. The decoder also reconstructed a segmentation map that has a skip connection to attain spatial data. The segmentation head block provides an output of a binary mask with an accurate diagnosis of DME regions. These extracted features are then optimized using Henry's Gas Swarm Optimization (HGSO) model. The HGSO is used to select the feature set by simulating gas particles to determine the most relevant feature from it. This optimized feature set is classified using the XGBoost algorithm, which is a robust model that effectively classifies the healthy and DME-affected regions. Thus, the experimental result shows that the proposed method has attained a more precise and reliable classification of DME disease in all the metrics than the traditional methods. Therefore, the proposed method enhances the diagnostic capabilities and management of diabetic eye diseases efficiently.

1. INTRODUCTION

In recent decades, DME has been a severe visual disease that causes vision impairment among individuals with diabetes [1]. This disease mostly occurs due to fluid accumulation in the macula region, which is in the central part of the retina. It is caused due to a leakage from blood vessels that also leads to high blood sugar levels, which is known as diabetic retinopathy (DR) [2]. If this issue is untreated, then the DME can lead to vision loss and even lead to blindness in the human eye. DME diseases also affect the cardiovascular system, neurons, immunological and digestive disorders [3]. Severe DME is reported among various types, has 4.2% and 7.9% in patients with DM type 1 and 1.4% and 12.8% in type 2 DM.

The symptoms of DME include blurred vision, fade-out colour appearance shadows in the vision field [4]. Some patients might experience it in one eye initially, then it develops in both eyes. The subtle onset of these symptoms often leads to a delayed diagnosis that needs a regular eye examination, especially for diabetic patients. To protect the retina of the eye and its vision, earlier detection of DME is required and treated in the initial stage for effective management [5].

Based on statistics of an Early Treatment Diabetic Retinopathy Study (ETDRS), DME is considered by the thickness of retinal hard oozes, microaneurysm bleeding and macular haemorrhage [6]. So, the Screening process is important for an early diagnosis and treatment to reduce its complexity. In imaging techniques, the most commonly used screening processes for DME are Fundus photography and OCT [7]. Fundus imaging is used to capture detailed images of the retina that provide a broad retinal view and potential abnormalities. Also, OCT imaging has transformed the diagnosis of DME, which provides high-resolution cross-sectional images of the retina. The OCT imaging can enable the visualization of retinal thickness, fluid accumulation, and other changes based on DME. It supports to identification of the severity and extent of the edema provides proper guidance for treatment decisions and also monitors the growth of the disease [8].

Though the OCT is better at identifying, it has an inaccuracy in determining the region relevant to DME effectively in an earlier stage. There is a huge demand for an efficient and accurate tool to assist in the early diagnosis and management of DME. Also, Manual examination and clarification of retinal images by specialists are very time-consuming and also lead to inconsistency [9]. Therefore, an automated system can provide consistent and rapid analysis to address these challenges for DME disease. In recent times, an advanced tool of medical imaging is machine learning (ML) and deep learning (DL) techniques are a promising solution to attain its effectiveness [10].

The ML/DL methods are a subset of artificial intelligence that has a more success rate in image recognition tasks in the medical imaging field [11]. Some of the most peculiar methods used in medical imaging are: Convolutional Neural Networks (CNNs) are the most popular DL architectures that are used to process a spatial hierarchy in images through their layered structure. Some other advanced models, like Inception Networks, are especially known for their multi-scale feature extraction ability where inception modules are used to allow the network to capture features at multiple different scales simultaneously. Recently, attention mechanisms have been added to an architecture to help the network prioritize an image. This supports to detection of the subtle signs of DME that have a greater indication of disease in it. Another popular model is Transfer learning which pre-trains a neural network on a large dataset. All these methods have a specific characteristic in OCT images to improve diagnostic accuracy and reduce the need for widely labelled data.

Therefore, to attain a highly accurate and optimal feature extraction and classification, the proposed work presented an Enhanced U-Net with a hybrid of Swin Transformer (SwinT) and SC-Conv for feature extraction, HGSO-based feature selection and XG Boost classification, respectively. This proposed method has attained better performance than existing methodologies. The remaining part of this article contributes related work in section 2, and section 3 discusses a preliminary part of it. Section 4 describes the material and methods of the proposed architecture, and Section 5 provides a result and discussion with an experimental comparison of the proposed and existing methods. The section ended with a conclusion that is followed by references.

2. RELATED WORKS

Several U-Net-based models have been proposed for medical image segmentation with the integration of different feature learning modules. Ding et al. [12] proposed a Multi-layer Deep Feature Extraction Network (MDE)-Net**.** It integrates a Hybrid Convolutional Feature Extraction (HCFE) module in the encoder to replace traditional convolutional blocks. This module enhances multi-scale feature extraction and expands the receptive field for better segmentation accuracy. Yadav et al. [13] introduced a hybrid model combining EfficientNetB7 as the encoder and UNet++ as the decoder. This model uses transfer learning with AdvProp pre-trained weights for improved feature extraction. In addition, it uses multi-scale feature fusion with skip connections for refined segmentation masks.

Magdy et al. [14] developed the PolyRes-Net for medical image segmentation. It combines Multi-Level Residual Blocks (MLR-blocks) in the encoder and attention gates in the decoder. The network's innovation lies in its Multi-Scale Feature Aggregation (MSFA) block, which is used to consolidate features across decoder steps for improved segmentation performance. Hao et al. [15] proposed MEFP-Net, a Multi-Scale Edge Feature Perception Network. This model includes an additional encoder branch with Global Information Extraction Modules (GIEMs) and Multi-Scale Adaptive Feature Fusion Modules (MAFFMs) to capture both global contextual and detailed features. The Atrous Pooling Dense Perception Module (APDPM) further improves the boundary feature representation of the images.

Karimi et al. [16] presented DEU-Net, a Dual-Encoder U-Net architecture combining a convolutional encoder and a transformer encoder. This dual-branch design concurrently extracts local features and global contextual information for accurate segmentation. Wisaeng [17] proposed U-Net++DSM for skin lesion segmentation. It integrates the Deep Supervision Mechanism (DSM) with the U-Net++ architecture to learn the features deeply.

Various DL models have been proposed for analysing DME. Each model has unique architectures and techniques for improved classification. Kumar et al. [18] presented a DenseNet121 method that was used to extract a feature vector to identify DME patients. The extracted features are processed using fully connected layers before and then moved for classification. Then the final layer provides a classification result for a DME. This DenseNet121 method achieved a classification accuracy of 86.4%, which showed its effectiveness in feature extraction and classification in DME disease.

Zubair et al. [19] addressed issues like fovea localizing, blood vessel extraction, and segmenting lesions by using a hybrid of improved image subtraction, Gabor wavelet filtering and fuzzy c-means clustering methodologies. This hybrid model achieved a high accuracy of 96.17% for optic disc detection, 98.60% for fovea localization, 97.85% for exudates segmentation, and 98.80% for DME classification, respectively. This method attained a superior performance by enhancing retinal image analysis.

da Costa et al. [20] investigated a VGG-19 network that was pre-trained using an ImageNet dataset to classify OCT images. These images are categorised as choroidal neovascularization (CNV), Drusen, DME and Normal. The VGG-19 model achieved an accuracy of 82.60% and an area under the receiver operating characteristic curve (AUROC) of 92.03%. This performance attained a strong performance in classifying various retinal types' status.

Rodríguez-Miguel et al. [21] combined CNNs with recurrent neural networks (RNNs) to classify a DME using OCT image cubes. This hybrid method is used to enable the model to capture both spatial and temporal dependencies within the OCT data. This method acquired its robustness in classification among DME and normal cases.

Hughes-Cano et al. [22] presented transfer learning for a DME classification using multiple imaging modalities like OCT, scalogram, spectrogram and fundus images. This Transfer Learning achieved higher performance in OCT and scalogram images that have 93% AUC and 89% F1-score respectively.

Moreno-Martínez et al. [23] focused on DME classification to detect a dexamethasone implant treatment. The classification system, developed by ESASO, that used to assess the treatment output and reduce the DME's severity. This method provided the most valuable determination with an efficient result that supports a treatment.

Ambure et al. [24] designed an automatic DME detection and grading from retinal images of the OCT dataset. It is processed in three stages such as macula localization, exudate detection and macular coordinate grading. The CNN models are applied to improve the accuracy of DME grading and enhance clinical treatment decisions as soon.

Kiciński and Gawęcki [25] presented an OCT-based classification of various types like perifoveal, central, and peripheral retinal thickness (RT) and choroidal thickness (CT) in DME patients. It evaluates correlations among RT, and CT and also provides an analysis of retinal changes in DME.

Wu et al. [26] developed a DL model to classify a morphological pattern of DME in OCT images. The model achieved high accuracy rates of 90.2%, 95.4% and 95.9% for all these patterns respectively. It showed its capability to identify and classify the DME and normal data effectively.

Kaymak and Serener [27] used the AlexNet algorithm to classify OCT images. It is classified into various categories like healthy, dry AMD, wet AMD and DME. The AlexNet model achieved a greater accuracy of 99.6% and overcame all the issues of previous methods with its effectiveness. This method is used to distinguish among various retinal conditions with high precision.

Wu et al. [28] presented a DME classification model using a Squeeze-and-Excitation attention method. This method allowed us to focus more on channel features and ignore the less relevant features. This attention process attained a higher accuracy and showed the model's ability to classify DME accurately.

Tang et al. [29] used multitask CNNS to classify DME into centre-involved DME (CI-DME), non-CI-DME (non-CI-DME) and normal. The residual network (ResNet) method is used to perform feature extraction and classification efficiently. The superior performances of ResNet enhance the overall segmentation and classification accuracy of DME.

3. PRELIMINARIES

This section provided a preliminary of the proposed work that contains a detailed description of SC-Convs and SwinT.

3.1 SC-Convs

In the clustered convolutions, the feature extraction process is carried out uniformly across multiple parallel branches with each branch operating independently. To form the final output, the outputs from all parallel branches are concatenated. In Contrast, the SC-Convs are used to split the learnable convolutional filters into multiple portions. Also, each portion is designated a specific role rather than being treated equally. The architecture of SC-Convs is given in Figure 1.

3.2 Design workflow

Let's consider the number of input channels (C) is equal to the number of output channels (C) as simple. A set of filter kernels K with dimensions (C, C, H, W) where H and W represent the spatial height and width. It is divided into four portions and each one is responsible for a unique function. Assuming C can be divided by 2 with four filter sets of {K1, K2, K3, K4} then it values each with dimensions as (C/2, C/2, H, W).

Let X1, and X2 are two input portions where each follows a distinct pathway to attain various types of data. The first pathway involves a SC operation using {K1, K2, K3} on X1 that outputs as Y1. The second pathway performs a direct convolution on X2 with K1 that is used to attain the original spatial data and output as Y2. These Y1 and Y2 are concatenated to form the final response Y.

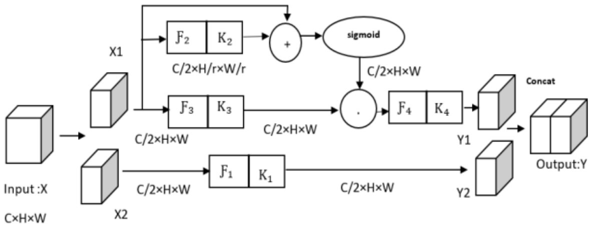


Figure 1. SC-Convs architecture

3.3 SC process

The SC operation is used to collect data information for each spatial location effectively by processing a feature transformation into the original scale space and smaller latent space. The original scale space is the feature maps that retain the same resolution as the input. Then the smaller latent space is processed a downsampling. The transformed features in the smaller latent space are transformed into an original space because of their broader view.

The process applied an average pooling with a filter size of r x r and a stride of r to the input X1 that provides transformed output (T1) as AvgPool (X1).

Next, a feature transformation on T1 is performed using K2:

$$X'1 = UP\big(F2(T1)\big) = Up(T1 * K2) \tag{1}$$

where, Up (·) denotes bilinear interpolation and $F2$ indicates a feature map down-sampling operation.

The calibration operation can then be expressed as:

$$Y'1 = F3(X1) \cdot \sigma(X1 + X'1) \tag{2}$$

$$F3(X1) \ = \ X1 \ * \ K3 \tag{3}$$

where, σ represents the sigmoid function, $F3$ indicates a feature map down-sampling operation and '·' denotes element-wise multiplication. Using X'1 as residuals to form the calibration weights, then the final output after calibration is given by:

$$Y1 = F4(Y'1) = Y'1 * K4 \tag{4}$$

where, $F4$ indicates a feature map up-sampling operation.

Therefore, the SC-Convs technique improves the feature extraction process by integrating diverse data from different scales. This method allocated a specific role to each portion of the filters and used a smaller latent space to guide transformations. This method attained more effectiveness in capturing relevant features and improving the overall performance of extractions.

3.4 SwinT architecture

The SwinT architecture is developed by its tiny version that divides an input RGB image into non-overlapping patches via a patch-splitting module similar to the Vision Transformer (ViT). Each patch is treated as a "token" with its features represented by the raw RGB pixel values concatenation. For implementation, a patch size of 4×4 is used, resulting in a feature dimension of 4×4×3=48. These raw features are projected to an arbitrary dimension (C) using a linear embedding layer. The architecture of SwinT is shown in Figure 2(a).

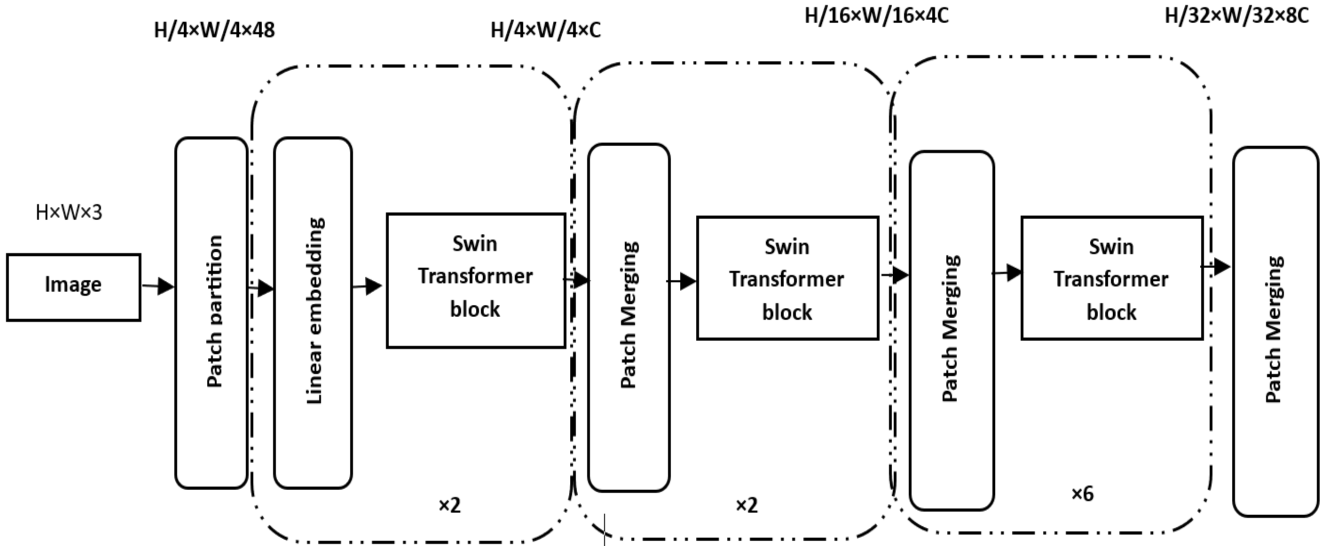
Several SwinT blocks include modified self-attention techniques that are applied to these patch tokens. These blocks are used to maintain the number of tokens (H/4 × W/4) that contain a "Stage 1" architecture along with the linear embedding layer.

To provide a hierarchical representation, the patch merging layers as the network deepens are used to reduce the number of tokens. The initial patch merging layer concatenates the features of each group of 2×2 neighboring patches. Then it is moved to a linear layer on the 4C-dimensional feature concatenation. This minimises the number of tokens by 4 factors that is 2× downsampling of resolution with the output dimension set to 2C. SwinT blocks then performed a feature transformation to maintain an H/8×W/8 resolution. All the patch merging and feature transformation sequences are in "Stage 2".

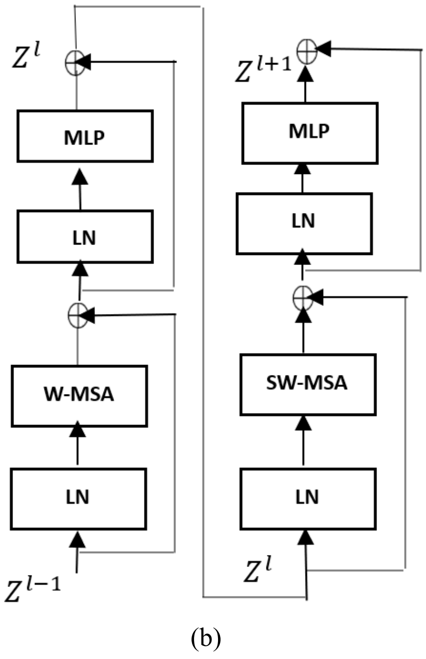
This process is repeated to attain a "Stage 3" and "Stage 4" with a resolution of H/16×W/16 and H/32×W/32 respectively. All these four stages provided a hierarchical representation with feature map resolutions. Therefore, the SwinT architecture can replace the backbone networks in different vision activities. The successive connections of SwinT blocks are given in Figure 2(b).

3.5 SwinT Blocks

The SwinT module consists of a shifted window using a Multi-head Self Attenuation (MSA) module that is followed by a 2-layer Multi-Layer Perception (MLP) with in-between GELU nonlinearity. Each MSA module and MLP is processed before by a LayerNorm (LN) layer and then the residual connection is applied after it.



(a)



(b)

Figure 2. (a) SwinT architecture; (b) Two successive SwinT blocks

4. MATERIALS AND METHODS

4.1 Dataset description

A set of OCT image datasets is acquired by the Singapore Eye Research Institute (SERI) and the Chinese University of Hong Kong (CUHK) respectively. This SERI dataset has a DME case of 16 numbers and normal cases of 16 numbers. The CUHK dataset has DME cases of 4 16 numbers and normal cases of 79 16 numbers. Every volume has 128 B-scan slices of 1024×512 pixels. All OCT datasets are used to train the data with grades and label it as normal or DME-affected. This evaluation is done by measuring a retinal thickening, oozes of hard, intraretinal cystoid space formation and subretinal fluid.

The OCT datasets used in this work are sourced from publicly available repositories and research institutions. These datasets were collected under strict ethical protocols approved by their respective institutional review boards. All patient information was anonymized before use to ensure confidentiality and adherence to international data protection standards.

4.2 Methodology

In this method, DME detection involves pre-processing, feature extraction, feature selection and classifications to ensure an accurate diagnosis. Initially, the collected data is pre-processed to provide the data fit for proposed computations. Then the processed data are used for a feature extraction by using an Enhanced U-Net architecture. This proposed model consists of convolutional layers, a SwinT module and SC-Convs that are used to attain both local and global features from the OCT images. Next, the extracted features are selected using the HGSO algorithm. It optimises the combination of features and selects the most relevant features for classification. Finally, the selected features are classified using an XGBoost algorithm that uses gradient boosting to create a robust model that accurately distinguishes between healthy and DME regions in the retina. The overall workflow is given in Figure 3.

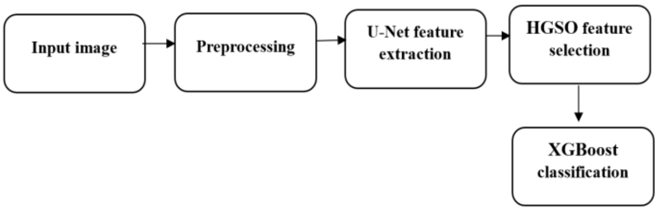


Figure 3. Proposed block diagram

4.3 Feature extraction

The Enhanced U-Net architecture is associated with a SwinT module and SC-Convs to process an accurate extraction of features in DME using OCT images. The processed OCT image is served as an input to the enhanced U-Net model. Initially, this UNet processes an Encoder and Decoder to attain a feature extraction. The structure of proposed UNet is given in Figure 4.

4.4 Encoder

In the encoder phase, it is used to extract a hierarchical feature that contains multiple layers of convolution, SwinT blocks and SC-Convs blocks that are explained below.

Conv-1: it is an initial layer that is used to provide a filter and extract features of edges and textures from a pre-processed image. This Conv layer modified the raw pixel values into a set of feature maps.

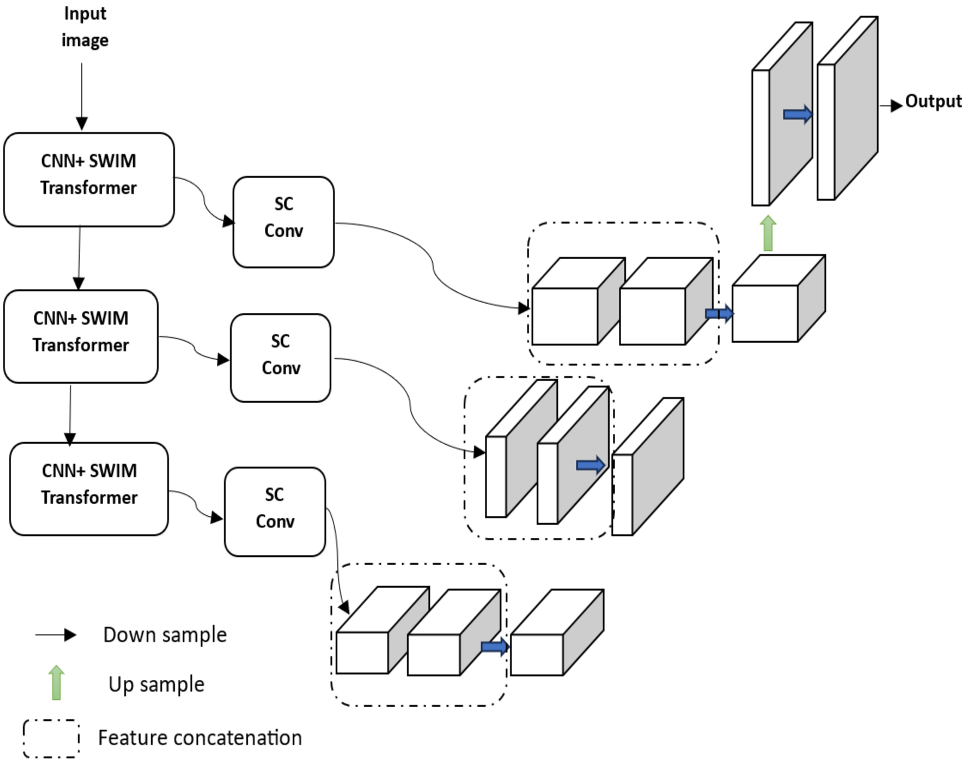


Figure 4. Proposed enhanced UNet with SwimT and SC-Convs architecture

SwinT-1: The output of Conv layer-1 is processed using a SwinT block-1. This transformer block provided hierarchical vision transformers to attain both local and global image data. It is used to divide the feature maps into several non-overlapping patches. Then it applies self-attention to process a long-range dependency for enhancing the feature representations.

SC-convs-1: The features obtained from SwinT block are further refined by SC-conv-1 block. This block is used to modify the feature maps dynamically which marks an important feature region and suppresses the less relevant ones to improve feature quality.

Conv-1: Another convolutional layer follows, further processing the features and preparing them for the next hierarchical level.

Similarly, in the architecture design, the series of conv-2, SwinT-2 and SC-Convs-2 are used to enhance and refine features. After the Conv-2 block, the SwinT-2 block captures more complex and detailed features and then moves to the SC-Convs-2 block that fine-tunes these features to make them stronger and more useful. Next again third conv layer is applied to process the features then the third block of SwinT provides an even deeper into the data to extract high-level features. After that, the final SC-Convs block refines these most important features for enhancement. By carrying out this process, this model helps to identify and extract the significant features accurately.

4.5 Decoder

After the encoder, the decoder is used to segment map reconstruction from the hierarchical extraction of features. It includes upsampling layers, feature concatenation and a convolutional process.

UpSampling: This layer is used to increase the feature maps' spatial resolution. The methods like transposed convolutions or interpolation are used to restore the original image dimensions. It helps to generate a segmentation map that matches the input image size.

Feature Concatenation: in this block, the upsampled features are concatenated with the Skip connections from an encoder. This block is used to retain the spatial data and refine the lost data during the encoding. The process of skip connections accesses a high-resolution feature to enhance segmentation accuracy.

Conv Layers: it processes the combined feature maps that refine the segmentation boundaries and improve the final output.

Segmentation Head: The segmentation head is a final block that provides a probability map output that indicates the presence of DME. This layer has a sigmoid or softmax activation function. It is used to convert the processed feature maps into a binary mask that shows affected regions by DME.

Therefore, the novel Enhanced U-Net architecture with SwinT and SC-Convs is proposed to perform feature extraction for DME features from OCT images. This method achieved a high performance in OCT image segmentation that offered a robust tool to diagnose and manage diabetic eye diseases. These extracted feature data are fed as input to perform a Feature Selection process.

In comparison with existing U-Net variants, DEU-Net uses deformable convolutions to adapt spatial sampling locations. It increases geometric flexibility but lacks the capacity for long-range contextual modeling. MDE-Net applies multi-scale dilated convolutions to capture contextual information at varying receptive fields, but its global modeling is implicit and limited. The proposed Enhanced U-Net integrates Swin Transformers which apply self-attention within shifted windows and hierarchical patch merging. This enables the model to learn explicit global dependencies and inter-region relationships. The proposed model based on self-attention within a Swin block can be expressed as follows:

$$Attention\ (Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d}} + B\right)V \qquad (5)$$

where, Q,K,V are query, key, and value matrices, and B is a learnable relative position bias. Similarly, DEU-Net handles local deformation through spatial adaptivity but may miss fine-grained intensity variations critical in OCT imaging. MDE-Net captures multi-scale edges but fails to handle contrast variation in DME. The proposed model uses SC-Convs to extract spatial-local features guided by attention cues. These convolutions highlight edge regions, lesion boundaries, and textural transitions. This supports finer localization of fluid pockets and structural retinal changes. It can be expressed as follows:

$$Y_{i,j} = \sum_{(m,n)\in\Omega} \alpha_{m,n} \cdot X_{i+m,j+n} \qquad (6)$$

where, $\alpha_{m,n}$ is the attention-weighted kernel coefficient around pixel (i,j), dynamically learned. Unlike DEU-Net and MDE-Net, the proposed U-Net architecture fuses Swin features with SC-Convs in a hierarchical manner to learn both global abstraction and local sensitivity at each level of the decoder. This multi-level fusion allows the model to simultaneously attend to macro-level context (via Swin) and micro-level details (via SC-Convs). This is crucial for segmenting pathological regions with high inter-patient variability.

4.6 Feature selection using HGSO

The feature selection process is presented to identify and retain the most important features from the Enhanced U-Net extracted data. Selecting relevant features improves overall performance and increases the ability of the classification process. Therefore, to access a feature selection, the meta-heuristics model of HGSO is used which is motivated by using Henry's law of gases [30]. It means the amount of dissolved gas in a liquid is proportional to its partial pressure above the liquid. Then the HGSO simulates the gas particle's behavior (i.e., candidate solutions in a search space to provide an optimal solution. HGSO is used for feature selection due to its advantages in handling non-linearity and high-dimensional feature spaces derived from OCT images. Unlike Principal Component Analysis (PCA) or Least Absolute Shrinkage and Selection Operator (LASSO), HGSO is not based on linear assumptions and can capture complex inter-feature dependencies relevant to DME diagnosis. It applies a population-based search mechanism that ensures effective exploration and exploitation for optimal feature subset selection. This results in improved classification performance and reduced overfitting. The HGSO algorithm to attain a feature selection is discussed below.

Initialization: an initial process of HGSO is to set the number of gas particles (N), types of gases and other constants. Then Initialize gas particles randomly within the search space.

Gas Particle Division: it processes to partition the gas particles into clusters based on their types using Henry's constant values. Every cluster represents a subset of the feature space.

Cluster Evaluation: it evaluates the fitness of gas particles in every cluster. It is based on how well a feature particularly contributes. Then, it is used to estimate the best solution of gas particles in every cluster and the overall best in a swarm.

Iterations and Updates: For every gas particle, update its position using Henry's law which simulates the gas's diffusion and solubility. It also adjusts Henry's coefficients and gas particles to refine its positions. It processes a Re-initialization of Worst Particles to explore new regions of the search space. Continuously iterate and evaluate the best gas particle in each cluster and the overall swarm. Then Repeat it until the maximum number of iterations is reached.

Optimal Feature Set: After iterations are attained successfully, it returns the best gas particle with an optimal set of features. The pseudocode for the **HGSO based** feature selection is given below:

```
Function HGSO_FeatureSelection():
    Initialize parameters:
        N = Number of gas particles (population size)
        MaxIterations = Maximum number of iterations
        HenryConstant = Constant for gas solubility
        OtherConstants = Any necessary constants
        FeatureSpace = Set of features to choose from
        Initialize gasParticles randomly within the search space

    // Step 1: Gas Particle Division
    Divide gas particles into clusters based on types using Henry's constant values:
        For each gasParticle in gasParticles:
            Assign gasParticle to a cluster based on solubility and Henry's constant

    // Step 2: Cluster Evaluation
    Evaluate the fitness of each gas particle within the clusters:
        For each cluster:
            For each gasParticle in the cluster:
                Calculate fitness as classification accuracy based on the feature set selected by the gasParticle

    // Step 3: Iteration and Updates
    For iteration = 1 to MaxIterations:
        For each gasParticle in gasParticles:
            // Update the position of each gas particle using Henry's law:
            Update gasParticle position based on gas diffusion and solubility

            // Adjust Henry's coefficient and re-initialize worst particles:
            If gasParticle is the worst in its cluster:
                Reinitialize gas-particle to explore new regions of the feature space

        // Evaluate the fitness again for all gas particles in every cluster
            Update the best gas particle in each cluster and overall swarm

    // Step 4: Optimal Feature Set
    After reaching maximum iterations:
        Identify the best gas particle with the optimal set of features
    Return the optimal feature set
```

4.7 End function

Initially, set the parameters for the search space, gas particles, and constants related to gas solubility.: Partitions the gas particles into clusters based on types. Evaluate the fitness of each gas particle based on how well each feature contributes to the classification task. Updates the gas particles' positions according to the principles of gas solubility, reinitializes worst-performing particles, and iterates until the maximum number of iterations is reached. Once the iterations are complete, the algorithm returns the optimal set of features that best contribute to classification.

4.8 XGBoost classification

After the feature selection, the XGBoost classification is performed with the chosen features given by the HGSO model. The XGBoost method is based on the ML method that is used to classify a DME and normal Cases from input data. It operates by building an ensemble of decision trees in a sequential manner, where every new tree attempts to an error correction made by the previous ones. This is based on gradient boosting that optimizes a differentiable loss function by adding decision trees sequentially. This process is used to reduce the residual errors and improve its regularization techniques to prevent overfitting, efficient missing data handling and speed implementation.

In DME classification from OCT images, XGBoost provides a robust to classify each image region as either healthy or DME-affected. Its ability to handle large datasets and provide high performance in DME image classification with an accurate and reliable diagnostic.

To increase clinical interpretability, attention maps generated by the SwinT are used to visualize the regions of OCT images. It is used for the model's decision-making. These visual explanations help clinicians to understand lesion-specific focus areas. In addition, XGBoost's inherent feature importance ranking is used to identify the most influential features in the final classification. This dual interpretability model bridges the gap between model predictions and clinical reasoning. It guarantees transparency and supports informed decision-making for DME diagnosis.

5. RESULT AND DISCUSSION

The experimental result of the proposed UNet is evaluated and compared with a traditional method. For analysis, the OCT dataset images are acquired and used for this work. This dataset is used as a training-to-testing split of 70% and 30% respectively. The SwinT is configured with a patch size of 4×4, a window size of 7, and a learning rate of 1e-4 using the Adam optimizer. For SC-Convs, kernel sizes of 3 and dilation rates of 2 are assigned. HGSO is executed for 100 iterations with a population size of 30. Initially, the proposed U-Net is compared with other U-Net models for segmentation performance analysis. The training and validation of the loss rate for varying epochs is given in Figure 5. Initially, both training and validation loss decrease rapidly which denotes the effective learning of the proposed U-Net. The metrics used for the analysis are Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and Precision and Sensitivity (Recall) rates. DSC is the measure of overlap between predicted segmentation and the ground truth. IoU is the measured area of overlap between predicted and actual regions. The precision is defined as a fraction of correctly predicted positive pixels out of all predicted positive pixels. The sensitivity rate measures the ability of the model to correctly identify positive pixels. The comparison of the proposed U-Net with other models is given in Table 1.

Table 1. Segmentation performance analysis of proposed U-Net

Method	DSC	IoU	Precision	Sensitivity
Proposed	0.936	0.929	0.947	0.957
MDE-Net	0.903	0.851	0.906	0.937
EfficientNetB7	0.889	0.822	0.878	0.900
PolyRes-Net	0.878	0.806	0.866	0.892
MEFP-Net	0.865	0.791	0.819	0.874
DEU-Net	0.830	0.777	0.810	0.873
U-Net++DSM	0.779	0.771	0.784	0.865

The proposed UNet achieved a superior performance compared to other U-Net variants in medical image segmentation. The proposed approach achieved the highest DSC (0.936) and IoU (0.929) indicating exceptional segmentation accuracy. It attained a Precision of 0.947 and a Sensitivity of 0.957 which shows its reliability in relevant region identification by minimizing false positives. Also, other models like MDE-Net, H-DenseUNet, and EfficientNetB7 show lower precision and sensitivity. Overall, the U-Net shows better performance in terms of all parameters.

The proposed Enhanced U-Net provided extracted features that are optimised and selected using an HGSO to perform a better feature selection. This process provides a greater result in classification that attained a superior performance in DME detection than traditional methods. The comparison is obtained for standard classification metrics like accuracy, precision, Specificity, recall and F1-score respectively is given in Table 2.

The proposed work demonstrates superior performance across various metrics compared to other models including GAN, Transfer Learning, ResNet50, InceptionNet, AlexNet, and DenseNet121 which are given in

- Precision (Figure 6(a)): The proposed model achieves a precision of 98.75% that surpasses GAN (97.12%), Transfer Learning (96.75%), ResNet50 (96%), InceptionNet (96.7%), AlexNet (95.44%), and DenseNet121 (92%).
- Recall (Figure 6(b)): With a recall of 97.67%, the proposed model outperforms GAN (96.35%), Transfer Learning (95.43%), ResNet50 (94.95%), InceptionNet (93.35%), AlexNet (92.85%), and DenseNet121 (91.47%).
- Specificity (Figure 6(c)): The proposed model attains a specificity of 98.52% which is higher than GAN (96.89%), Transfer Learning (95.93%), ResNet50 (95%), InceptionNet (94.35%), AlexNet (93.55%) and DenseNet121 (89.58%)
- F1 Score (Figure 6(d)): The proposed model has an F1 score of 98.32%, better than GAN (97.53%), Transfer Learning (96.24%), ResNet50 (98%), InceptionNet (97.12%), AlexNet (93%), and DenseNet121 (88.25%).
- Accuracy (Figure 6 (e)): Also accuracy of 99.75% is achieved by a proposed model that outperforms GAN (98.46%), Transfer Learning (97%), ResNet50 (96.82%), InceptionNet (98%), AlexNet (99.6%), and DenseNet121 (86.4%).

The ablation study is conducted to evaluate the impact of different architectural components on the overall classification performance of DME detection. The obtianed values are given in Table 3. Initially, a basic CNN combined with XGBoost achieved moderate performance across all metrics. Replacing the CNN with a U-Net backbone improved model performance by enabling better structural understanding. The integration of SwinT and SC-Conv modules further increased precision and F1-score due to enhanced spatial attention and contextual feature extraction. Finally, the proposed model with HGSO achieved the highest performance, with 98.75% precision, 97.67% recall, and 99.75% accuracy, respectively.

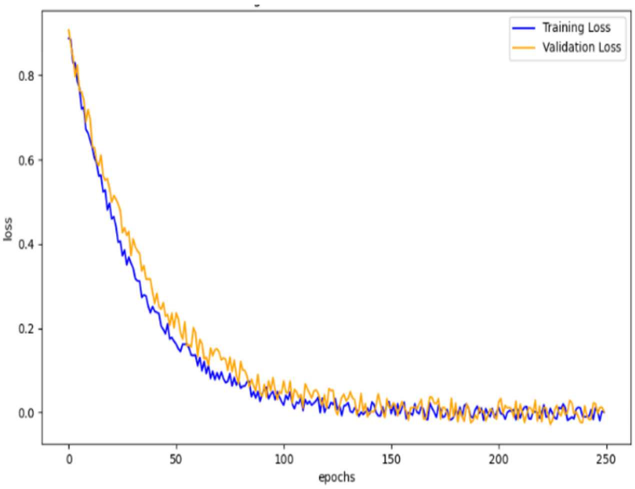
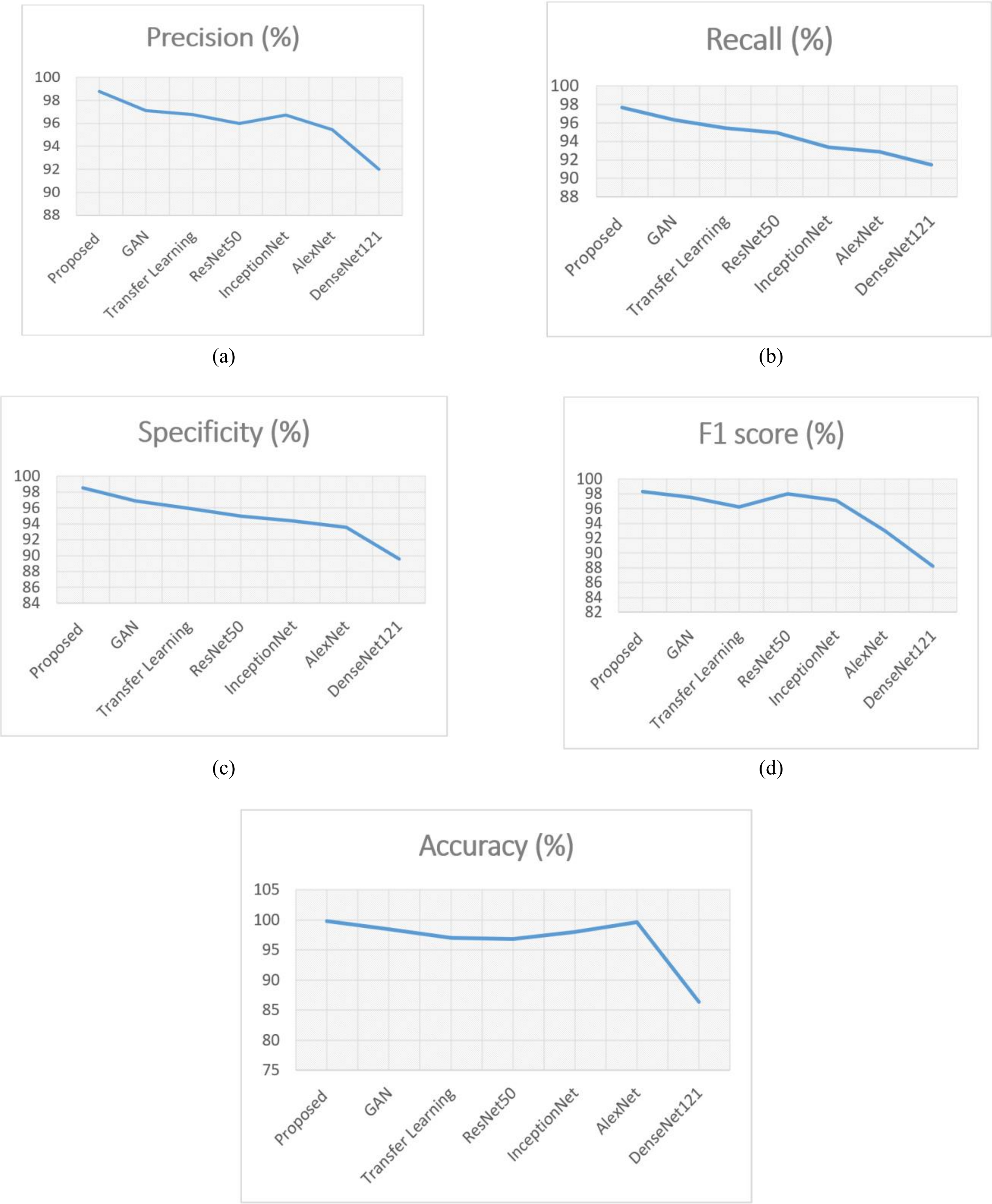


Figure 5. Loss validation of proposed U-Net

Table 2. Performance table of classification metrics

Metrics	Proposed	GAN	Transfer Learning	ResNet50	InceptionNet	AlexNet	DenseNet121
Precision (%)	98.75	97.12	96.75	96	96.7	95.44	92
Recall (%)	97.67	96.35	95.43	94.95	93.35	92.85	91.47
Specificity (%)	98.52	96.89	95.93	95	94.35	93.55	89.58
Accuracy (%)	99.75	98.46	97	96.82	98	99.6	86.4
F1 score (%)	98.32	97.53	96.24	98	97.12	93	88.25



(a)

(b)

(c)

(d)

(e)

Figure 6. Performance of classification metrics a) Precision, b) Recall, c) Specificity, d) F1 score and e) Accuracy

Table 3. Ablation study on classification model components

Model Variant	Feature Extractor	Feature Selection	Classifier	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
CNN+XGBoost	CNN	-	XGBoost	90.62	87.40	88.98	91.20
U-Net (basic) +XGBoost	U-Net	-	XGBoost	93.85	90.95	92.37	94.35
Enhanced U-Net (Swin+SC-Conv)+XGBoost	Swin + SC-Conv	-	XGBoost	96.21	94.58	95.38	97.02
Enhanced U-Net+PCA+XGBoost	Swin + SC-Conv	PCA	XGBoost	97.10	95.28	96.18	98.05
Enhanced U-Net+HGSO+XGBoost	Swin + SC-Conv	HGSO	XGBoost	**98.75**	**97.67**	**98.32**	**99.75**

Table 4. Computational analysis of the enhanced U-Net model

Model	Training Time (hrs)	Inference Time (ms/image)
Enhanced U-Net	3.6	23
Enhanced U-Net+HGSO	4.2	26

Table 5. Comparative evaluation on Kermany [31] dataset

Metrics	Proposed
Precision (%)	97.35
Recall (%)	95.42
Specificity (%)	96.85
Accuracy (%)	98.25

Table 6. Comparative analysis of false negatives on different datasets

Dataset	Total DME Images	Test Split (%)	Test Set Size (DME Images)	False Negatives	False Negative Rate (%)
SERI+CUHK (Primary Dataset)	16	30	5	3	18.75
Kermany [31]	696	30	208	9	4.32

The computational cost analysis of the Enhanced U-Net model on NVIDIA GPU is given in Table 4. The integration of HGSO slightly increases training time, but the inference speed remains within acceptable clinical limits and proves the suitability for real-time implementation.

To further validate the model's robustness, the proposed model is also tested on the larger publicly available OCT dataset by Kermany [31], Mendeley Data, V2. This dataset includes 696 DME images, with a test split of 30% (208 images). The obtained results are given in Table 5. The proposed model maintains superior performance with high precision, recall, and accuracy rates, proving its strong generalization capability across different datasets.

False negatives analysis is very important in clinical settings where, missed detections of DME can lead to delayed treatment. The false negative rate analysis values are given in Table 6. In the analysis, 3 out of 16 DME images (18.75%) were falsely classified as negative in the SERI + CUHK dataset. In another dataset, 9 false negatives are observed out of 208 test images. The false negative rate percentage is 4.32%. This denotes that the model performs more reliably on larger and different datasets. It reduces the clinical risk associated with missed diagnoses.

6. CONCLUSION

The proposed work presented a novel methodology to process a feature extraction and also implement a classification DME from medical images. The proposed model involves the use of an Enhanced U-Net architecture that has SwinT and SC-Convs for robust feature extraction. These features are then optimized using HGSO and classified with XGBoost. Therefore, the proposed model includes its ability to capture both local and global features from OCT images through the Enhanced U-Net architecture attaining a detailed and accurate segmentation. The integration of HGSO optimizes feature selection used to enhance classification performance. The use of XGBoost further refines the classification process that provides the model robust and effective in clinical applications. The numerical validation shows that the proposed model achieved higher results than existing models. The proposed model achieved a precision of 98.75%, recall of 97.67%, specificity of 98.52%, accuracy of 99.75%, and F1 score of 98.32%. Overall, this approach provides a highly accurate solution for diagnosing and managing DME to provide substantial development over existing methods. Future research will focus on integrating multimodal imaging data and expanding datasets through multicenter collaborations to improve model robustness. Additionally, there are plans to develop explainability tools for increased clinical interpretability and validation in prospective studies.

REFERENCES

[1] Hui, V.W., Szeto, S.K., Tang, F., Yang, D., et al. (2022). Optical coherence tomography classification systems for Diabetic Macular Edema and their associations with visual outcome and treatment responses-an updated review. Asia-Pacific Journal of Ophthalmology, 11(3): 247-257. https://doi.org/10.1097/APO.0000000000000468

[2] Parodi Battaglia, M., Iacono, P., Cascavilla, M., Zucchiatti, I., Bandello, F. (2018). A pathogenetic classification of diabetic macular edema. Ophthalmic Research, 60(1): 23-28. https://doi.org/10.1159/000484350

[3] Arf, S., Sayman Muslubas, I., Hocaoglu, M., Ersoz, M. G., Ozdemir, H., Karacorlu, M. (2020). Spectral domain optical coherence tomography classification of diabetic macular edema: A new proposal to clinical practice. Graefe's Archive for Clinical and Experimental Ophthalmology, 258: 1165-1172. https://doi.org/10.1007/s00417-020-04640-9

[4] Ruia, S., Saxena, S., Cheung, C.M.G., Gilhotra, J.S., Lai, T.Y. (2016). Spectral domain optical coherence tomography features and classification systems for diabetic macular EDEMA: A review. The Asia-Pacific Journal of Ophthalmology, 5(5): 360-367. https://doi.org/10.1097/APO.0000000000000218

[5] Jampol, L.M. (2020). Classifications of diabetic macular edema. European Journal of Ophthalmology, 30(1): 6-7. https://doi.org/10.1177/1120672119889532

[6] MD, F.B., Pognuz, R., MD, A.P., MD, A.P., MD, F.M.,

MD, M.A. (2003). Diabetic macular edema: Classification, medical and laser therapy. In Seminars in Ophthalmology. Taylor & Francis, 18(4): 251-258. https://doi.org/10.1080/08820530390895262

[7] Al-Bander, B., Al-Nuaimy, W., Al-Taee, M.A., Williams, B.M., Zheng, Y. (2016). Diabetic macular edema grading based on deep neural networks. In Proceedings of The Ophthalmic Medical Image Analysis International Workshop. University of Iowa, 3(2016): 121-128. https://doi.org/10.17077/omia.1055

[8] Chauhan, M.Z., Rather, P.A., Samarah, S.M., Elhusseiny, A.M., Sallam, A.B. (2022). Current and novel therapeutic approaches for treatment of diabetic macular edema. Cells, 11(12): 1950. https://doi.org/10.3390/cells11121950

[9] Mathews, M.R., Anzar, S.M. (2021). A comprehensive review on automated systems for severity grading of diabetic retinopathy and macular edema. International Journal of Imaging Systems and Technology, 31(4): 2093-2122. https://doi.org/10.1002/ima.22574

[10] Chan, G.C., Muhammad, A., Shah, S.A., Tang, T.B., Lu, C.K., Meriaudeau, F. (2017). Transfer learning for diabetic macular edema (DME) detection on optical coherence tomography (OCT) images. In 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, Malaysia, pp. 493-496. https://doi.org/10.1109/ICSIPA.2017.8120662

[11] Kamble, R.M., Chan, G.C., Perdomo, O., Kokare, M., Gonzalez, F.A., Müller, H., Mériaudeau, F. (2018). Automated diabetic macular EDEMA (DME) analysis using fine tuning with inception-resnet-v2 on OCT images. In 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia, pp. 442-446. https://doi.org/10.1109/IECBES.2018.8626616

[12] Ding, X., Dong, L., Ji, Y., Qian, K.E. (2024). MDE-Net: Multi-Layer depth extraction network with attention mechanism for medical image segmentation. IEEE Access, 12: 177647-177662, https://doi.org/10.1109/ACCESS.2024.3506773

[13] Yadav, A.C., Kolekar, M.H., Sonawane, Y., Kadam, G., Tiwarekar, S., Kalbande, D.R. (2024). EffUNet++: A Novel architecture for brain tumor segmentation using flair MRI images. IEEE Access, 12: 152430-152443. https://doi.org/10.1109/ACCESS.2024.3480271

[14] Magdy, A., Ismail, K.N., Mohamed, M.H., Hassaballah, M., Mahmoud, H., Abdelsamea, M.M. (2024). PolyRes-Net: A polyhierarchical residual network for decoding anatomical complexity in medical image segmentation. IEEE Access, 13: 15312-15323. https://doi.org/10.1109/ACCESS.2024.3475877

[15] Hao, S., Yu, Z., Zhang, B., Dai, C., Fan, Z., Ji, Z., Ganchev, I. (2024). MEFP-Net: A dual-encoding multi-scale edge feature perception network for skin lesion segmentation. IEEE Access, 12: 140039-140052. https://doi.org/10.1109/ACCESS.2024.3467678

[16] Karimi, A., Faez, K., Nazari, S. (2023). DEU-NET: Dual-encoder U-Net for automated skin lesion segmentation. IEEE Access, 11: 134804-134821. https://doi.org/10.1109/ACCESS.2023.3337528

[17] Wisaeng, K. (2023). U-Net++ DSM: Improved U-Net++ for brain tumor segmentation with deep supervision mechanism. IEEE Access, 11: 132268-132285. https://doi.org/10.1109/ACCESS.2023.3331025

[18] Kumar, A., Tewari, A.S., Singh, J.P. (2022). Classification of diabetic macular edema severity using deep learning technique. Research on Biomedical Engineering, 38(3): 977-987. https://doi.org/10.1007/s42600-022-00233-z

[19] Zubair, M., Umair, M., Naqvi, R.A., Hussain, D., Owais, M., Werghi, N. (2023). A comprehensive computer-aided system for an early-stage diagnosis and classification of diabetic macular EDEMA. Journal of King Saud University-Computer and Information Sciences, 35(8): 101719. https://doi.org/10.1016/j.jksuci.2023.101719

[20] da Costa, I.C., Torres-Costa, S., Barbosa, G., Carvalho, E., Parente, M., Guerra, A., Ramião, N., Falcão, M. (2024). Deep learning network to distinguish between retinal vein occlusion and diabetic macular EDEMA. Investigative Ophthalmology & Visual Science, 65(7): 1611-1611.

[21] Rodríguez-Miguel, A., Arruabarrena, C., Allendes, G., Olivera, M., Zarranz-Ventura, J., Teus, M.A. (2024). Hybrid deep learning models for the screening of Diabetic Macular EDEMA in optical coherence tomography volumes. Scientific Reports, 14(1): 17633. https://doi.org/10.1038/s41598-024-68489-2

[22] Hughes-Cano, J.A., Quiroz-Mercado, H., Hernández-Zimbrón, L.F., García-Franco, R., Mijangos, J.R., López-Star, E., García-Roa, M., Lansingh, V.C., Olivares-Pinto, U., Thébault, S.C. (2024). Improved predictive diagnosis of diabetic macular EDEMA based on hybrid models: An observational study. Computers in Biology and Medicine, 170: 107979. https://doi.org/10.1016/j.compbiomed.2024.107979

[23] Moreno-Martínez, A., Blanco-Marchite, C., Andres-Pretel, F., López-Martínez, F., Donate-Tercero, A., González-Aquino, E., Cava-Valenciano, C., Panozzo, G., Copete, S. (2024). ESASO classification relevance in the diagnosis and evolution in diabetic macular edema patients after dexamethasone implant treatment. Graefe's Archive for Clinical and Experimental Ophthalmology, 262(9): 2813-2821. https://doi.org/10.1007/s00417-024-06473-2

[24] Ambure, A.V., Jadhav, N.S., Dudhankar, P.T., Jirafe, V.N., Kamathe, R.S., Borde, S.D. Automated diabetic macular edema (DME) grading. Journal of Technical Education, 47(1): 105-109.

[25] Kiciński, K., Gawęcki, M. (2024). Wide-Field optical coherence tomography in patients with diabetic macular EDEMA. Journal of Clinical Medicine, 13(14): 4242. https://doi.org/10.3390/jcm13144242

[26] Wu, Q., Zhang, B., Hu, Y., Liu, B., Cao, D., Yang, D., Peng, Q., Zhong, P., Zeng, X., Xiao, Y., Li, C., Fang, Y., Feng, S., Huang, M., Cai, H., Yang, X., Yu, H. (2021). Detection of morphologic patterns of diabetic macular edema using a deep learning approach based on optical coherence tomography images. Retina, 41(5): 1110-1117. https://doi.org/10.1097/IAE.0000000000002992

[27] Kaymak, S., Serener, A. (2018). Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. In 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, pp. 265-269. https://doi.org/10.1109/ICCP.2018.8516635

[28] Wu, T., Liu, L., Zhang, T., Wu, X. (2022). Deep learning-based risk classification and auxiliary diagnosis of macular EDEMA. Intelligence-Based Medicine, 6: 100053. https://doi.org/10.1016/j.ibmed.2022.100053

[29] Tang, F., Wang, X., Ran, A.R., Chan, C.K., et al. (2021). A multitask deep-learning system to classify diabetic macular EDEMA for different optical coherence tomography devices: A multicenter analysis. Diabetes Care, 44(9): 2078-2088. https://doi.org/10.2337/dc20-3064

[30] Hashim, F.A., Houssein, E.H., Mabrouk, M.S., Al-Atabany, W., Mirjalili, S. (2019). Henry gas solubility optimization: A novel physics-based algorithm. Future Generation Computer Systems, 101: 646-667. https://doi.org/10.1016/j.future.2019.07.015

[31] Kermany, D. (2018). Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley Data. https://doi.org/10.17632/rscbjbr9sj.2