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

Traitement du Signal
Vol. 40, No. 3, June, 2023, pp. 939-948
Journal homepage: http://iieta.org/journals/ts

An Approach to Classify and Segment Diabetic Retinopathy and Retinopathy of Prematurity

Madduri Vamsi Krishna, Battula Srinivasa Rao*

School of Computer Science and Engineering, VIT-AP University, Andhra Pradesh 522237, Amaravati, India

Corresponding Author Email: srinivas.battula@vitap.ac.in

https://doi.org/10.18280/ts.400310

Received: 5 January 2023
Accepted: 8 March 2023

Keywords:
retinal disease, DCGAN, deep learning, DeepLabv3+, diabetic retinopathy, retinopathy of prematurity

ABSTRACT

In recent years, retinal disorders have grown to be a serious public health issue. Retinopathy of Prematurity (ROP) and Diabetic Retinopathy (DR) are the foremost factors of vision impairments in children and youngsters correspondingly. These illnesses develop gradually and have no visible symptoms. To avoid vision damage, it is crucial to identify these conditions quickly and receive the appropriate medication. Therefore, a completely automated approach for identifying retinal disorders is needed. It is designed to reduce human contact for the identification of Diabetic Retinopathy (DR) and Retinopathy of Prematurity (ROP) while maintaining the excellent accuracy of the classification. This paper presents an enhanced deep learning model LeNet-5 for retinal disease categorization framework. To achieve the desired findings, the DeepLabv3+ based blood vessel segmentation is carried out. After segmenting the retinal vessels, the features relevant to DR and ROP are extracted using dual channel based Capsule Network (CapsNet). After that, LeNet-5 receives the CapsNet feature map for categorization. To increase the deep learning classifier's performance, the Deep Convolutional Generative Adversarial Network (DCGAN) based data augmentation technique is implemented. The system evaluated in MESSIDOR and private datasets obtained 99.29% and 99.12% accuracy for DR and ROP classification. When the attained results are compared with other existing techniques, it is seen that more successful findings are achieved.

1. INTRODUCTION

Retinal-based disorders can cause blurry vision and occasionally total blindness, are growing more prevalent among people of all ages and diminish the eye's capacity to see clearly [1-3]. Analyzing lesions and retinal structures including the blood vessels, fovea, macular region, optic cup, and Optic Disc (OD) can reveal important information about some of the main retinal illnesses and help with diagnosis. Some of the major eye conditions that can lead to blindness if not properly diagnosed are cataract, Retinopathy of Prematurity (ROP), Diabetic Macular Edema (DME), Age-related Macular Degeneration (AMD), Diabetic Retinopathy (DR), and glaucoma [4-6]. Such retinal disorders are typically detected by screening procedures that require considerable skill and professional attention.

DR is caused by the negative effects of diabetes on the eyes, while ROP is a defective blood vessel growth in the retina of premature babies or low birth weight babies. These disorders should be detected early on because their symptoms can be minor and create serious issues. Vision loss could happen if it is not addressed quickly. Only qualified ophthalmologists are capable of conducting the testing procedures needed to diagnose these disorders [7-10]. They examine and assess the retinal images that were collected to identify any anomalies. Even for experienced ophthalmologists, the procedure may take time. In the meantime, the issue of delayed disease identification may be made worse by the increased amount of patients with eye problems [11-14]. In contrast, the computer-aided automatic diagnosis model demonstrates considerable aptitude in resolving the aforementioned shortcomings, and a number of solutions have been put out in this field where deep learning appears to offer impressive outcomes [15, 16]. It has been demonstrated that deep learning algorithms perform better than traditional methods. It has outperformed all traditional image analysis techniques in a variety of computer vision and image analysis applications. To build automated computer-aided methods with applications in numerous fields, a number of deep-learning algorithms have been created to evaluate medical images [17-20]. The diagnosis of different eye illnesses like DR and ROP is one of the areas where we can use these algorithms.

Therefore, in this work, the effective deep learning-based technique LeNet-5 is implemented to classify Diabetic Retinopathy (DR) and Retinopathy of Prematurity (ROP). LeNet's design was quite straightforward; it only had 5 layers but produced results that were more accurate. To perform the classification, segmentation of blood vessels is an important concept. For this, DeepLabv3+ based technique is applied to segment the blood vessels from the retinal fundus images. It attains improved semantic segmentation accuracy, significantly lowers the cost of network training and processing complexity. Then, the features of the images are extracted by the dual-channel CapsNet. In order to address the drawbacks of CNN brought on by pooling, transformation matrices are employed in CapsNet to encode the inherent spatial relationship between a part and a whole of an image. These concepts have strengthened the CapsNet. To increase

the classifier's performance and to avoid overfitting, the dataset images are increased by Deep Convolutional Generative Adversarial Network (DCGAN).

• To achieve DR grading and ROP classification, LeNet-5 based effective deep learning technique structure is applied with the dual channel CapsNet based feature extractor.

• To achieve effective classification performance, DeepLabV3+-based segmentation technique is implemented.

• To train the deep learning model efficiently and increase the dataset images, Deep Convolutional Generative Adversarial Network (DCGAN) based data augmentation technique is applied.

• The proposed approach exhibits better results than the existing techniques for the classification of Diabetic Retinopathy and Retinopathy of Prematurity on the MESSIDOR and private datasets.

The remaining portion of this paper is organized as follows. In section 2, the existing techniques are reviewed. Section 3 illustrates the proposed methodology, while the results and discussion are presented in Section 4. Section 5 illustrates the conclusion and future works.

2. LITERATURE REVIEW

Many image classification algorithms were created in previous decades to assess the severity of DR and ROP using fundus pictures. Below is a detailed review of a few of them.

Using retinal fundus pictures, numerous researchers have worked on DR segmentation and detection. For instance, Maiti et al. [21] offered a deep learning method utilizing a modified neural network for autonomously identifying and segmenting the optic disc from fundus pictures. The VGG16 framework was utilized as an encoder in this study, and the decoder was built with a symmetric structure to the encoder for improved object segmentation. Yet, the vanishing gradient problem is one of the main issues with this network. That lessens the effectiveness of this strategy.

Abdelmaksoud et al. [22] recommended the conventional and DL approaches to accurately detect DR grades automatically using various colour fundus pictures. To begin with, they used a number of pre-processing methods to boost the image's quality. The segmentation process was then implemented using a customized U-Net DL model. Gray Level Run Length Matrix (GLRLM) was then used to extract the key characteristics from the image, which were then fed into the Support Vector Machine (SVM) classifier. Yet, low sensitivity (86%) attained by this network indicates the detection of inaccurate solutions.

By combining the benefits of cell membrane structures, Xue et al. [23] developed a dynamic membrane structure with hybrid architectures. For pixel-wise segmentations of Optic Disk (OD), Exudates (EXs), and Micro aneurysms (Mas) in DR, they designed a Mask R-CNN within every new membrane structure. Yet, the lesions' visualization is of low quality.

To identify zones I, II, and III with segmentation of the optic disc and blood vessels, Agrawal et al. [24] suggested a method using an ensemble of "U-Network" and "Circle Hough Transform". The created model is general and was practiced on mixed photos of various sizes. With photos of varying sizes taken by two separate imaging systems, it identifies zones. Nevertheless, this article does not report on identifying zones I and II in ROP pictures with and without macula.

Nisha et al. [25] presented the computer-based analysis system for the objective evaluation of severe illness in ROP. The suggested system starts out with a preliminary segmentation phase based on adaptive filtering, connection analysis, and picture fusion. The research suggests the use of new retinal properties, such as vessel density and leaf node count, to supplement the regularly utilized features, such as tortuosity and breadth. It depicts the irregular growth of the blood vessels. However, this approach focuses on applying image processing techniques to calculate only one or two kinds of characteristics. Henceforth, it fails to extract further crucial features.

A basic model of OD segmentation built around the attention gate is presented by Abaei Kashan et al. [26]. At the beginning, the photos were gathered, analyzed, and then fed into a brand-new deep convolutional neural network composed of attention-in-skip connections. A two-stage convolutional network makes up the architecture. In the initial phase, the raw image and image features are divided into two separate branches, each of which produces a different output. To move into the post-processing stage and find the area connected to the OD, the outputs were concatenated. The overall performance of this effort is strong. However, it results in a high number of false positives and inaccurate boundary tracing around the optic disc area.

Cao et al. [27] created the unified weakly supervised domain adaptation system with three parts. They are attention based multi-instance learning, instance progressive discriminator, and domain adaptation. An attention mechanism and a multi-instance learning strategy were employed to model the relationship between the patches and images in the target domain. In the meantime, it uses a jointly learning technique that combines all knowledge that is accessible from the source and target domains. Using the massive EyePACS dataset and the Messidor dataset, they evaluated how well the suggested paradigm for DR grading performed. However, the performance of the models is substandard. Because they had not suggested any technique to handle the imbalanced data problem.

Kalyani et al. [28] designed a modified capsule network for the identification and categorization of diabetic retinopathy. The features were first retrieved from the fundus images using convolution and the primary capsule layer, and the class capsule layer and softmax layer were subsequently employed to assess the images belonging to a particular class. They did not employ any segmentation techniques. For feature extraction, the entire retinal pictures were used. As a result, undesirable traits are also extracted.

A quadrant ensemble automatic DR grading method using the InceptionResnet-V2 based deep network was proposed by Bhardwaj et al. [29]. The data augmentation stage in the developed framework was combined with quadrant cropping, optical disc localization, and histogram equalization to enhance network performance. However, it did not manifest the neovascularization of the optic disc.

For ROP detection, Peng et al. [30] presented the three-stream parallel framework including EfficientNetB2, DenseNet121 and ResNet18 for feature extraction. It gathered rich and varied high-level features. Afterwards, the features from three streams were deeply combined by concatenation and convolution process. An ordinal classification technique was finally used in the classification stage. But this framework is computationally expensive due to the three feature extractors.

3. PROPOSED METHODOLOGY

The developed structure includes of five stages such as image pre-processing, image augmentation, vessel segmentation, extraction of important features, and classification. Initially, the input retinal images are obtained from the dataset and then the preprocessing and data augmentation tasks are applied on them. Afterward, segmentation of the optic disc and vessels in the retina is done. Then, feature extraction using dual channel CapsNet is performed. In this network, one channel detects the vessel centerlines and the vascular tree structure and then extracts characteristics for ROP classification based on the tortuosity and dilatation of the vessels and another channel extracts the features from the blood vessels for DR classification. At last, the combined features are sent to the classifier to perform the classification. Figure 1 depicts the proposed framework's overall structure.

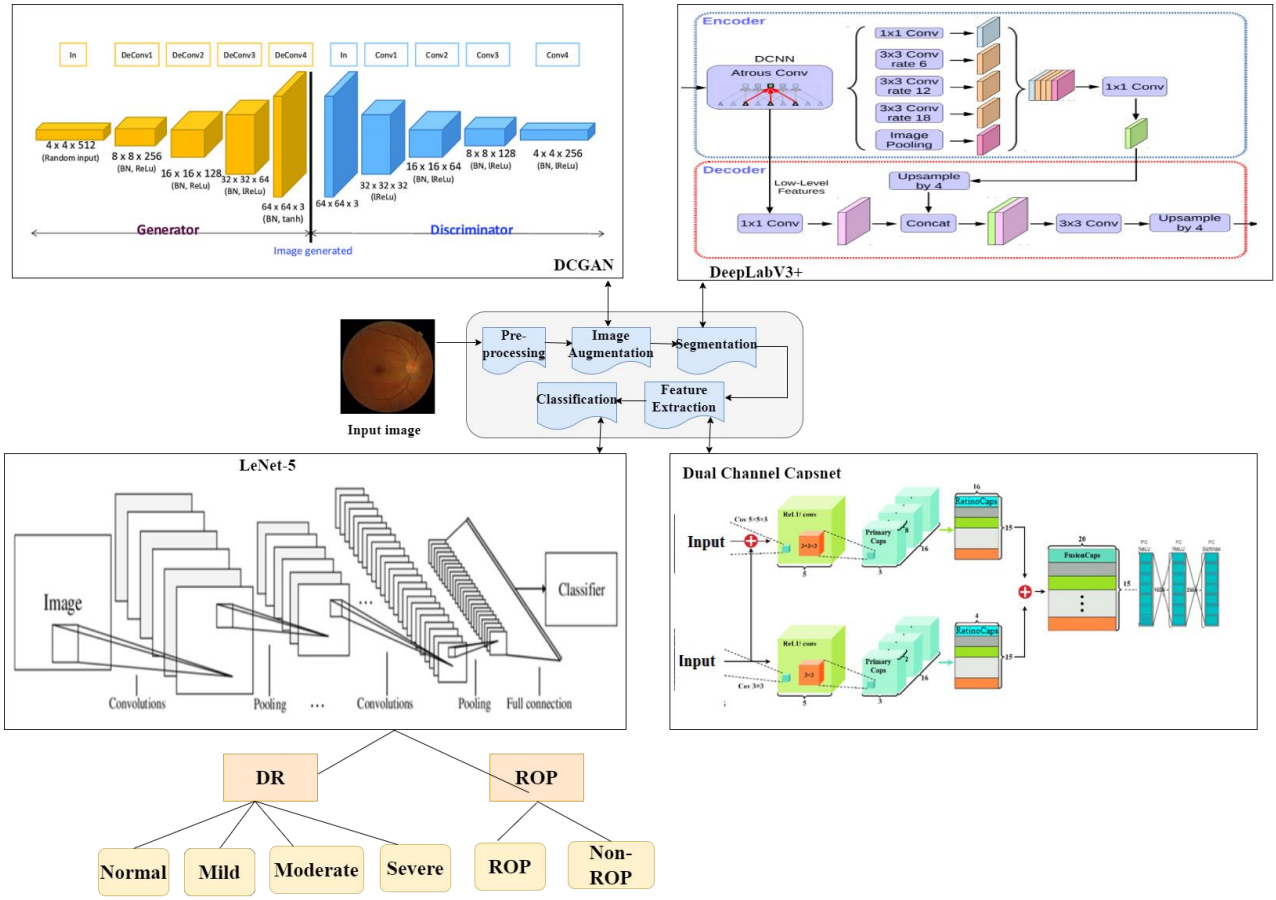

Figure 1. System architecture of the proposed system

3.1 Pre-processing

Preprocessing an image is crucial for enhancing input image's quality because poor quality images may affect the overall network performance. Moreover, it is critical to make sure that all the images are uniform and have their attributes upgraded. Every fundus image in the dataset has been scaled to 256×256 pixels and transformed to a greyscale format. Then the contrast of the fundus images is improved by Contrast Limited Adaptive Histogram Equalization (CLAHE). Gaussian filter is then applied to the images to remove the noise.

3.2 Data augmentation using DCGAN

An overview of the GAN-based image generation process is provided in this section. Deep Convolutional Generative Adversarial Network (DCGAN) is employed for this purpose. In this network, two deep convolutional neural networks are used as the generator and discriminator. Four convolutional layers are included in the generator phase. 8 kernels are included in the first three layers and one kernel is included in the last layer. The size of all the kernels in the generator is 3×3 and a stride of 1×1. The Discriminator component has four convolutional layers with 8, 16, 32, and 64 kernels with 2×2 kernel size and 2×2 stride. A convolutional strided layer was introduced to replace the max pooling layer. The outputs of the convolutional layers are concatenated by the flattening process, producing a vector with 3000 dimensions. The generator is created to generate images, to trick the discriminator. It requires a randomized input, such as 100-element random vector with 100 entries distributed randomly between 0.0 and 1.0 range. The generator receives this randomized input and creates the image using two-dimensional transposed convolution. Furthermore, transposed convolutional layers are used in the network to perform up sampling. Batch Normalization (BN) and LeakyReLU (LReLU) activation are used to further process the output of each transposed convolution layer except the last layer. After the convolutional layers, Batch Normalization aids in controlling the extracted features, while the LeakyRELU activation function protects against the vanishing gradient problem.

LeakyReLU activation function eliminates negative outputs while directly mapping positive inputs to outputs defined in

$F(x)$=max(0, x). Similar in operation, LeakyReLU only reduces negative values by a factor of 0.01 rather than eliminating them entirely. LeakyReLU's definition can be found in the following equation.

$$F(x)=\begin{cases} x, x \in \Re_0^+ \\ 0.01 \bullet x \in \Re^- \end{cases} \quad (1)$$

where, $\Re_0^+$, $\Re^-$ stands for Positive and Negative valued output ranges respectively.

The discriminator then uses the created image to determine if it is a part of the original dataset or not. In other words, it distinguishes between real images and false (made) ones. The generator's settings are changed, and the procedure is repeated if the discriminator concludes that the created image is fraudulent. In this manner, the generator makes continuous adjustments to its settings until the images it produces are undetectable to the discriminator from the images of the original dataset.

3.3 DeepLabv3+ based segmentation

The DeepLabV3+ was created based on the DeepLabV3 framework of the previous generation. It utilized atrous convolution for feature extraction of the image at any resolution and enhancing the Atrous Spatial Pyramid Pooling (ASPP) module which aims to detect convolutional features at multiple scales by using various atrous convolution rates. The encoder and decoder are two separate components of this network. The feature map's dimensionality is decreased and features are extracted by the encoder. To acquire the final semantic segmentation results, the decoder is mostly utilized which recovers feature map's resolution and edge information. Initially, the input image from DR Dataset is given to the encoder for vessel segmentation. After the obtained vessel segmentation map from the decoder part, the ROP image is given to the encoder section for further process. In this segmentation, the optic disc and vessels are segmented from the image.

The final few convolutional layers of the encoder's convolutional operation are substituted with the whole convolution to expand the receptive area while maintaining the feature map's resolution. To gather multi-scale semantic contextual data, the ASPP module included in DeepLabv3+ employs dilation convolution at various speeds. By using variable sampling rates, this approach resamples feature maps produced by the encoder. Next, the outputs are combined after a parallel convolution filter is applied to the feature maps at various atrous rates. The ASPP module consists of a max-pooling layer running in parallel with 3×3 convolutions with dilation rates of 6, 12, and 18 were employed. Batch normalization is also present to normalize data after each operation. In order to compress and integrate the acquired feature maps in every branch, a 1×1 convolution is utilized in the final step.

The decoding portion of the framework has a straightforward structure. A high-dimensional feature map with 256 channels is the ultimate result of the network's encoding section which is forwarded to the decoder part. Then the up sampling process is conducted using bilinear interpolation and the low-level features are concatenated with the same resolution. The network utilizes 1×1 convolution of the low-level features to minimize channels before the feature concatenation. The concatenated features are refined using a 3×3 convolution, and the final result is obtained by restoring the concatenated features to the size of the original picture using bilinear interpolation based up sampling process (4 times). DeepLabv3+ produces appropriate semantic segmentation results by utilizing these novel components.

3.4 Dual CapsNet based feature extraction

Dual channel based Capsule Network (CapsNet) is applied for the feature extraction of segmented images. Through two feature extraction modules, this network extracts the deep features from the segmented retinal pictures. Contrary to standard CNN, CapsNet represents feature intensity using vectors rather than scalars, effectively enhancing the representation and utilization of available information. Two 2-D capsule network channels, fully connected deep neural network, and a fusion network are all components of this design. One channel of 2-D capsule network is used for the feature extraction of DR segmented images and another one is used for the feature extraction of ROP segmented images. All the features are combined after fusion in Fusion Caps Module, a fully connected deep neural network layer. This fusion is used as input to the reconstruction layer.

The first two layers in this architecture include 16 and 32 kernels with 5×5 size and a stride of 1. The output from the second layer is subjected to maximum pooling with a stride of two. In the third layer, 128 filters with a stride of 1 and 9×9 kernel size are present. The primary capsule (PrimaryCaps) is the fourth layer which includes32 distinct capsules. Each capsule has a filter applied with a stride of 2 and a filter size of 9×9. Every Primary capsule utilizes a convolutional process (16 convolution channels, 2 filters per convolution, and ReLU activation), with 2 convolutional units each having a 3×3 kernel and a stride of 2. Using the dynamic routing algorithm technique, the capsules in the PrimaryCaps are forwarded to the RetinoCaps layer. Every capsule's unique opinions are calculated using the trainable weights, W by using dynamic routing. Assume that the 8-dimensional primary capsule's index is $i \in [1, N_{PC}]$, 16 dimensional output capsule's index is $j \in [1, N_{class}]$ and the dimension of 8 × 16 is W_{ij}. The following equation provides the primary capsule's unique viewpoint on the output capsule's j:

$$\hat{u}_{j/i} = u_i \times W_{ij}^{DC} \quad (2)$$

here, u_i denotes the ith primary capsule. We obtain an output block of shape N_{class}×16 for each primary capsule i. Another sort of weight, known as routing weights, b, with the dimension of N_{PC}×N_{class} is taken into account for the operation of dynamic routing. They are utilized to combine many viewpoints to create the final output capsule. The different viewpoints are combined to create the RetinoCaps output using the coupling coefficients c_{ij} which is defined from the following equation.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (3)$$

Two size 16 capsules are present in the RetinoCaps, and each of them receives information from the PrimaryCaps layer. Concatenation is utilized to combine the outputs from two

RetinoCaps. Concatenating the result of two feature maps is produced by the network from two input images. The concatenated feature matrix is then supplied to the reconstruction sub-network (decoder layer), which is made up of three fully connected layers: the first two layers contain ReLU activation function and final one has softmax activation. The dimensions of these layers are 512, 1024 and 2304 respectively.

3.5 LeNet-5 based classification

The feature maps obtained from the Dual channel CapsNet are passed to the LeNet-5 network for classification. LeNet is an easy-to-use and powerful algorithm for grayscale photos. Furthermore, it decreases structural complexity and raises the likelihood of real-time processing. There are eight layers in all in the LeNet-5. The input layer, two convolution layers, one fully linked layer, two subsampling layers, and an output layer are the layers that make up this structure. In this network, the subsampling layer is the network used for pooling.

The input layer of LeNet-5 receives the feature maps for further processing. Between the convolution layer and the fully connected layer, a dropout layer is added with a dropout rate of 0.8 to prevent overfitting. The convolutional layer's kernel size is set to be 5×5 and the subsampling layer's kernel size is 2×2. To minimize parameter training, the fully connected layer decreases the amount of neurons from 120 to 84. The input feature matrix is then decreased in two dimensions by two times based on the 2×2 pool size using the feature selection technique in the subsampling layer while maintaining the primary properties of the data. Eq. (4) illustrates the expression of the pooling layer, where pool (.) denotes the maximum pooling operation. Typically, the output of the lth layer is defined as a_n^l while the output of the previous layer is written as a_n^{l-1}, where n refers the nth sample:

$$a_n^l = pool(a_n^{l-1}) \tag{4}$$

The last layer in this structure is fully connected layer. The ReLU activation function in every neuron is used to connect the neurons of preceding layer. Then the result of these neurons is sent to the output layer, which can incorporate local information with the ability to distinguish between classes. As a result, the fully connected layer plays a part in traditional classifiers. The output of fully connected lth layer is shown in Eq. (5). In that, b^l is the offset term and convolutional kernel is denoted by w^l.

$$a_n^l = f(w^l \cdot a_n^{l-1} + b^l) \tag{5}$$

The final output layer is softmax layer. This layer selects the output class which has maximum probability. This function is shown in the following equation.

$$a_n^L = o_n^v = soft\max(w^L \cdot a_n^{L-1} + b^L) \tag{6}$$

o_n^v (n=1, 2, 3, ……, N; v=1, 2, 3, ……., V) is the resulting probability of nth instance for v classes. If t_n^v indicates the predicted result probability of nth instance n v classes, then the local and global error formula of nth sample, E_n is given in the following equations.

$$E_n = \frac{1}{2}\sum_{v=1}^{V}\left\|t_n^v - o_n^v\right\|_2^2 \tag{7}$$

$$E = \sum_{n=1}^{N} E_n \tag{8}$$

Finally, the softmax layer produces the output based on the maximum probability of the classes. In the proposed work, DR classification is the multi-class classification problem (0, 1, 2, 3)=which contains four classes and ROP is the binary= classification (0, 1) problem. At last, the classifier produces= maximum probability based result.

4. RESULTS AND DISCUSSION

To analyze the effectiveness of the proposed system, numerous investigations are carried out on the DR and ROP datasets. This section presents various tests that illustrate the objective and quantitative outcomes of the suggested strategy for choosing the most instructive retinal samples, segmenting blood vessels, and categorizing DR and ROP. These experimental steps are carried out utilizing a variety of statistical metrics, which are further discussed in the following sections. The Keras framework, which is based on Tensor Flow, and the Python programming language were used to create the proposed system. All tests were conducted using a 2 GB NVIDIA GeForce 930 MX graphics card. Eighty percent of the datasets were used for training, while twenty percent were used for testing. The suggested networks are trained over 100 iterations using the Adam optimizer with a batch size of 8, and an L2 regularization of 0.0005.

4.1 Datasets description

4.1.1 Messidor dataset

The dataset comprises of 1200 retinal fundus colour images that were collected by three ophthalmologic institutions in the cities Brest, Etienne, and Paris of the country France. This is done by utilizing a colour video 3CCD (Charge-Coupled Devices) camera installed on a Topcon TRC-NW6 Non-Mydriatic Retinal Camera with a 45 degree field of view. At 2304*1536, 2240*1488, or 1440*960 pixels, images were captured in 8 bits per colour plane (R, G, and B). In this dataset, the images are categorized into four labels. They are Normal, Mild, Moderate and Severe.

4.1.2 ROP dataset

The ROP Collaboration Group consists of 30 hospitals spread across China, was responsible for collecting the ROP photos. Five professional children ophthalmologists individually labeled the images taken throughout the screening method as "Disease" and "Not Diseased." The pictures were captured by RetCam 3 or 2 by a skilled technician. This data set is available on IEEE data port. In this dataset, the images are categorized into two labels. They are disease and non-disease. Disease category includes 8244 images and 11298 images are included in non-disease category.

4.2 Segmentation outcome

To analyze the efficiency of the suggested technique for DR and ROP segmentation, the experiments are conducted on the

basis of the two tasks. In the first task, the performance of DR segmentation is analyzed. In the second task, the performance of ROP segmentation is evaluated. The performance metrics named Positive Predictive Value (PPV), Sensitivity (SEN), and Dice Similarity Coefficient (DSC) were used for this assessment. The similarity between the predicted and Ground Truth (GT) is measured using DSC. The percentage of correct positive predictions versus both correct and faulty positive predictions is known as PPV. The true positive rate is known as SEN.DSC is a quantitative indicator used to analyse the similarities between 2 sets of data.

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}, \quad PPV = \frac{TP}{TP + FP}, \quad SEN = \frac{TP}{TP + FN} \tag{9}$$

where, *TP*=True Positive, *FP*=False Positive, *FN*=False Negative are the parameters used for performance assessment in the classification of classes.

Task 1: DR segmentation

The proposed approach divides the tiny blood vessels that are most suitable for classification. The fundus image's minute details, including the branch of blood vessels, are not skipped throughout the segmentation process. The greyscale conversion has a significant impact on how well the proposed approach segments data. The image's grey scale conversion allowed for the acquisition of the minute blood vessel features. Figure 2 displays the segmentation model's visualization outcomes.

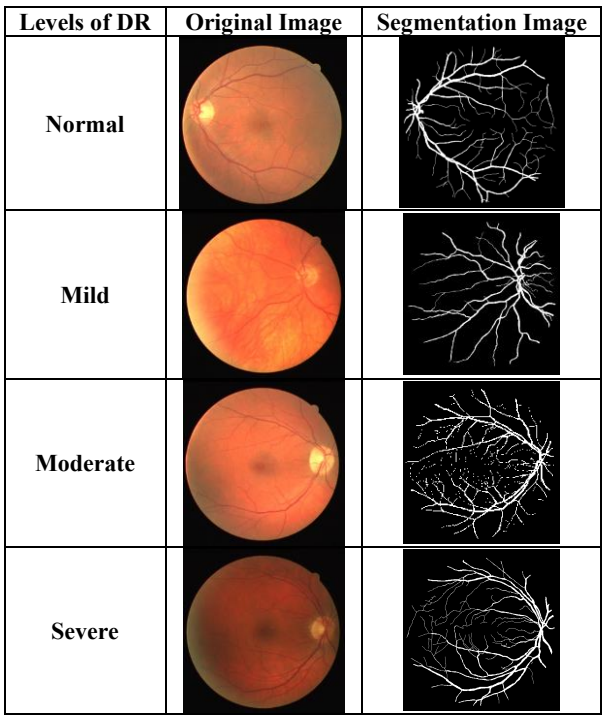

Figure 2. Segmentation results of DR

To evaluate the DR segmentation task, several traditional semantic segmentation techniques are utilized. The comparisons of the quantitative results are shown in Table 1 and the graphical representation of this table is shown in Figure 3.

Table 1. Comparison of DR segmentation results

Techniques	DSC (%)	PPV (%)	SEN (%)
VGG-16 [21]	93.73	-	96.59
U-Net [22]	82.7	80	86
Mask R-CNN [23]	96.7	-	98.1
M-Net [31]	94.63	-	96.19
Proposed	**98.56**	**98.78**	**99.23**

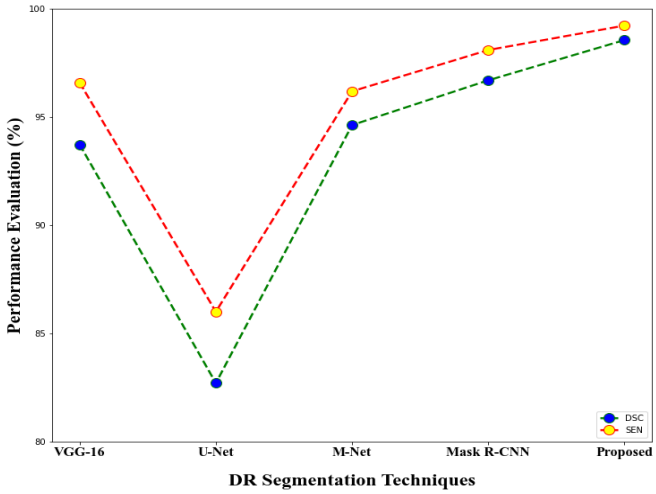

Figure 3. Comparison of SEN and DSC performance metrics

The M-Net shows better results than the traditional U-Net in terms of DSC and SEN. Moreover, the result of VGG-16 is also similar to the M-Net with 93.73% DSC and 96.59% SEN. However, these results are slightly lower than the Mask R-CNN technique. It achieves 96.7% DSC and 98.1% SEN which reveals that this technique contains a high percentage of true positive pixels. Even though these values are lower than our proposed approach which attains 98.56% DSC, 98.78% PPV, and 99.23% Sensitivity.

Furthermore, the segmentation results are refined by the suggested segmentation algorithm's efficient decoder, which also boosts feature density and enhances the receptive field. Additionally, conditional random fields can be used to increase classification accuracy. As a result, the suggested technique produces more effective segmentation results than other methods that were compared.

Task 2: ROP segmentation

In the ROP segmentation process, the blood vessels and the optic disc segmentation are performed. In this segmentation, the proposed approach segments blood vessels to their fine detail which improves the performance of the classifier to achieve the best results. Similarly, the optic disc segmentation is accomplished and the obtained results are close to the ground truth images. The quantitative results of ROP segmentation compared to other standard approaches are given in Table 2.

Table 2. Comparison of ROP segmentation results

Techniques	DSC (%)	PPV (%)	SEN (%)
U-Net [24]	84.4	-	-
Otsu Thresholding [25]	72.29	-	73.85
Attention Gate-Unet [26]	94.22	-	-
SCIRD-TS [32]	85	-	-
Proposed	**98.45**	**98.86**	**99.12**

For ROP segmentation, the performance of SCIRD-TS and Traditional U-Net is very similar and achieves 85% and 84.4% DSC values respectively. In this process, the proposed solution has a sensitivity of 99.12% and a PPV of 98.86%. It describes our method produce superior segmentation results. Compared to all the other methods, the Attention Gate-Unet provides better DSC value (94.22%). However, it is not higher than the proposed approach's DSC value (98.45%).

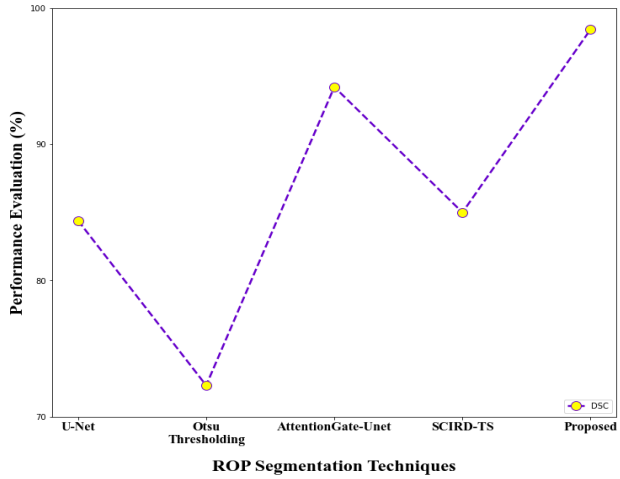

Figure 4. Comparison of DSC values in ROP segmentation

The line graph shown in Figure 4 represents the comparison of DSC values. From the image, it is observed that the performance of Otsu thresholding is very poor compared to other deep neural networks. It reveals that the performance of deep neural networks is better than traditional techniques.

4.3 Classification results

To analyze the effectiveness of the proposed approach for DR and ROP classification, the experiments are conducted on the basis of the two tasks. In the first task, the performance of DR classification process is analyzed. In the second task, the performance of ROP classification process is evaluated. The f-score, recall, specificity, precision, and accuracy of the proposed approach were computed individually for each dataset. These metrics are computed based on false negatives, false positives, true negatives, and true positives values. When a clinically-labeled image is correctly categorized, true positive is recorded, and when it is inaccurate, a false negative F2 is recorded. Similar to this, when a healthy retina is correctly categorized as abnormal, it results in a false positive and a true negative.

Task 1: DR classification

In this classification process, the DR is classified into four categories; they are Severe, Moderate, Mild and Normal. For the Messidor dataset, the f1-score, specificity, sensitivity, precision, and accuracy of the proposed system with data augmentation were estimated to be 99.33%, 99.22%, 99.30%, 99.76%, and 99.29%respectively. The DCGAN based data augmentation technique reduced overfitting and increased the performance of the classifier. The obtained quantitative results are shown in Table 3.

From the above table, it is observed that the proposed approach classified the DR images and obtained efficient results for all classes. In all the four classes, the proposed approach classified moderate class images with high accuracy (99.86%) compared to all other classes. However, in terms of other metrics, the normal class images have obtained better results with 99.91% precision, 99.67% sensitivity, 99.56% specificity and 99.89% f1-score. The graphical representation of this multi-class DR classification comparison is shown in Figure 5.

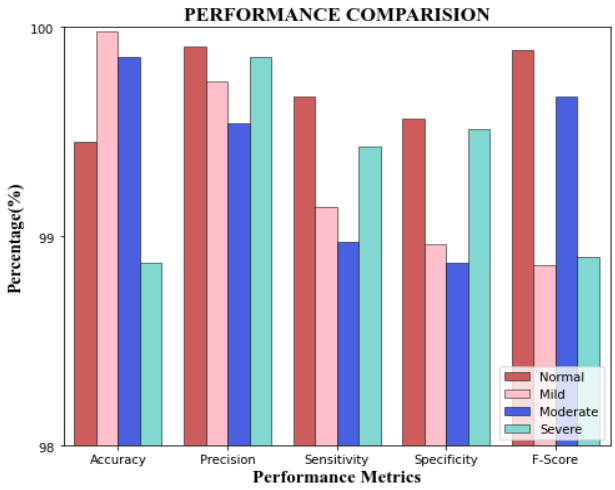

Figure 5. Comparison of four classes in DR classification

To fully assess our model, Table 4 compares our suggested methodology with a number of current standard techniques that were applied to the Messidor dataset. The suggested strategy outperforms existing methods in experiments conducted on the Messidor database, showing superior performance in both deep learning and traditional classification methods. Additionally, it shows that our method works well for DR grading issues involving multiple classes. It seems that our method performs favorably when compared to the prior state-of-the-art, despite the fact that the experimental setting in these techniques is a little different.

The comparison of proposed approach with other techniques in terms of accuracy is shown in Figure 6. From the figure, it is observed that the performance of traditional CapsNet and ResNet-50 is superior to other techniques such as WADNet and Quadrant-based Ensemble InceptionResnet-V2 (QEIRV2). However, the proposed system achieves 99.29% accuracy which is higher than all the other techniques. It shows the capacity of our proposed approach in multi-class classification of DR.

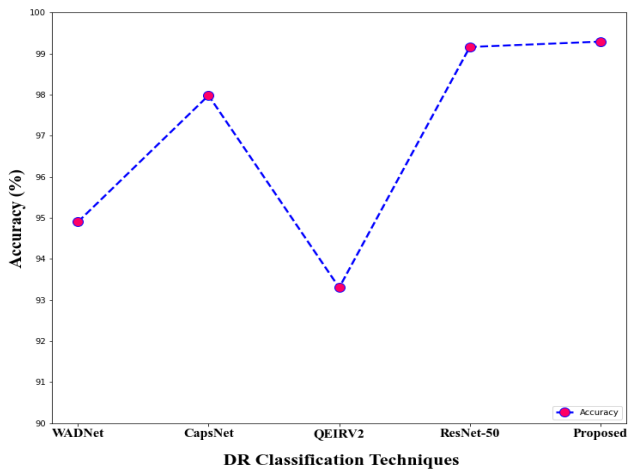

Figure 6. Accuracy comparison of DR classification

Table 3. Multi-class classification results of DR

Classes	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-score (%)
Normal	99.45	99.91	99.67	99.56	99.89
Mild	98.98	99.74	99.14	98.96	98.86
Moderate	99.86	99.54	98.97	98.87	99.67
Severe	98.87	99.86	99.43	99.51	98.9

Table 4. Comparison of classification results of DR

Techniques	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-score (%)
WADNet [27]	94.9	-	92.7	95.7	-
CapsNet [28]	97.98	95.62	96.11	-	95.82
QEIRV2 [29]	93.3	-	-	-	-
ResNet-50 [33]	99.16	99.72	99.16	99.16	-
Proposed	**99.29**	**99.76**	**99.30**	**99.22**	**99.33**

Task 2: ROP classification

The ROP classification is the binary class classification problem and the obtained results are shown in Table 5. Afterward, the classifier results are compared with the existing methods and it is shown in Table 6.

As seen from Table 6, the proposed classifier achieves 99.12% accuracy, 98.89% precision, 98.97% sensitivity, 99.32% specificity, 99.56% F-score respectively. These results show that the proposed approach achieves the highest values than other standard machine and deep learning techniques. The important cause to achieving these results is DCGAN based data augmentation approach. It provides a high number of training samples which is advantageous to model training, particularly for the highly sophisticated deep network.

Table 5. Binary classification results of ROP

Classes	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-score (%)
ROP	98.93	98.85	98.89	99.23	99.49
Non-ROP	99.31	98.92	99.05	99.41	99.63

Table 6. Comparison of ROP classification results

Techniques	Dataset	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F-score (%)
DNN [30]	Private	98.27	90.92	90.55	-	90.43
SVM [34]	Private	-	-	95	93	-
Q-SVM [35]	Private	93.2	95.1	89.7	96.1	92.3
CNN [36]	Private	-	-	94	-	-
Proposed	Private	**99.12**	**98.89**	**98.97**	**99.32**	**99.56**

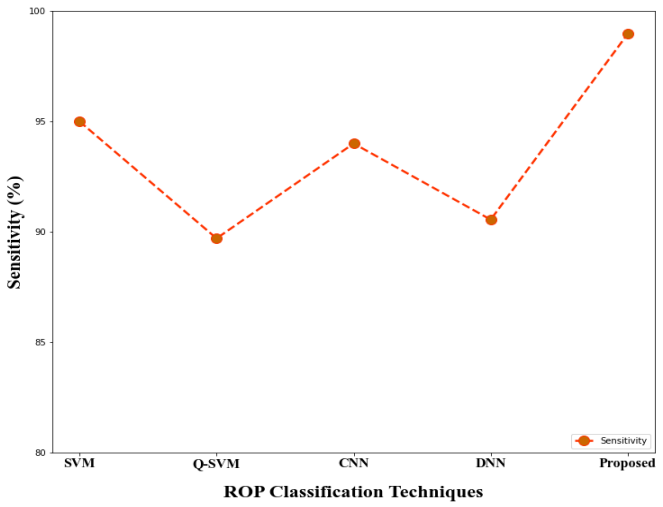

Figure 7. Comparison of sensitivity in ROP classification

In terms of sensitivity, the Support Vector Machine (SVM) and Convolutional Neural Network (CNN) attain comparable values. The accuracy of Q-SVM (93.2%) is lower than Deep Neural Network (DNN) (98.27%). Although the precision value of Q-SVM (95.1%) is higher than DNN (90.92%). From the results, it is identified that the classification of deep learning network is better than machine learning. Compared to CNN and DNN, the sensitivity of CNN (94%) is greater than DNN (90.55%). Visual illustration of sensitivity comparison is displayed in Figure 7.

4.4 Data augmentation evaluation

Table 7 provides an overview of how data augmentation influenced the performance of the proposed classifier. It is clear from the table that applying the data augmentation technique enhances classifier performance.

Table 7. Analysis of data augmentation technique

Classification	With DCGAN (%)	Without DCGAN (%)
DR Accuracy	99.29	98.56
ROP Accuracy	99.12	98.41

With the DCGAN-based data augmentation strategy, the classifier achieves 99.29% accuracy during DR categorization. It accomplishes only 98.56% without DCGAN. Similarly, the system uses data augmentation approaches to achieve 99.12% accuracy in ROP categorization. In comparison to training images without data augmentation, the quantity of samples for training produced by this method is much higher. By using this data augmentation technique, different sets of training samples

are generated for each epoch. Therefore, the proposed classifier perform better and attain maximum reliability in both categories by the use of the DCGAN based data augmentation approach.

5. CONCLUSION

Retinal diseases like DR and ROP are the life-altering conditions that affects many people and creates vision impairment. Therefore, automatic screening of DR and ROP is very essential. For this, a deep learning structure is developed for segmentation and classification of DR and ROP. The proposed methods training was conducted with the augmented dataset and obtained significant results for all the performance metrics. The proposed model delivers 99.29% accuracy, 99.76% precision, 99.30% sensitivity, 99.22% specificity and 99.33% F1-score for DR classification and the ROP classification attains the 99.12% accuracy, 98.89% precision, 98.97% sensitivity 99.32% specificity, and 99.56% f1-score respectively. Moreover, the segmentation performance of the proposed technique is also satisfactory and attains the best results. In the future, this work will be extended to locate other vision problems like cataract detection and glaucoma detection with a transfer learning approach.

REFERENCES

[1] Pugal Priya, R., Saradadevi Sivarani, T., Gnana Saravanan, A. (2022). Deep long and short term memory based Red Fox optimization algorithm for diabetic retinopathy detection and classification. International Journal for Numerical Methods in Biomedical Engineering, 38(3): e3560. https://doi.org/10.1002/cnm.3560

[2] Silva, P.S., Lewis, D., Cavallerano, J., Ashraf, M., Jacoba, C.M.P., Doan, D., Wang, F.S., Sun, J.K., Aiello, L.P. (2022). Automated machine learning (AutoML) model for diabetic retinopathy (DR) image classification from ultrawide field (UWF) retinal images. Investigative Ophthalmology & Visual Science, 63(7): 2095-F0084.

[3] Jayanthi, J., Jayasankar, T., Krishnaraj, N., Prakash, N. B., Sagai Francis Britto, A., Vinoth Kumar, K. (2021). An intelligent particle swarm optimization with convolutional neural network for diabetic retinopathy classification model. Journal of Medical Imaging and Health Informatics, 11(3): 803-809. https://doi.org/10.1166/jmihi.2021.3362

[4] Lu, C., Hanif, A., Singh, P., et al. (2022). Federated learning for multicenter collaboration in ophthalmology: improving classification performance in retinopathy of prematurity. Ophthalmology Retina, 6(8): 657-663. https://doi.org/10.1016/j.oret.2022.02.015

[5] Cole, E., Valikodath, N.G., Al-Khaled, T., et al. (2022). Evaluation of an artificial intelligence system for retinopathy of prematurity screening in Nepal and Mongolia. Ophthalmology Science, 2(4): 100165. https://doi.org/10.1016/j.xops.2022.100165

[6] Zhang, R., Zhao, J., Xie, H., Wang, T., Chen, G., Zhang, G., Lei, B. (2022). Automatic diagnosis for aggressive posterior retinopathy of prematurity via deep attentive convolutional neural network. Expert Systems with Applications, 187: 115843. https://doi.org/10.1016/j.eswa.2021.115843

[7] Canayaz, M. (2022). Classification of diabetic retinopathy with feature selection over deep features using nature-inspired wrapper methods. Applied Soft Computing, 128: 109462. https://doi.org/10.1016/j.asoc.2022.109462

[8] Bhardwaj, C., Jain, S., Sood, M. (2021). Hierarchical severity grade classification of non-proliferative diabetic retinopathy. Journal of Ambient Intelligence and Humanized Computing, 12: 2649-2670. https://doi.org/10.1007/s12652-020-02426-9

[9] Huang, Y.P., Basanta, H., Kang, E.Y.C., Chen, K.J., Hwang, Y.S., Lai, C.C., Campbell, J.P., Chiang, M.F., Chan, R.V.P., Kusaka, S., Fukushima, Y., Wu, W.C. (2021). Automated detection of early-stage ROP using a deep convolutional neural network. British Journal of Ophthalmology, 105(8): 1099-1103. http://dx.doi.org/10.1136/bjophthalmol-2020-316526

[10] Ara, R.K., Matiolański, A., Dziech, A., Baran, R., Domin, P., Wieczorkiewicz, A. (2022). Fast and efficient method for optical coherence tomography images classification using deep learning approach. Sensors, 22(13): 4675. https://doi.org/10.3390/s22134675

[11] Garifullin, A., Lensu, L., Uusitalo, H. (2021). Deep Bayesian baseline for segmenting diabetic retinopathy lesions: Advances and challenges. Computers in Biology and Medicine, 136: 104725. https://doi.org/10.1016/j.compbiomed.2021.104725

[12] Bengani, S. (2021). Automatic segmentation of optic disc in retinal fundus images using semi-supervised deep learning. Multimedia Tools and Applications, 80: 3443-3468. https://doi.org/10.1007/s11042-020-09778-6

[13] Sikder, N., Masud, M., Bairagi, A.K., Arif, A.S.M., Nahid, A.A., Alhumyani, H.A. (2021). Severity classification of diabetic retinopathy using an ensemble learning algorithm through analyzing retinal images. Symmetry, 13(4): 670. https://doi.org/10.3390/sym13040670

[14] Da Rocha, D.A., Ferreira, F.M.F., Peixoto, Z.M.A. (2022). Diabetic retinopathy classification using VGG16 neural network. Research on Biomedical Engineering, 38(2): 761-772. https://doi.org/10.1007/s42600-022-00200-8

[15] Mungloo-Dilmohamud, Z., Heenaye-Mamode Khan, M., Jhumka, K., Beedassy, B.N., Mungloo, N.Z., Peña-Reyes, C. (2022). Balancing data through data augmentation improves the generality of transfer learning for diabetic retinopathy classification. Applied Sciences, 12(11): 5363. https://doi.org/10.3390/app12115363

[16] Abaei Kashan, A., Maghsoudi, A., Shoeibi, N., Heidarzadeh, M., Mirnia, K. (2022). An automatic optic disk segmentation approach from retina of neonates via attention based deep network. International Journal of Engineering, 35(4): 715-724. https://doi.org/10.5829/ije.2022.35.04A.11

[17] Singh, P., Campbell, J.P., Ostmo, S., et al. (2021). External validation of a deep learning algorithm for plus disease classification on a multinational ROP dataset. Investigative Ophthalmology & Visual Science, 62(8): 3266-3266.

[18] Hanif, A., Yıldız, İ., Tian, P., Kalkanlı, B., Erdoğmuş, D., Ioannidis, S., Dy, J., Kalpathy-Cramer, J., Ostmo, S., Jonas, K., Paul Chan, R.V., Chiang, M.F., Campbell, J.P. (2022). Improved training efficiency for retinopathy of

prematurity deep learning models using comparison versus class labels. Ophthalmology Science, 2(2): 100122. https://doi.org/10.1016/j.xops.2022.100122

[19] Zia, F., Irum, I., Qadri, N.N., Nam, Y., Khurshid, K., Ali, M., Ashraf, I., Khan, M.A. (2022). A multilevel deep feature selection framework for diabetic retinopathy image classification. CMC-Computers Materials & Continua, 1(1): 12-20. http://dx.doi.org/10.32604/cmc.2022.017820

[20] Saranya, P., Prabakaran, S., Kumar, R., Das, E. (2022). Blood vessel segmentation in retinal fundus images for proliferative diabetic retinopathy screening using deep learning. The Visual Computer, 38(3): 977-992. http://dx.doi.org/10.1007/s00371-021-02062-0

[21] Maiti, S., Maji, D., Dhara, A.K., Sarkar, G. (2022). Automatic detection and segmentation of optic disc using a modified convolution network. Biomedical Signal Processing and Control, 76: 103633. https://doi.org/10.1016/j.bspc.2022.103633

[22] Abdelmaksoud, E., El-Sappagh, S., Barakat, S., Abuhmed, T., Elmogy, M. (2021). Automatic diabetic retinopathy grading system based on detecting multiple retinal lesions. IEEE Access, 9: 15939-15960. https://doi.org/10.1109/ACCESS.2021.3052870

[23] Xue, J., Yan, S., Qu, J., Qi, F., Qiu, C., Zhang, H., Chen, M., Liu, T., Li, D., Liu, X. (2019). Deep membrane systems for multitask segmentation in diabetic retinopathy. Knowledge-Based Systems, 183: 104887. https://doi.org/10.1016/j.knosys.2019.104887

[24] Agrawal, R., Kulkarni, S., Walambe, R., Kotecha, K. (2021). Assistive framework for automatic detection of all the zones in retinopathy of prematurity using deep learning. Journal of Digital Imaging, 34: 932-947. https://doi.org/10.1007/s10278-021-00477-8

[25] Nisha, K.L., Sreelekha, G., Sathidevi, P.S., Mohanachandran, P., Vinekar, A. (2019). A computer-aided diagnosis system for plus disease in retinopathy of prematurity with structure adaptive segmentation and vessel based features. Computerized Medical Imaging and Graphics, 74: 72-94. https://doi.org/10.1016/j.compmedimag.2019.04.003

[26] Abaei Kashan, A., Maghsoudi, A., Shoeibi, N., Heidarzadeh, M., Mirnia, K. (2022). An automatic optic disk segmentation approach from Retina of Neonates via Attention Based Deep Network. International Journal of Engineering, 35(4): 715-724. https://doi.org/10.5829/ije.2022.35.04A.11

[27] Cao, P., Hou, Q., Song, R., Wang, H., Zaiane, O. (2022). Collaborative learning of weakly-supervised domain adaptation for diabetic retinopathy grading on retinal images. Computers in Biology and Medicine, 144: 105341. https://doi.org/10.1016/j.compbiomed.2022.105341

[28] Kalyani, G., Janakiramaiah, B., Karuna, A., Prasad, L.N. (2021). Diabetic retinopathy detection and classification using capsule networks. Complex & Intelligent Systems, 1-14. https://doi.org/10.1007/s40747-021-00318-9

[29] Bhardwaj, C., Jain, S., Sood, M. (2021). Deep learning–based diabetic retinopathy severity grading system employing quadrant ensemble model. Journal of Digital Imaging, 34: 440-457. https://doi.org/10.1007/s10278-021-00418-5

[30] Peng, Y., Zhu, W., Chen, Z., Wang, M., Geng, L., Yu, K., Zhou, Y., Wang, T., Xiang, D., Chen, F., Chen, X. (2021). Automatic staging for retinopathy of prematurity with deep feature fusion and ordinal classification strategy. IEEE Transactions on Medical Imaging, 40(7): 1750-1762. https://doi.org/10.1109/TMI.2021.3065753

[31] Wang, L., Gu, J., Chen, Y., Liang, Y., Zhang, W., Pu, J., Chen, H. (2021). Automated segmentation of the optic disc from fundus images using an asymmetric deep learning network. Pattern Recognition, 112: 107810. https://doi.org/10.1016/j.patcog.2020.107810

[32] Castro, D.L., Tegolo, D., Valenti, C. (2018). A fast multiresolution approach useful for retinal image segmentation. In ICPRAM, pp. 340-345.

[33] Sambyal, N., Saini, P., Syal, R., Gupta, V. (2021). Aggregated residual transformation network for multistage classification in diabetic retinopathy. International Journal of Imaging Systems and Technology, 31(2): 741-752. https://doi.org/10.1002/ima.22513

[34] Nisha, K.L., Sreelekha, G., Sathidevi, P.S., Mohanachandran, P., Vinekar, A. (2019). A computer-aided diagnosis system for plus disease in retinopathy of prematurity with structure adaptive segmentation and vessel based features. Computerized Medical Imaging and Graphics, 74: 72-94. https://doi.org/10.1016/j.compmedimag.2019.04.003

[35] Attallah, O. (2021). DIAROP: automated deep learning-based diagnostic tool for retinopathy of prematurity. Diagnostics, 11(11): 2034. https://doi.org/10.3390/diagnostics11112034

[36] Chen, J.S., Coyner, A.S., Ostmo, S., Sonmez, K., Bajimaya, S., Pradhan, E., Valikodath, N., Cole, E.D., Al-Khaled, T., Paul Chan, R.V., Singh, P., Kalpathy-Cramer, J., Chiang, M.F., Campbell, J.P. (2021). Deep learning for the diagnosis of stage in retinopathy of prematurity: accuracy and generalizability across populations and cameras. Ophthalmology Retina, 5(10): 1027-1035. https://doi.org/10.1016/j.oret.2020.12.013