



# Optoelectronic Retinal Images for the Prediction of Diabetic Macular Edema Based on a Hybrid Deep Transfer Learning Technique

Theenathayaalan Alavanthar<sup>1</sup>, Muthusamy Madheswaran<sup>2\*</sup>

<sup>1</sup> Department of Biomedical Engineering, Muthayammal Engineering College, Rasipuram 637408, India

<sup>2</sup> Department of Electronics and Communications Engineering, Muthayammal Engineering College, Rasipuram 637408, India

Corresponding Author Email: [principal@mec.edu.in](mailto:principal@mec.edu.in)

Copyright: ©2024 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410311>

## ABSTRACT

**Received:** 17 May 2023  
**Revised:** 28 January 2024  
**Accepted:** 1 March 2024  
**Available online:** 26 June 2024

### Keywords:

*diabetes mellitus, diabetic macular edema, U-Nets, VGG-16, extreme learning feedforward networks, transfer learning*

Diabetes is characterized by elevated levels of glucose in the blood, which can lead to complications like Diabetic Macular Edema (DME), causing permanent vision loss. A novel HET-EYE-NETS which is built on the ensemble transfer learning networks with extreme feedforward model for the prediction of DME is proposed. The proposed algorithm pre-processes the color optoelectronic retinal images and classifies the severity of DME by the three-stage pipeline model. In the first stage, DME is segmented by the U-Nets, features of segmented DME are extracted by AlexNet layers and finally severity is predicted by the extreme learning feedforward layers. The extensive experimentation is carried out using IDRiD and MESSIDOR database images. During this process, performance measures like precision, recall, F1-score, specificity, and accuracy are computed and analyzed. In addition, data augmentation is employed to address the issue of data imbalance problem in IDRiD and MESSIDOR database images. The proposed HET-EYE-NETS model achieved an average accuracy of 99.1%, precision of 99.2%, recall of 99% and F1-score of 0.9920. Results proved that proposed HET-EYE-NETS model outperforms existing learning models, demonstrating its potential for early diagnosis of DME.

## 1. INTRODUCTION

Diabetes is a major health threat which affects up to 7.2% of the population world-wide and this will increase up to 650 million by the end of 2040 [1, 2]. Among the diabetics, one third of all diabetics develop Diabetic Retinopathy (DR) and the most complicated stage of DR is Diabetic Macular edema (DME). DME typically manifests itself when the retinal vessels are impacted by the accumulation of fluid [3, 4] and causes vision loss. It affects nearly 2.8% of the population which is estimated to increase even to 10% of the total global population. DME affects around 26.7 million individuals, and it is anticipated that this figure will increase to almost 50 million by 2025 [5-8].

Despite having effective screening for early diagnosis of DME in developed countries, avoiding the false prediction of DME has always been a challenge for diagnosticians. Because of the limited number of ophthalmologists available in developing nations, it is difficult to keep up with the constantly growing DME cases [9, 10]. Also, in developing nations, the provision of appropriate and timely treatment at an affordable cost is another problem in the healthcare industry. Under these circumstances, automated diagnosis frameworks can lower the diagnostic costs and reduce the workloads of ophthalmologists. These systems can also be able to manage the lack of ophthalmologists by restricting referrals to just those cases that require immediate evaluation. To reduce DME cases, it will be essential to reduce the time to diagnosis and effort that

ophthalmologists spend on diagnosis.

Propelled by the above challenges, several imaging diagnosis systems have been developed based on Optoelectronic retinal images using machine and deep learning algorithms [11-20]. A two-stage method is used to identify and classify the severity of DME using colour fundus images [11]. A supervised learning technique is used to carry out the process of DME detection. The feature extraction strategy captures the global characteristics and differentiates the DME and normal images.

A unique model is discussed in the study of Lee et al. [12] to accomplish automated image analysis by combining deep neural networks with machine learning. Optical Coherence Tomography (OCT) provides deep and rich data when combined with labels produced from the electronic medical record. The diagnosis of DR based on Convolutional Neural Networks (CNN) is discussed in the study of Perdomo et al. [13]. It combines images of the eye fundus with the location of exudates for the automated classification of DME. A deep CNN for the classification of DR is discussed in the study of He et al. [14]. The classification of DR, DME, and multi-label are carried out by three different CNNs independent of one another. All CNN's features are fused for effective classification.

A neural network system based on recurrent attention mechanisms is described in the study of Shaikh et al. [15] which helps reduce the computations of processing overhead required when executing convolution filter operations on high-

resolution images. Two different medical images are employed for the classification tasks: brain tumor classification using magnetic resonance imaging and predicting the severity of DME using fundus images.

A novel cross-disease attention module (AM) is developed in the study of Li et al. [16] to classify DME and DR. It is accomplished by investigating the intrinsic link between diseases by using image-level supervision. The disease-specific AM allows for the selective learning of relevant characteristics for diseases, and their internal relationships are captured using the disease-dependent AM.

An efficient framework to correctly locate and classify disease lesions is discussed in the study of Nazir et al. [17]. In contrast to the current DR and DME classification methods, the system can successfully classify low-intensity and noisy images by extracting representative key points from such images. Retinal fundus and OCT images are used to create an automated framework to classify DR and normal in the study of Hassan et al. [18]. It utilizes deep ensemble learning where a deep CNN is used to identify the input OCT and fundus images. Subsequently, the second layer extracts the essential feature descriptors required for the classification.

The automatic detection of AMD and DME is described in the study of Kaymak and Serener [19] using a deep learning technique. It classifies the input image into wet or dry AMD, DME, and healthy. The effectiveness of Iowa Detection Programme for automated DR detection using publicly available fundus images is discussed in the study of Abramoff et al. [20].

These automated diagnosis frameworks can decrease the cost and workloads, as well as the shortage of ophthalmologists. Also, these methods play a pivotal role in reducing the DME cases and reducing the clinician’s efforts. But these methods need its improvisation in terms of achieving accurate segmentation, more subtle feature extraction along with the high-speed classification layers that can act as potential tool for high certainty DME diagnosis systems to be used by the ophthalmologists across the globe.

Motivated by this challenge, HET-EYE-NETS model which automatically analyses the fundus images is developed for the prediction of DME. Pre-processing on the images is

performed in three stages such as morphological filtering, pixel intensive testing and image enhancement. The IDRiD and MESSIDOR database images are enhanced using augmentation techniques to address class imbalance problem, hence improving the HET-EYE-NETS model’s performance. After pre-processing, the images are fed to a three-stage pipeline architecture. In the 1<sup>st</sup> stage, the DME region is segmented from the fundus images, and in the 2<sup>nd</sup> stage, image features are extracted and finally detect the DME using high speed classification layers.

The HET-EYE-NETS model is completely built on the principle of transfer learning ensemble with high accurate Extreme Learning Machines (ELM). It is the first of its kind model used for an early diagnosis of the DME from optoelectronic retinal images.

## 2. MATERIALS AND METHODS

The HET-EYE-NETS model is a hybrid transfer learning-based system for DME classification. It consists of three important modules: data preprocessing, U-Net based macular segmentation, and classification by ELM. Figure 1 shows the architecture of HET-EYE-NETS using optoelectronic retinal images for the prediction of DME.

### 2.1 Image datasets

To build robust and accurate classification models, it is important to have a uniform distribution of the datasets. Class imbalance is a common problem in image/signal datasets. It significantly impacts the model’s performance when training the network and often creates an overfitting problem for smaller datasets. To address this issue, data augmentation is employed in which each image undergoes a series of transformations such as flips, jittering, scaling, and rotation, producing a uniform arrangement of data to train the network. The proposed HET-EYE-NETS model uses two publicly available datasets, IDRiD [21] and MESSIDOR [22]. The specifications of these databases are given in Table 1 and Table 2 respectively.

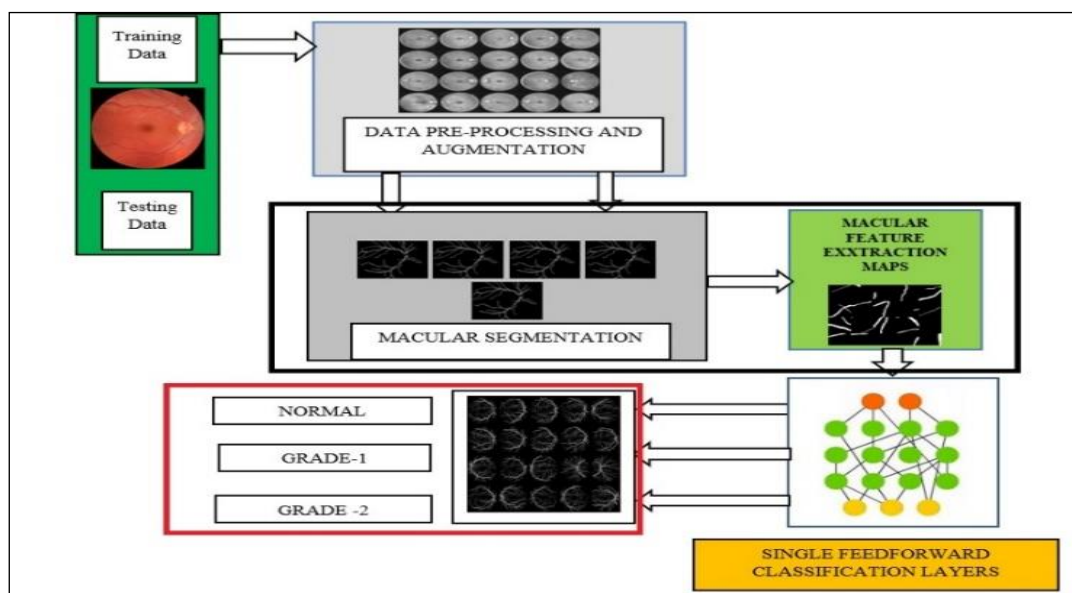


Figure 1. Architecture of HET-EYE-NETS model

**Table 1.** Specification of images in the IDRiD database

Specification	Description
No. of images	516
Source of Images	Images captured at an eye clinic in Nanded, Maharashtra, India
Resolution Used	4288×4288
Grade-1 Images	41
Grade-0 Images	242
Grade-2 Images	243

**Table 2.** Specification of images in the MESSIDOR database

Specification	Description
No. of images	1200
Source of Images	laboratory of medical information processing, Paris, France
Resolution Used	1440×960, 2240×1488, 2304×1536
Grade-1 Images	75
Grade-0 Images	974
Grade-2 Images	151

## 2.2 Data preparation

Pre-processing plays an important role in enhancing the accuracy of the training model by minimizing background noise. It removes the different noise levels in the images, thus making data more consistent for training. Since the image datasets mentioned above have different resolutions, a pre-processing technique is adopted to create the standardized datasets. The proposed model has three pre-processing techniques. First, the morphological filtering technique, which filters background noises, is employed. In the second stage, intensive testing is applied on fundus images to remove the inconsistent and noisy pixels. Finally, image histogram methods to enhance the image quality are adopted.

## 2.3 HET-EYE-NETS

The proposed methodology works on the three different pipelined stages. Macular segmentation is the first stage, followed by feature extraction maps and classification of different macular grades. This research proposes the ensemble layers of transfer learning based on CNN for effective segmentation and feature extraction.

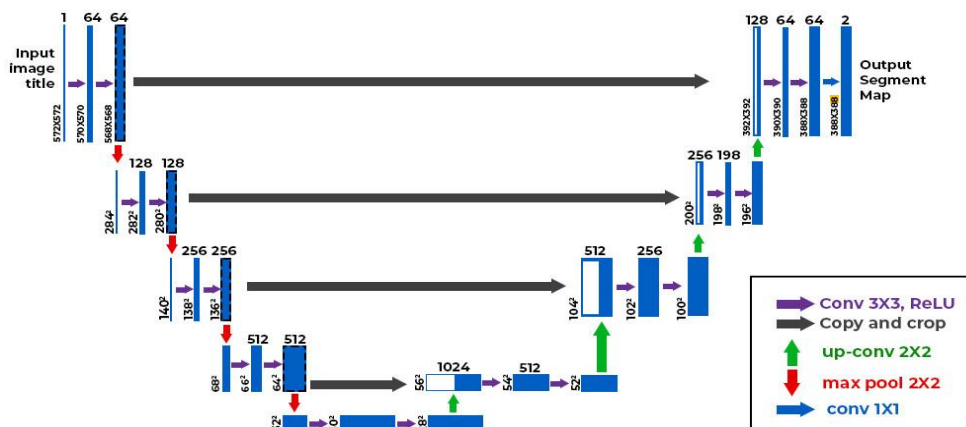
### 2.3.1 Transfer learning mechanism

Various studies have established that transfer learning-based CNN is better than traditional CNN training from scratch. Transfer learning is used in image classification [23], skin cancer diagnosis [24], brain cancer diagnosis [25], and lung cancer diagnosis [26]. Training from scratch suffers from computational overhead and complexity when larger datasets are involved. In the medical field, expert annotation is also an expensive issue. To overcome these problems, transfer learning is adopted to train CNN effectively. In transfer learning, CNN first learns features in one setting and uses the same settings in another task. For effective segmentation and feature extraction, the proposed methodology uses an ensemble of U-Nets and AlexNet.

### 2.3.2 Macular U-Net segmentation

In HET-EYE-NETS model, the U-Net network segments macular from the eye images. U-Net captures local and global characteristics using an encoder-decoder architecture. The encoder gathers contextual information, whereas the decoder provides precise localization. It enables hierarchical feature learning and the skip connections in U-Net helps to preserve fine-grained details. U-Net has lower parameter count than other deep learning networks. Due to its popularity and effectiveness in various computer vision applications, the proposed system employs U-Net for macular segmentation.

The working of U-Nets can be divided into two components. The first approach is the contracting path, which employs a conventional CNN architecture. Every block in the contracting route is comprised of two consecutive 3×3 convolution filters, which are then followed by a Rectified Linear Unit (ReLU) unit and a pooling layer. This pattern is repeated several times to enhance the effectiveness of the training. The unique characteristic of this framework is the use of an extension route in which 2×2 up-convolution is employed to up sample the feature maps. Then, the feature maps in the contracting path are cropped and merged onto the up sampled feature map. Next, there are two consecutive 3×3 convolution filters and ReLU activation. Finally, 1×1 convolution filters are used to reduce the feature maps to the desired channels and then the segmentation results are generated. Cropping is used to eliminate extraneous contextual information and to segment the objects from the surrounding overlapping area. In this work, U-Net is used to separate the macular regions. It has advantages for image segmentation in smaller datasets and effectively mitigating the issue of overfitting. Figure 2 shows the U-Net framework used for macular segmentation.

**Figure 2.** U-Net framework used for macular segmentation

### 2.3.3 Feature extraction

The inputs to the feature extraction layers are the segmented images from the U-Net framework. In the second stage, AlexNet extracts the feature maps from the segmented images. AlexNet success in the ImageNet challenge marked a significant breakthrough in the deep learning revolution. The parallel processing capabilities allow the model to be trained more efficiently than other architectures. Also, the model generalization is improved by local response normalization by AlexNet across local groups of neurons and incorporated ReLU function. It generally comprises of 5 convolutional layers and 3 Fully connected layers. The primary subtleties of each layer in the network are shown in Figure 3.

### 2.3.4 Feedforward layers classification layers

In the third stage, extracted feature maps are used to train the model to classify the different grades of the macular images. The proposed methodology uses ELM [27] for the high-speed and accurate classification of different grades. ELM employs a single hidden layer, which does not necessarily need to be tuned. It utilizes the kernel function to achieve high precision, resulting in improved performance. They have low training error and improved approximation. ELM is mainly used in classification tasks due to its utilization of auto-tuning for biases, weights and non-zero activation functions.

In ELM, the hidden layer's neurons must use an activation function (for instance, the sigmoid function) that is differentiable, while the output layer's activation function of the output layer remains linear. In ELM, it is not necessary to tune the hidden layers' weight and they are assigned randomly including the bias weights. The presence of hidden nodes is significant in this case. However, it is not necessary to adjust them, and the parameters of the hidden neuron may be generated in advance, that is, before handling the training set data. The working principle of ELM is discussed in the study of Wang et al. [28]. A single-hidden layer ELM is defined in Eq. (1).

$$f_L(x) = \sum_{i=1}^L \beta_i h_i(x) = h(x)\beta \quad (1)$$

where,  $x$  is the input features. The output hidden layer ( $h(x)$ ) and output weight vector ( $\beta$ ) are defined in Eq. (2), and Eq. (3) respectively.

$$\beta = [\beta_1, \beta_2, \dots, \beta_L]^T \quad (2)$$

$$h(x) = [h_1(x), h_2(x), \dots, h_L(x)] \quad (3)$$

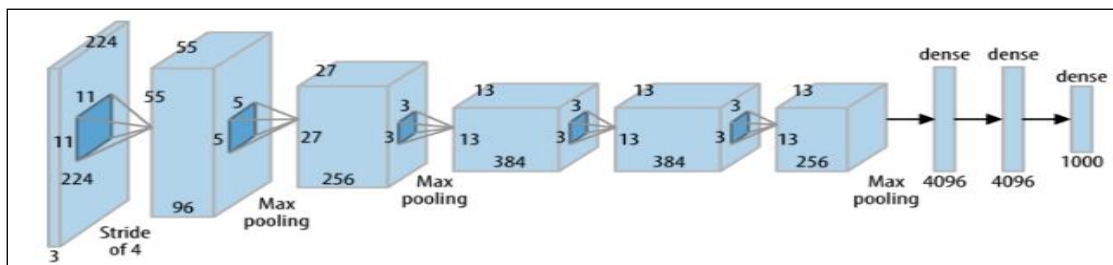


Figure 3. AlexNet framework used for feature extraction

To determine the ELM's target vector, the output hidden layer in Eq. (3) is redefined in Eq. (4)

$$H = \begin{bmatrix} h(x_1) \\ h(x_2) \\ \dots \\ h(x_N) \end{bmatrix} \quad (4)$$

Eq. (5) represents the minimal non-linear least square method used by the ELM.

$$\beta' = H^* O = H^T (HH^T)^{-1} O \quad (5)$$

where,  $H^*$  is the Moore–Penrose generalized inverse and the above Eq. (5) can also be rewritten as

$$\beta' = H^T \left( \frac{1}{C} HH^T \right)^{-1} O \quad (6)$$

Hence, the output function can be defined in Eq. (7)

$$f_L(x) = h(x)\beta = h(x)H^T \left( \frac{1}{C} HH^T \right)^{-1} O \quad (7)$$

The different grades of images are classified based on the mathematical Eq. (7). At the output layer, sigmoid function is used to classify DME images. The ELM parameters for the HET-EYE-NETS architecture are shown in Table 3.

Table 3. ELM parameters

Parameters	Setting
#Input neurons	#Number of features
# hidden layers	1
Hidden weights	-50 (lower bound) 50 (upper bond)
Hidden biases	-50 (lower bound) 50 (upper bond)
#Output neurons	#Number of classes
Activation function	Sigmoid (output layer) ReLU (hidden layers)

The abovementioned hyperparameters are applied to train the ELM model to achieve optimized prediction of DME. The output matrix  $H$  of hidden layer's is generated using the randomly selected biases and weights, along with the activation functions. Also, the  $\beta'$  matrix is computed using the training data. Finally,  $T = H\beta'$  gives the classification of DME.



### 3. RESULTS AND DISCUSSION

To evaluate the HET-EYE-NETS’s performance, standard performance measures have been employed. The performance metrics such as precision, recall, F1-score, specificity, and accuracy are calculated using the mathematical expression as presented in Table 4.

**Table 4.** Computation of performance metrics

Performance Metrics	Mathematical Expression
Sensitivity or recall	$\frac{TP}{TP + FN}$
Precision	$\frac{TN}{TP + FP}$
F1-Score	$\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
Specificity	$\frac{TN}{TN + FP}$
Accuracy	$\frac{TN + TP}{TN + FP + TP + FN}$

In the above Table 4, TN is True Negative values, TP is True Positive values, FN is False Negative values and FP is False Positive. Within the scope of this investigation, 5-fold cross-validation is used to generalize the HET-EYE-NETS model’s performance and to evaluate the classification measures. Both datasets are partitioned into five equal-sized sets while ensuring that each set represents the independent data by preserving random seeds across the iteration. Four partitions are employed for training, and for testing, the remaining partition is used. These five steps are iterated for

two datasets for which average classification performance is evaluated.

Due to class imbalance in both datasets, data augmentation is utilized to balance the number of images in each category. As the images in Grade-1 category is less than 100, the application of data augmentation with different rotation and flipping processes increases to 600 images per category. The complete algorithm was implemented using Tensor flow 2.1 backend with Keras libraries, which runs on the PC with i9 CPU, 16 GB RAM which operates at 3.4 GHZ, and NVIDIA TITAN GPU. This study is based on the three-stage HET-EYE-NETS, which uses the ELMs as a key classification mechanism. It is found that the HET-EYE-NETS model has uniform performance in classifying DME images into different grades of severity. Tables 5 and 6 show the performance metrics calculated for the HET-EYE-NETS model using different datasets. Figure 4 shows the accuracy and loss curves of the U-Net framework used for the macular segmentation for IDRiD dataset.

To prove the excellence of HET-EYE-NETS system, their performances are compared with different learning models using different datasets. Tables 7-12 present the comparative analysis between HET-EYE-NETS system and existing algorithms using IDRiD and MESSIDOR datasets to classify the severities of DME images.

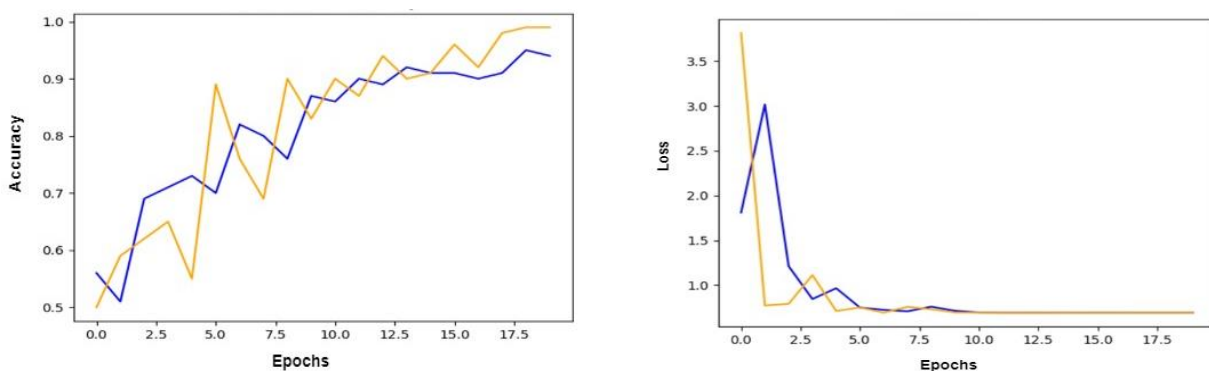
Tables 7-12 show that the proposed algorithm has shown uniform and high performances when handling the different datasets. Also, it has been proved that HET-EYE-NETS system has outperformed other existing algorithms for classifying severity levels of DME. The inclusion of data preprocessing techniques and three-stage working mechanisms have made the proposed algorithm exhibit superior performances from the other algorithms even greater than the hybrid learning mode –DMENETS.

**Table 5.** Performance metrics of the HET-EYE-NETS model on the IDRiD dataset sets

Classification Mode	Performance Analysis				
	Accuracy	Precision	Recall	Specificity	F1-Score
Normal	0.991	0.992	0.991	0.992	0.991
Grade-1	0.992	0.9923	0.992	0.9923	0.992
Grade-2	0.9912	0.9937	0.9912	0.9924	0.9924

**Table 6.** Performance metrics of the HET-EYE-NETS model on the MESSIDOR datasets

Classification Mode	Performance Analysis				
	Accuracy	Precision	Recall	Specificity	F1-Score
Normal	0.9920	0.9930	0.9929	0.994	0.9922
Grade-1	0.9918	0.9927	0.9928	0.9936	0.9926
Grade-2	0.9921	0.9928	0.9920	0.9941	0.9924



**Figure 4.** Macular segmentation’s accuracy and loss curves of the U-Net frameworks

**Table 7.** Performance comparison of different algorithms using IDRiD datasets for Grade 0 detection

Algorithms	Performance Analysis				
	Accuracy	Precision	Recall	Specificity	F1-Score
ResNETS-50	0.89	0.873	0.867	0.873	0.869
VGGnets-19	0.91	0.92	0.91	0.90	0.91
U-Nets	0.85	0.856	0.84	0.834	0.84
DenseNETS	0.82	0.83	0.821	0.85	0.823
SqueezeNETS	0.83	0.84	0.789	0.801	0.82
GoogleNETS	0.867	0.863	0.864	0.872	0.865
AlexNETS	0.867	0.873	0.828	0.865	0.857
Conventional CNN	0.80	0.789	0.782	0.778	0.790
DMENETS	0.95	0.954	0.955	0.9567	0.9598
HET-EYE-NETS	0.9912	0.992	0.991	0.992	0.9919

**Table 8.** Performance comparison of different algorithms using IDRiD datasets for Grade 1 detection

Algorithms	Performance Analysis				
	Accuracy	Precision	Recall	Specificity	F1-Score
ResNETS-50	0.89	0.873	0.867	0.873	0.869
VGGnets-19	0.921	0.91	0.90	0.90	0.903
U-Nets	0.862	0.820	0.84	0.82	0.820
DenseNETS	0.800	0.821	0.800	0.802	0.801
SqueezeNETS	0.80	0.84	0.789	0.801	0.820
GoogleNETS	0.780	0.863	0.864	0.872	0.792
AlexNETS	0.899	0.873	0.867	0.882	0.88
Conventional CNN	0.79	0.789	0.780	0.77	0.785
DMENETS	0.953	0.963	0.962	0.9578	0.967
HET-EYE-NETS	0.9912	0.992	0.991	0.992	0.9919

**Table 9.** Performance comparison of different algorithms using IDRiD datasets for Grade 2 detection

Algorithms	Performance Analysis				
	Accuracy	Precision	Recall	Specificity	F1-Score
ResNETS-50	0.884	0.872	0.870	0.873	0.870
VGGnets-19	0.918	0.919	0.91	0.90	0.915
U-Nets	0.872	0.863	0.84	0.834	0.832
DenseNETS	0.80	0.828	0.821	0.85	0.83
SqueezeNETS	0.823	0.845	0.789	0.801	0.799
GoogleNETS	0.878	0.822	0.864	0.872	0.8734
AlexNETS	0.842	0.819	0.828	0.865	0.834
Conventional CNN	0.782	0.778	0.782	0.778	0.785
DMENETS	0.960	0.972	0.955	0.9567	0.9689
HET-EYE-NETS	0.9912	0.992	0.991	0.992	0.9919

**Table 10.** Performance comparison of different algorithms using MESSIDOR images for normal detection

Algorithms	Performance Analysis				
	Accuracy	Precision	Recall	Specificity	F1-Score
ResNETS-50	0.82	0.834	0.842	0.812	0.802
VGGnets-19	0.89	0.845	0.888	0.884	0.866
U-Nets	0.823	0.856	0.832	0.810	0.820
DenseNETS	0.759	0.7679	0.789	0.745	0.723
SqueezeNETS	0.803	0.812	0.82	0.802	0.806
GoogleNETS	0.845	0.8456	0.878	0.854	0.893
AlexNETS	0.822	0.873	0.828	0.865	0.857
Conventional CNN	0.745	0.789	0.782	0.778	0.770
DMENETS	0.9566	0.954	0.9534	0.9543	0.9523
HET-EYE-NETS	0.9912	0.992	0.991	0.992	0.9919

**Table 11.** Performance comparison of different algorithms using MESSIDOR images for Grade 1 detection

Algorithms	Performance Analysis				
	Accuracy	Precision	Recall	Specificity	F1-Score
ResNETS-50	0.82	0.834	0.842	0.812	0.802
VGGnets-19	0.89	0.845	0.888	0.884	0.866
U-Nets	0.823	0.856	0.832	0.810	0.820
DenseNETS	0.759	0.7679	0.789	0.745	0.723
SqueezeNETS	0.803	0.812	0.82	0.802	0.806

GoogleNETS	0.845	0.8456	0.878	0.854	0.893
AlexNETS	0.822	0.873	0.828	0.865	0.857
Conventional CNN	0.745	0.789	0.782	0.778	0.770
DMENETS	0.9566	0.954	0.9534	0.9543	0.9523
HET-EYE-NETS	0.9912	0.992	0.991	0.992	0.9919

**Table 12.** Performance comparison of different algorithms using MESSIDOR images for Grade 2 detection

Algorithms	Performance Analysis				
	Accuracy	Precision	Recall	Specificity	F1-Score
ResNETS-50	0.82	0.834	0.842	0.812	0.802
VGGnets-19	0.89	0.845	0.888	0.884	0.866
U-Nets	0.823	0.856	0.832	0.810	0.820
DenseNETS	0.759	0.7679	0.789	0.745	0.723
SqueezeNETS	0.803	0.812	0.82	0.802	0.806
GoogleNETS	0.845	0.8456	0.878	0.854	0.893
AlexNETS	0.822	0.873	0.828	0.865	0.857
Conventional CNN	0.745	0.789	0.782	0.778	0.770
DMENETS	0.9566	0.954	0.9534	0.9543	0.9523
HET-EYE-NETS	0.9912	0.992	0.991	0.992	0.9919

**4. CONCLUSION AND FUTURE SCOPE**

A HET-EYE-NETS system is developed for screening DME using optoelectronic retinal images in this research. It consists of a three-stage pipeline system to achieve high performance. The novel ensemble transfer learning mechanism is used for effective segmentation and extracting features maps. To achieve the highest accuracy, ELM-based classification layers detect different severity grading. The proposed algorithm overcomes the overfitting problems and fairly solves the classification problems due to the different imaging systems. Extensive experimentation is conducted on using IDRiD and MESSIDOR database images and compared with existing learning models. Results demonstrate that the HET-EYE-NETS system has performed better than existing learning models and proves resilient to various noisy images and resolutions independent of any image acquisition in terms of precision, recall, F1 score, specificity and accuracy. The HET-EYE-NETS differs from other architectures by combining two architectures to extract features by AlexNet from the segmented region by U-Net architecture.

In the future, the proposed methodology could be verified on real-time datasets obtained from medical institutions, and the interpretability of the proposed model needs improvisation to ensure an accurate screening of DME from raw optoelectronic retinal images. There are many challenges such as data privacy and security, data standardization, interoperability, data quality, and ethical considerations associated when using real-time datasets. Also, it is necessary to address the following challenges: secure data sharing protocols, data governance and standards, interoperability standards and data quality assurance. The following strategies can be implemented to improve model interpretability. Feature engineering, and decision boundaries can provide a more intuitive understanding of the HET-EYE-NETS's behavior.

**REFERENCES**

[1] Ciulla, T.A., Amador, A.G., Zinman, B. (2003). Diabetic retinopathy and diabetic macular edema: Pathophysiology, screening, and novel therapies. *Diabetes Care*, 26(9): 2653-2664. <https://doi.org/10.2337/diacare.26.9.2653>

[2] Alagirisamy, M. (2021). Micro statistical descriptors for glaucoma diagnosis using neural networks. *International Journal of Advances in Signal and Image Sciences*, 7(1): 1-10. <https://doi.org/10.29284/ijasis.7.1.2021.1-10>

[3] Zhang, X.W., Thibault, G., Decencière, E., Marcotegui, B., Laÿ, B., Danno, R., Cazuguel, G., Quelled, G., Lamard, M., Massin, P., Chabouis, A., Victor, Z., Erginay, A. (2014). Exudate detection in color retinal images for mass screening of diabetic retinopathy. *Medical Image Analysis*, 18(7): 1026-1043. <https://doi.org/10.1016/j.media.2014.05.004>

[4] Zheng, Y.F., He, M.G., Congdon, N. (2012). The worldwide epidemic of diabetic retinopathy. *Indian Journal of Ophthalmology*, 60(5): 428-431. <https://doi.org/10.4103/0301-4738.100542>

[5] Sivaprasad, S., Oyetunde, S. (2016). Impact of injection therapy on retinal patients with diabetic macular edema or retinal vein occlusion. *Clinical Ophthalmology*, 10(2016): 939-946. <https://doi.org/10.2147/OPHTH.S100168>

[6] Davidson, J.A., Ciulla, T.A., McGill, J.B., Kles, K.A., Anderson, P.W. (2007). How the diabetic eye loses vision. *Endocrine*, 32(1):107-116. <https://doi.org/10.1007/s12020-007-0040-9>

[7] Wilkinson, C.P., Ferris III, F.L., Klein, R.E., Lee, P.P., Agardh, C.D., Davis, M., Verdager, J.T. (2003). Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*, 110(9): 1677-1682. [https://doi.org/10.1016/S0161-6420\(03\)00475-5](https://doi.org/10.1016/S0161-6420(03)00475-5)

[8] Mookiah, M.R.K., Acharya, U.R., Chua, C.K., Lim, C.M., Ng, E.Y.K., Laude, A. (2013). Computer-aided diagnosis of diabetic retinopathy: A review. *Computers in Biology and Medicine*, 43(12): 2136-2155. <https://doi.org/10.1016/j.combiomed.2013.10.007>

[9] De Souza, N., Cui, Y., Looi, S., Paudel, P., Shinde, L., Kumar, K., Berwal, R., Wadhwa, R., Daniel, V., Flanagan, J., Holden, B. (2012). The role of optometrists in India: An integral part of an eye health team. *Indian Journal of Ophthalmology*, 60(5): 401-405. <https://doi.org/10.4103/0301-4738.100534>

[10] Thomas, R., Paul, P., Rao, G.N., Muliylil, J.P., Mathai, A. (2005). Present status of eye care in India. *Survey of Ophthalmology*, 50(1): 85-101.

- <https://doi.org/10.1016/j.survophthal.2004.10.008>
- [11] Deepak, K.S., Sivaswamy, J. (2012). Automatic assessment of macular edema from color retinal images. *IEEE Transactions on Medical Imaging*, 31(3): 766-776. <https://doi.org/10.1109/TMI.2011.2178856>
- [12] Lee, C.S., Baughman, D.M. Lee, A.Y. (2017). Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. *Ophthalmol Retina*, 1(4): 322-327. <https://doi.org/10.1016/j.oret.2016.12.009>
- [13] Perdomo, O., Otalora, S., Rodríguez, F., Arevalo, J., González, F.A. (2016). A novel machine learning model based on exudate localization to detect diabetic macular edema. *Lecture Notes in Computer Science*, 1: 137-144. <https://doi.org/10.17077/omia.1057>
- [14] He, J., Shen, L.L., Ai, X.F., Li, X.C. (2019). Diabetic retinopathy grade and macular edema risk classification using convolutional neural networks. In 2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, pp. 463-466. <https://doi.org/10.1109/ICPICS47731.2019.8942426>
- [15] Shaikh, M., Kollerathu, V.A., Krishnamurthi, G. (2019). Recurrent attention mechanism networks for enhanced classification of biomedical images. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, pp. 1260-1264. <https://doi.org/10.1109/ISBI.2019.8759214>
- [16] Li, X.M., Hu, X.W., Yu, L.Q., Zhu, L., Fu, C.W., Heng, P.A. (2019). CANet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. *IEEE Transactions on Medical Imaging*, 39(5): 1483-1493. <https://doi.org/10.1109/TMI.2019.2951844>
- [17] Nazir, T., Nawaz, M., Rashid, J., Mahum, R., Masood, M., Mehmood, A., Ali, F., Kim, J., Kwon, H.Y., Hussain, A. (2021). Detection of diabetic eye disease from retinal images using a deep learning based Centernet model. *Sensors*, 21(16): 5283. <https://doi.org/10.3390/s21165283>
- [18] Hassan, B., Hassan, T., Li, B., Ahmed, R., Hassan, O. (2019). Deep ensemble learning based objective grading of macular edema by extracting clinically significant findings from fused retinal imaging modalities. *Sensors*, 19(13): 2970. <https://doi.org/10.3390/s19132970>
- [19] 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>
- [20] Abràmoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C., Niemeijer, M. (2016). Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative Ophthalmology & Visual Science*, 57(13): 5200-5206. <https://doi.org/10.1167/iovs.16-19964>
- [21] Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabudhe, V., Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (IDRID): A database for diabetic retinopathy screening research. *Data*, 3(3): 25. <https://doi.org/10.3390/data3030025>
- [22] Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., Charton, B., Klein, J.C. (2014). Feedback on a publicly distributed image database: The Messidor database. *Image Analysis and Stereology*, 33(3): 231-234. <https://doi.org/10.5566/ias.1155>
- [23] Bichri, H., Chergui, A., Hain, M. (2023). Image classification with transfer learning using a custom dataset: Comparative study. *Procedia Computer Science*, 220(1): 48-54. <https://doi.org/10.1016/j.procs.2023.03.009>
- [24] Hosny, K.M., Kassem, M.A., Foad, M.M. (2018). Skin cancer classification using deep learning and transfer learning. In 2018 9th Cairo International Biomedical Engineering Conference (CIBEC), Cairo, Egypt, pp. 90-93. <https://doi.org/10.1109/CIBEC.2018.8641762>
- [25] Veni, N., Manjula, J. (2022). Modified visual geometric group architecture for MRI brain image classification. *Computer Systems Science and Engineering*, 42(2): 825-835. <https://doi.org/10.32604/csse.2022.022318>
- [26] Sajja, T.K., Devarapalli, R.M., Kalluri, H.K. (2019). Lung cancer detection based on CT scan images by using deep transfer learning. *Traitement du Signal*, 36(4): 339-344. <https://doi.org/10.18280/ts.360406>
- [27] Huang, G.B., Zhu, Q.Y., Siew, C.K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3): 489-501. <https://doi.org/10.1016/j.neucom.2005.12.126>
- [28] Wang, B.T., Huang, S., Qiu, J.H., Liu, Y., Wang, G.R. (2015). Parallel online sequential extreme learning machine based on MapReduce. *Neurocomputing*, 149: 224-232. <https://doi.org/10.1016/j.neucom.2014.03.076>