ORIGINAL PAPER

Deep Learning–Based Diabetic Retinopathy Severity Grading System Employing Quadrant Ensemble Model

Charu Bhardwaj[1] [ORCID] · **Shruti Jain[1]** · **Meenakshi Sood[2]**

Abstract
The diabetic retinopathy accounts in the deterioration of retinal blood vessels leading to a serious compilation affecting the eyes. The automated DR diagnosis frameworks are critically important for the early identification and detection of these eye-related problems, helping the ophthalmic experts in providing the second opinion for effectual treatment. The deep learning techniques have evolved as an improvement over the conventional approaches, which are dependent on the handcrafted feature extraction. To address the issue of proficient DR discrimination, the authors have proposed a quadrant ensemble automated DR grading approach by implementing InceptionResnet-V2 deep neural network framework. The presented model incorporates histogram equalization, optical disc localization, and quadrant cropping along with the data augmentation step for improving the network performance. A superior accuracy performance of 93.33% is observed for the proposed framework, and a significant reduction of 0.325 is noticed in the cross-entropy loss function for MESSIDOR benchmark dataset; however, its validation utilizing the latest IDRiD dataset establishes its generalization ability. The accuracy improvement of 13.58% is observed when the proposed QEIRV-2 model is compared with the classical Inception-V3 CNN model. To justify the viability of the proposed framework, its performance is compared with the existing state-of-the-art approaches and 25.23% of accuracy improvement is observed.

Keywords Diabetic retinopathy · Deep neural network · Convolution neural network · Hand-crafted features · InceptionResnet-V2 · Data augmentation

Introduction

Diabetic retinopathy (DR) is caused due to intensive visual disability occurring due to prolonged disintegration of blood vessels present in retinal region. DR becomes more complex to cure at its elevated stages; therefore, initial recognition of the problem is significant. Early detection of DR is essential for clinical prognosis in order to provide the treatment and to further reduce the disease advancement. Early-stage DR detection can be categorized into four significant classifications: non-retinopathic, mild, moderate, and severe stage. The automated DR evaluation methods are adopted by several researchers, and these approaches present different strategies for severity recognition and its grading into several stages [1, 2]. The treatment of DR plan for various patients varies with the severity level of DR criticality. Patients suffering from zero or mild DR indicate the requirement of ordinary treatment with screening, whereas the patients suffering from moderate and severe indications of DR may refer to the vitrectomy and laser treatment. According to severity level, instant and timely treatment of patient is significant [1].

The strategies of DR screening through fundus images are broadly utilized because of their easiness, suitable acquisition, and better visibility of lesions. The increase in diabetic patient has strengthened the scope of advanced proficient ophthalmologists for initiating the requirement of automatic DR diagnosis frameworks. The indications of potential DR are not observable through naked eye;

✉ Charu Bhardwaj
  cbcharubhardwaj215@gmail.com

  Shruti Jain
  shruti.jain@juit.ac.in

  Meenakshi Sood
  meenakshi@nitttrchd.ac.in

1 Department of Electronics and Communication Engineering, JUIT Waknaghat, Solan, HP, India

2 Department of CDC, NITTTR, Chandigarh, India

therefore, a system for automatic early detection of DR is the foremost requirement to investigate the characteristics and pattern of DR [1]. The computerized visualization for diabetic retinopathy is critical to deal with, thereby, providing the burden on ophthalmologists for recognition of patients who require immediate eye care and explicit treatment [2]. High clinical pertinence of DR arrangement for better diagnosis leads various researchers to design an automatic DR diagnosis framework.

At present, the machine learning approaches presents different computer-aided possibilities for the automated classification and analysis of DR. The DR recognition utilizes different feature extraction methods to extract valuable data from the input fundus images. The feature extraction is carried out manually along with the variabilities in visual attributes of various lesions and should be robust for DR lesion variations [3]. The automated frameworks of DR recognition may implement the manual feature extraction technique for the detection of lesion. It ensures that the lesions can be detected during isolation as well as with the combination of other lesions, providing a second opinion to the ophthalmologist for decision making and further assessment. The machine learning–based algorithms are capable of categorizing lesion classification depending upon the decision boundary along with the activation functions. These machine learning approaches can neither adjust these decision boundaries by implementing non-linear values and nor proficient for effectual learning, thereby restricting their abilities for settling the difficult

tasks. Additionally, machine learning technique can be improved by implementing feature engineering, which itself is a laborious process requiring the proficient domain awareness. The feature attributes being used by machine learning techniques should have been recognized by the domain professionals to reduce data complexity and to study the output classification attributes. The deep learning has evolved as a forward leap to automate the feature engineering procedure by including feature learning effectively while learning the features in an incremental manner. Deep learning approach is considered as end-to-end solution finding technique without partitioning the process into various parts and afterward joining at the last phase, as done for machine learning. The deep neural network (DNN) designs have outperformed manual grading frameworks in numerous applications. On the other hand, convolutional neural networks (CNN) has accomplished improvement for image characterization and recognition, and they are subsequently implemented for DR diagnosis frameworks.

The basic CNN model is presented in Fig. 1 composes of input layer, followed by the stack of convolution layer, and the number of pooling layers, followed by the fully connected layer and the output layer.

The architecture of CNN model involves the combination of feature attribute extractor and classifier. Feature extraction is carried out from the input layer utilizing both convolution and pooling layers. The incremental increase in number of convolution layers is utilized to make CNN extract more complex feature attributes. The initial network layers are
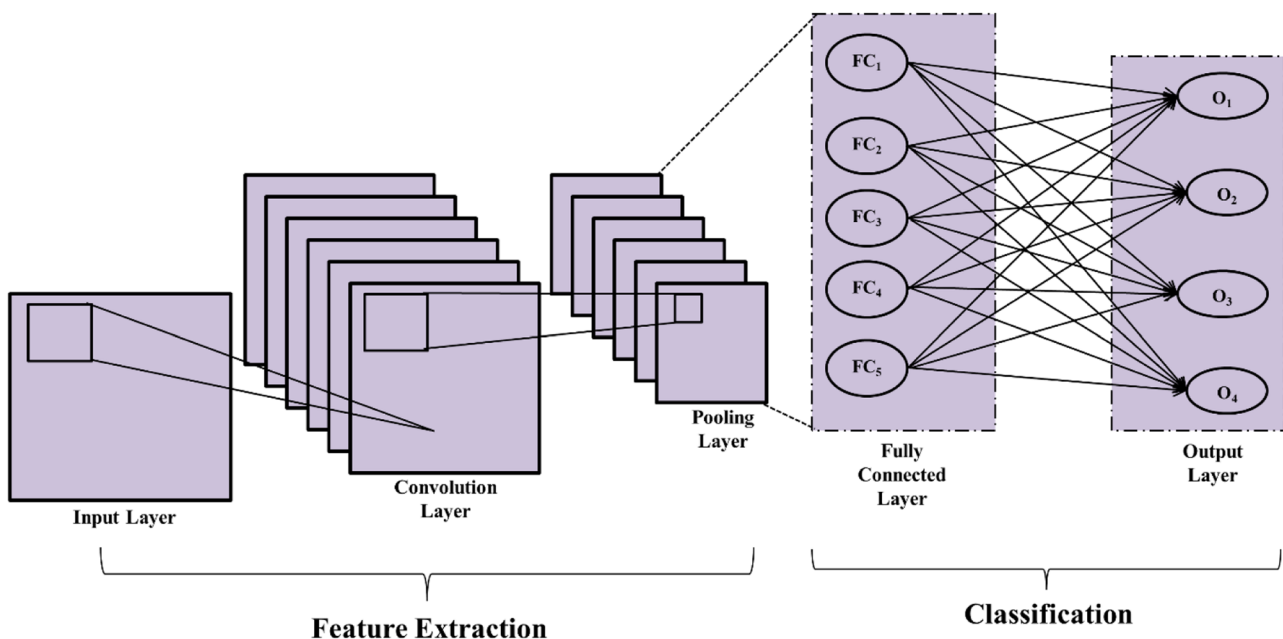

Fig. 1 Convolutional neural network architecture

liable for perceiving the edge properties present in fundus images; on the other hand, the deeper layers of convolution layers is responsible for learning the feature attributes for fundus images recognition into different DR severity grades. The feature vector attained from fully connected layer is forwarded for the classification, and then, the classification outcomes are acquired at the output layer.

Unlike machine learning, CNN-based characterization strategies give better execution because of its property of scale, rotational property, and its wider field of view. These properties of CNN make it reasonable for DR grading on the basis of retinal fundus images, as these images differ in terms of various sizes and various fields of view. Regardless of the numerous advances made to utilize CNN in DR diagnosis, these frameworks still face challenges for majority of healthcare applications. This article attempts to address these challenges by implementing a CNN approach on the basis of transfer learning.

A significant amount of time along with huge dataset is needed for training a DNN model out of scratch, and representation is challenging for healthcare domain where fixed number of data is accessible. Therefore, considering the task specific application, the transfer learning approach is exceptionally better, permitting the utilization of pre-trained structures previously utilized for addressing the similar domain-specific problem. The transfer learning technique permits the learning of another task on the basis of previous knowledge making the learning process quicker and more precise without requiring a large amount of the training data.

Different literatures for DR severity categorization using CNN uncovered that the utilization of transfer learning gives better results comparative to the construction of novel CNN structures [4]. The well-known dataset utilized for transfer learning approach–based pre-trained CNN frameworks is ImageNet dataset [5], which comprises a large number of images for training. This huge dataset is utilized for training, and the non-exclusive feature attributes are identified by the fundamental layers of CNN and the dataset-explicit feature attributes are recognized by the final layers.

Related Work

In the recent years, various researchers have contributed towards the automation of DR diagnosis utilizing different techniques. This section presents two main aspects: first section summarizes the work carried out for the automatic DR diagnosis utilizing conventional machine learning–based methods followed by the second section presenting recent work carried out in this domain composing of convolutional neural network frameworks.

Traditional Machine Learning–Based Approaches

The conventional DR classification techniques include significant stages of three types: one is image processing, followed by stage of feature extraction, and the final stage is classification of severity level. The studies for automated DR diagnosis include separate segment of feature extraction before transferring towards final stage of classifier. The classification of DR at its initial phase utilizing Machine Learning Bagging Ensemble Classifier (ML-BEC) is studied [6]. Their strategy involves feature extraction in its initial stage followed by stage of classification by utilizing features extracted in the first stage. Srivastava et al. [7] uses various kernel strategies in order to recognize the features of MA and HM. Seoud et al. [8] presented a study for the recognition of different severity levels of DR by using random forest classifiers for fundus image classification. Sankar et al. [9] presented a study for the identification of non-DR fundus images in terms of various DR grades such as mild grade, moderate, and extreme grade. Seoud et al. [10] studied shape-based feature for hemorrhages and MA detection, and their work presents superior classification outcomes. Pires et al. [11] investigates the automated DR recognition utilizing through applying image recognition in initial phases. Meta-characterization method is utilized to estimate the patient requirement for providing instant healthcare services at clinics to give an automated recognition instrument in remote locations. Antal et al. [12] presented an ensemble-based framework for improving MA recognition method in terms of its reliability. A combination of pre-processing segments and candidate extractors utilizing 0.90 ± 0.01 AUC (area under curve) is accomplished for the DR and non-DR classifications of retinal structures dependent on the existence and non-existence of MAs. In [13], Mansour et al. discussed an overview determining the accuracy along with the efficiency of computer-aided diagnosis (CAD) frameworks utilizing conventional along with evolutionary methodologies. Their study reveals that the implementation of DR-CAD framework where the approach of optimization and evolutionary methods plays a significant part in pre-processing utilizing enhancement filters, segmentation phase, dimensionality reduction phase, feature extraction, and its selection along with the classification. Dai et al. [14] presented a study using interleaved deep mining procedure, for studying the identification of MA in order to address the restrictions of present techniques which neglects to report inter and intra class varieties for the detection of MA. Their study presents a report for MA region identification during fundus image classification by utilizing hybrid interleaved-deep mining procedure. Rahim et al. [15] presented a

new model for automatic DR recognition for early-stage recognition of MA. The adopted procedure includes fuzzy histogram for pre-processing and the feature extraction before severity level classification.

Conventional methodologies deploy different handcrafted feature extraction procedures for the extraction of clinically relevant feature attributes from fundus images. The handcrafted feature attributes are restricted, and creating new-effective features is difficult and does not result in better outcomes. This feature extraction method is laborious and may lead to misclassification. The decision of choosing suitable handcrafted feature attributes requires examination of different boundary limitations with prior information by ophthalmic experts.

The advancement and capabilities of DNN replaced DR diagnosis approaches based on machine learning and present huge execution improvement in various medical image processing applications [16].

DCNN–Based Approaches

Recently, with the evolution of deep neural network (DNN)–based techniques which have accomplished creative results in different fields particularly in healthcare applications. An advancement came into existence in medical image examination and its classification performance, by introducing new bio-image approach dependent on ensemble convolutional neural networks (CNNs). An automated computer–aided framework on the basis of deep visual feature for diagnosis was discussed by Abbas et al. [17] for the classification of various DR severity grades without employing pre-processing phases. The deep visual feature attributes were inferred utilizing deep learning–oriented multilayer semi-supervised method. Deep learning method in combination with the domain information for the recognition of red lesion and random forest classification is implemented for the classification of retinal images on the basis of their severity grade. Wang et al. [18] present a deep learning–based scheme by implementing regression activation maps in accordance with the pooling layers for estimating the region of interest–based grading of severity levels. Yu et al. [19] presented a new characterization approach for analysis of image quality depending upon the human visual framework which is combination of saliency charts and CNNs to acquire features both supervised and unsupervised which further subjected to support vector machine (SVM) classifier as an input. Gao et al. [20] utilized deep convolutional neural network (DCNN) framework for severity level grading of DR utilizing labeled dataset. The framework is capable of achieving accuracy of 88.72% for evaluating severity levels grading of DR through fundus images. Additionally, the network was tested for its deployment on cloud based platform for providing DR diagnostic facilities to various

healthcare centers. A combination of Gaussian mixer model (GMM) and visual geometry group network (VGGNet-19) is adopted for obtaining better outcomes in terms of accuracy [21]. Singular value decomposition (SVD) along with the feature selection platform of principle component analysis (PCA) was utilized at FC7 and FC8 layers termed as fully connected to accomplish better accuracy in terms of classification. The network performs better for AlexNet dataset and achieves high percentage accuracy of classification and achieves better average computation time for standard KAGGLE dataset. Transfer learning model on the basis of CNN was proposed by Li et al. [22] for the classification of DR by adjusting the boundaries of pre-trained framework. The observed results present that transfer learning model on the basis of CNN achieves better results of classification for small datasets. The CNN-based framework is applied on colored fundus images for DR recognition, and 95% of validation sensitivity is achieved [23]. The experiments reveal that the histogram equalization along with the fidelity of dataset improves the recognition of features and transfer learning utilizing pertained models presents accuracy of 74.5%, 68.8%, and 57.2% for 2-ary, 3-ary, and 4-ary for the GoogleNet transfer framework. The CNN oriented exudates localization scheme for diabetic macular edema (DME) detection by analyzing datasets of two fundus images: one is MESSIDOR, and other is E-Ophtha datasets [24]. The classification model of DME delivers the 77% of accuracy along with 0.78 of validation harm, thereby improving the diagnosis of disease. An algorithm is presented for the diagnosis of DR on the basis of Alexnet CNN which achieves 88.3% accuracy using MESSIDOR dataset fundus images [25]. Wang et al. [26] proposed an algorithm on the basis of CNNs Zoom-in-Net for diagnosis of DR which presents accuracy of 91.1% which is superior when compared with the existing state of art approaches for similar MESSIDOR dataset. One study applies the Inception-V3 pre-trained scheme for dealing with the experiment of marked training data inadequacy for the detection of DR. The minor sub-sample from Kaggle DR database was utilized for model training, and the accuracy of framework was tested and verified on another sample dataset. The transfer learning method can be implemented in order to overcome additional deep learning–based problems of healthcare domain. Additionally, an automatic diagnosis framework is proposed for the detection of DR on the basis of smartphone using Inception-based CNN model and classification strategy of binary decision tree. The significance of this technique into smartphone delivers a suitable DR diagnostic facility even for the experts of non-ophthalmic. One study presents a DR image recognition approach on the basis of deep CNN framework which composes of six various approached to increase the performance DR identification [27]. Without depending the fine-tuning, their architecture performs better

than the existing techniques on MESSIDOR similar dataset presents 5.26× of lesser parameters and reduction in overall computation cost of the system. The authors have presented a study comparing the traditional machine learning approaches and deep learning approaches on the basis of CNN [28]. CNN model offers improved agreement to the outcome of extremely trained human grading specialists while making use of various datasets. Inception-V3 CNN framework offers superior performance for each performance metrices and results accuracy performance of 89% for EyePACS dataset and accuracy of 81.6% for MESSIDOR dataset. A cross-disease consideration network was developed that absorbs particular features and discovers the internal relationship among the diabetic macular edema (DME) and the DR [29]. The fivefold cross-validation presents significant improvement in performance when estimated using the challenge of IDRiD and MESSIDOR dataset. The authors proposed an approach for the automatic detection of DR and its grading using CNNs [30]. In pre-processing step, the retinal fundus images are processed and resized and then subjected towards computational layers. A maximum accuracy of around 90.9% is attained for the images of MESSIDOR dataset forming the competency of CNN-based technique for the DR grading.

The literature review suggests that neural network–oriented strategies have shown superior outcomes for majority of applications including NN-based DR grading frameworks. The CNN model requires enormous amount of data to guarantee appropriate convergence without data accessibility and overfitting which is restricted in medical domain, especially in DR detection. The DR severity grading methods present in the literature exploits the handcrafter feature attributes which are laborious, require proficient expert knowledge, and may lead to misclassification. Despite of numerous progresses of DNN implementation in DR diagnosis, still these frameworks present difficulties for healthcare applications. DCNNs have been demonstrated prevalent for the image classification and furthermore introduced in DR diagnosis frameworks. Thus, to address the limitations of the reviewed work, this article aims at assessing the suitability of CNN-based methods for DR classification and grading due to its advantageous contribution to the medical imaging field.

This paper proposes a fundus imaging–based DR grading framework for the categorization of fundus images into various classifications based on retinal pathological changes using CNNs. A quadrant-based ensemble model utilizing InceptionResnet-V2 CNN model (QEIRV-2) is presented in this work for the identification and characterization of NPDR severities generalized over a standard MESSIDOR dataset. The novelty of the proposed framework lies in the integration of a pre-processing phase inclusive of fundus image enhancement, removal of optical disc (OD), quadrant cropping, and augmentation of

data in the model implementation pipeline. The quadrant ensemble methodology makes the framework more versatile to tiny lesions that exist in retinal fundus images which were not noticeable earlier in original fundus images. The innovation of the proposed QEIRV-2 model offers better accuracy by obtaining enormous trainable parameters, and the exploratory outcomes acquired exhibit the ability of the proposed framework for proficient DR determination. The proposed framework offers better performance by applying data augmentation approach and hence providing the scale flexibility, rotation, and invariance in field of view when validated on a latest IDRiD dataset. Additionally, the comparative analysis of QEIRV-2 model utilizing other standard model presents that the outcome accomplished by our proposed framework outperforms the present state-of-the-art methodologies justifying its generalization and viability.

The remainder of this paper is structured as follows: "Material and Methods" describes the materials and techniques required for convolution neural networks based application for diagnosis of DR. Proposed method is described in "Proposed Methodology" including the architectural layout of the proposed CNN architecture. The observed experimental results are discussed in "Experimental Results and Discussions" which is followed by the description of conclusion and future consideration in "Conclusion."

Material and Methods

This research work has presented a capable DR classification solution by categorization of fundus image as per the degree of disease severity. The objective is accomplished by separating the proposed method into various consecutive stages represented in Fig. 2.

Acquisition of Data

Acquisition of data is the essential phase for the execution of the proposed framework. Publically accessible fundus image datasets, for example, Digital Retinal Images for Vessel Extraction (DRIVE) [31], Structured Analysis of Retina (STARE) [32], DIAbetic RETinopathy DataBase-calibration level 1 (DIARETDB1) [33], Retinal Online Challenge (ROC [34], Methods for Segmentation and Indexing Techniques Dedicated to Retinal Ophthalmology Retinal Ophthalmology (MESSIDOR) [35], and Indian Diabetic Retinopathy Image (IDRiD) [36], are utilized for various DR diagnostic tasks. All the datasets are applicable for different diagnostic applications and consists of large number of fundus images, and they vary in image count, their sizes, field of view (FOV), and various annotations. This research work uses the methods for evaluating the segmentation and indexing approaches in the area of retinal
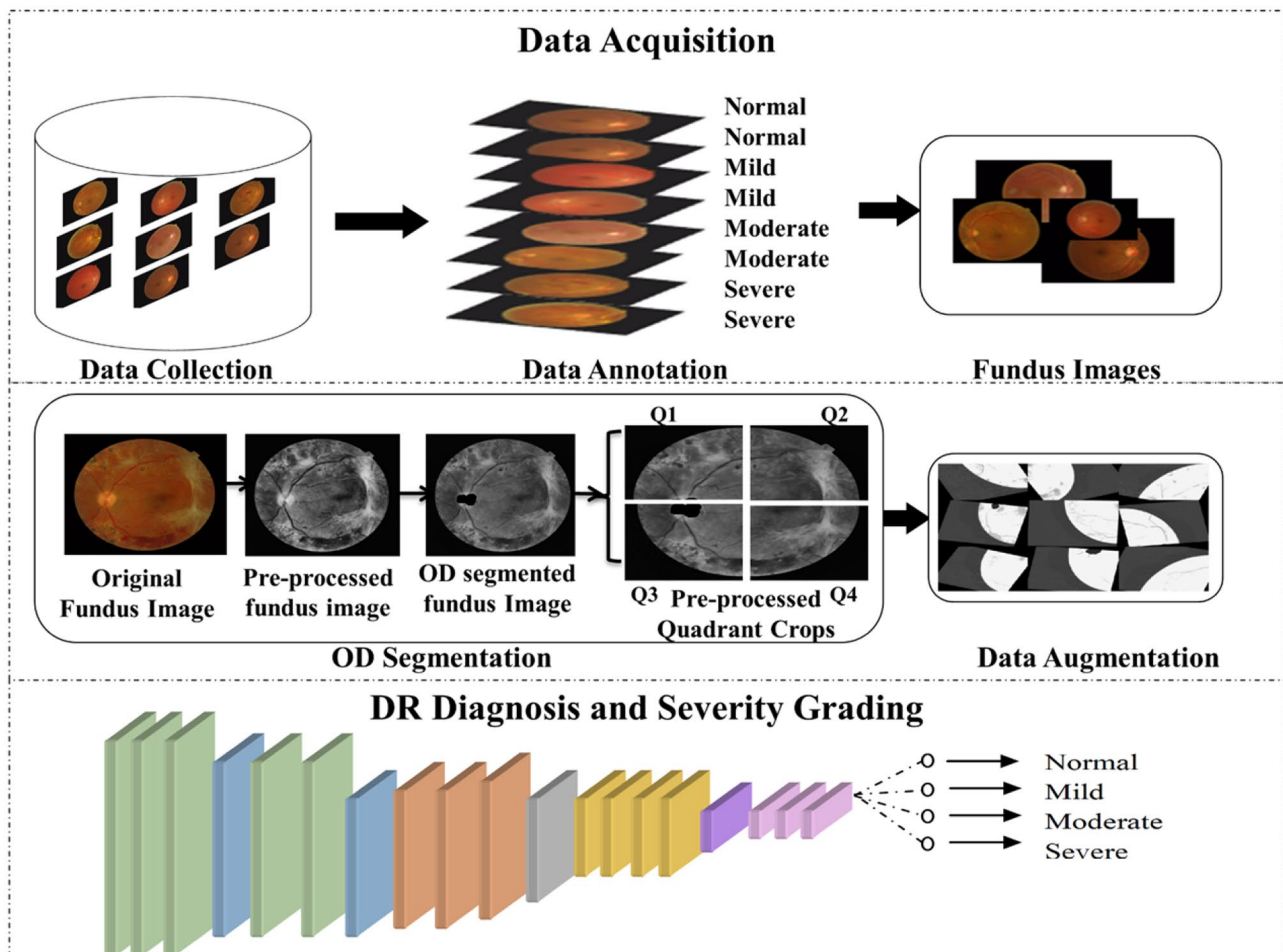
Fig. 2 Consecutive steps involved in the Quadrant-based Ensemble InceptionResnet-V2 (QEIRV-2) framework for severity grading

ophthalmology utilizing MESSIDOR dataset [35]. MESSIDOR is a DR grading dataset which consists of fundus images in which the labels of every image demonstrate the four severity levels. It composes of 1200 fundus images collected at 45° FOV having differing pixel sizes of 1440×960, 2240×1488, or 2304×1536. Out of the totality of 1200 images: 546 are non-diseased images with no signs of DR, 153 numbers of images consist of mild NPDR symptoms, moderate symptoms can be seen in 247 images, and 254 images consist of severe symptoms of NPDR severity.

The generalization capability of the proposed model is validated using the latest Indian Diabetic Retinopathy Image Dataset (IDRiD) [36] in this research work. The sample images are increased using the retinal fundus images from IDRiD (Porwal, et al. 2018) dataset in order to realize the disease progression. The dataset comprises clinical fundus images of thousands of patients examined at Eye Clinic located in Nanded (M.S.), India, during 2009–2017. Kowa VX-10a digital fundus camera was used to capture these images at 50° FOV while maintaining a distance of 39 mm between the camera lens and the eye. IDRiD dataset contains 454 images with NPDR

severities, out of which 168 have no DR sign, 25 images show mild NPDR signs, 168 have moderate, and 93 images have severe NPDR symptoms.

OD Segmentation

OD segmentation is accomplished in three stages including a pre-processing step (denoising and contrast improvement) boundary localization and augmentation of data.

Denoising and Contrast Enhancement

In this paper, the experimentation was conducted utilizing the colored fundus images collected from the benchmark dataset. Before forwarding the data towards the network, pre-processing phase is crucial in accomplishing optimal fundus image classification into retinopathic or non-retinopathic. Pre-processing phase is useful in identification of lesion and in order to differentiate among genuine lesions and non-lesions feature attributes of fundus images. Therefore, the pre-processing is important before extraction of features to recognize DR lesions. In the

pre-processing phase, normalization of size and color, denoising, contrast limited adaptive histogram equalization (CLAHE), and ODlocalization steps are utilized. All the fundus images are resized considering fixed dimensions of 600×600 pixels. Size normalization is followed by the normalization of color, as the fundus images collected utilizing various cameras have distinctive color temperatures and fluctuating illuminations [20, 37]. The color normalization utilized for this work incorporates the mean estimation operation over the value of each channel of the colored fundus image $I(x, y)$. The color normalization expression is presented by Eq. (1).

$$R_{xi,yi} = min\left\{ \frac{R_{x,y}}{mean\ (R)} \times 255 \right\}$$
$$G_{xi,yi} = min\left\{ \frac{G_{x,y}}{mean(G)} \times 255 \right\} \quad (1)$$
$$B_{xi,yi} = min\left\{ \frac{B_{x,y}}{mean(B)} \times 255 \right\}$$

where Rx, y, Gx, y, Bx, y is the value of pixels corresponding to R, G, B channels and the normalized pixels are denoted as $Rxi, yi, Gxi, yi, Bxi, yi$. The color normalization method provides a stable range of input by approximating the input from individual image channels for model learning.

Image denoising is a significant step to suppress the noise isolated without causing the blurred effect in the retinal fundus image. The genuine features of image such as edges or discontinuities are appropriately distinguished from the isolated noise using the median filtering operation expressed by Eq. (2):

$$I'(i, j) = median\left\{ I[x, y] \right\}, (x, y) \epsilon N \quad (2)$$

where $I'(i,j)$ is the filtered denoised image, $I[x, y]$ is the normalized image, and D represents the neighborhood centered around location $[i, j]$ in the normalized image [37].

OD Boundary Localization

The denoising of image is trailed with the step of optical disc localization, as it is very essential for extracting diagnosis of DR pathologies from the retinal anatomical structures. The boundary localization for optical disc is obtained through the conversion of pre-processed image into HSV plane for intensity value extraction. The operation of morphological closing is implemented and the biggest circular portion is confirmed as the OD portion. After the removal of OD, the segmented fundus images are obtained by separating the segmented regions from the pre-processed fundus images [38–40].

Data Augmentation

The success of neural networks depends on adequate training data, but at the same time, this necessity is not satisfied in majority of applications. For the domain of medical image, more training data implies additional annotations which are very

expensive due to the shortage of experienced ophthalmologists. The other disadvantage is the imbalance among the images of various diseased classes. In order to mitigate these restrictions and improve the ability of the network, the process of data augmentation is implemented. The step of data augmentation includes flipping (horizontal and vertical), 90–180° random rotation, and random zooming that ranges in between [0.85, 1.15] [41]. The process of data augmentation is useful in expanding the training samples as well as strengthening the size of class. The combination of image preprocessing and data augmentation makes the neural network insusceptible towards variation attenuation, insufficient illumination, and changing orientations [42–44].

DR Diagnosis System Setup

The convolutional deep neural networks are demonstrated better for the tasks of image recognition as well as image classification and are, therefore, implemented in DR diagnosis frameworks. The convolution neural network (CNN) addresses the issue of classification of DR utilizing hierarchical and discriminative automatic learning of features without earlier knowledge from clinical expertise. For this purpose, LeCun et al. [42, 45] presented a distinctive CNN framework which results superior outcomes for image recognition-based tasks. Convolutional layer, pooling layer, and complete connected layer are the three essential segments of CNN.

For an input image X of size, $i \times j \times c$, $i \times j$ is the image size and c represent the number of channels. The function calculated through convolutional layer is represented as Eq. (3).

$$f(X) = \sum_{n=0}^{c} w_n * X_n + bias \quad (3)$$

where w represents the weight vector, n denotes the number of notes, and * represents convolution operator.

Down-sampling is refined by max pooling layer considering a small sliding window of a particular stride size. The information feature map is partitioned into smaller patches, and the maximum of every patch is evaluated utilizing max operation. The intermediate reduction in dimensionality is achieved by embedding the max pooling layer in between the successive convolution layers. The function of non-differentiable, rectified linear unit (RELU) is implemented to introduce the non-linearity because of its steady gradient output response for positive values of input. Therefore, contribution of RELU is a typical choice for CNNs [41] which is represented as Eq. (4).

$$RELU(X) = \begin{cases} 0\ if\ X < 0 \\ X\ otherwise \end{cases} \quad (4)$$

The sequence of convolutional and pooling layers is trailed by the fully connected layers. Every node of this layer is in direct contact with the other node, thus increasing the number of connections among the successive layers. This process increases the complex computations and also rises the abundant network parameters. This issue is settled utilizing dropout strategy which drops out the intermediate connections from some nodes and, therefore, reduces the complex parameters. The softmax activation function is utilized at the output layer that permits output interpretation straightforwardly, as the value of probability range for this function is in between 0 to 1. The output softmax layer is further reduced to four probability counts indicating the four degrees of DR severity grading. The function of softmax activation is represented by Eq. (5).

$$Softmax(X) = \frac{e^{yi}}{\sum_i e^{yi}} \tag{5}$$

where e^{yi} represents the exponential function for the output vector yi.

The fine-tuning of CNN framework is essential to utilize the arrangement of pre-trained weights and exploit these attributes for other datasets. The fine-tuning of pre-trained CNN model utilizing the objective data can improve the performance of model which fits to the task of classification. Image attributes among the source and objective class can be adjusted by varying the output layer size to a value equivalent to the target classifications [43]. Various layers of CNN framework are helpful in learning numerous features at different levels of hierarchy. The features are extracted in an incremental manner from each of the convolutional layer which is further classified in fully connected layer. Low-level feature attributes are observed at starting convolutions; however, the high-level feature attributes are observed through deeper convolutions. In this paper, pre-trained InceptionResnet-V2-based CNN model is utilized as extractor of feature and afterward the task of classification of performed [44].

Performance Indices

The performance of the CNN model is assessed for the following metrics: performance accuracy, time complexity, and cross-entropy cost function. Time complexity is defined as the time needed to train the CNN model. The accuracy gives the correct prediction of test data utilizing the trained model. The value of accuracy provides the amount of correctly predicted data from the overall predictions. The cross-entropy loss assesses the exhibition of characterization model, and it ranges from 0 to 1 where 0 is the value of cross-entropy loss.

Accuracy is represented as Eq. (6) regarding false positive (FP), false negative (FN), true negative (TN), and true positive (TP), whereas the cross-entropy loss is described by Eq. (7).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$CrossEntropy = \frac{1}{\sum_{t=1}^{M} y_{o,t} \log(p_o, t)} \tag{7}$$

where M is class number, t represents correct classification for the observations o, p represents the likelihood of prediction, and y represents binary indicator (that may be 0 or 1) for the correct classification.

Proposed Methodology

Different CNN models have been anticipated for various tasks of classifications which include Lecun et al. [42, 45], AlexNet [46], VGGNet [47], ResNet [48], and Inception Architecture [49]. The authors of this article utilized the InceptionResnet-V2 version of benchmark Inception network as Inception CNN model, and its derivatives have indicated huge improvement for the application of image classification and recognition.

InceptionResnet-V2 model

InceptionResnet-V2 corresponds to 164-layer deep network that was at first intended to characterize 1000 object classes [49]. The network was trained over large number of images, and it may be retrained over a smaller set of data while preserving the information about the training model. This network attains the exceptionally accurate classification outcomes without requiring massive training comparative to its counterparts. The mathematical analysis is done for the computation of output size and trainable parameters of the network. The standard input image size for InceptionResnet-V2 is given by Eq. (8).

$$h_1 \times w_1 \times c_1 = 229 \times 229 \times 3 \tag{8}$$

where h_1 = height, w_1 = width and c_1 = channel number associated with the input image.

The initial convolutional filter consists of $K = 32$ filters of $f \times f = 3 \times 3$ size, stride (S) size of 2, and valid padding (P) of 35. Its output size is calculated using Eq. (9).

$$For\left(h_2 = \frac{h_1 - f + 2P}{S} + 1\right), \left(w_2 = \frac{w_1 - f + 2P}{S} + 1\right), (c_2 = K)$$

$$\xrightarrow{\text{yields}} \left(h_2 = \frac{229 - 3 + (2 \times 35)}{2} + 1\right), \left(w_2 = \frac{229 - 3 + (2 \times 35)}{2} + 1\right), (c_2 = 32) \tag{9}$$

Output size of Convolutional Filter $= h_2 \times w_2 \times c_2 = 149 \times 149 \times 32$

The number of trainable parameters in each layer is computed using the weight information (w_c) and the bias (b_c). For filter size $f=3$, number of filters $K=32$, and number of channels $c_1 = 3$, trainable parameters are calculated using the equation expressed by Eq. (10).

$$w_c = f^2 \times c_1 \times K \text{ and } b_c = K$$

Number of Trainable Parameters $= w_c + b_c$

$$\overset{yields}{\rightarrow} \left(3^2 \times 3 \times 32\right) + 32 = 896 \text{ parameters} \qquad (10)$$

Number of Trainable Parameters in the final Fully Connected layer $= w_c + b_c$

Similarly, the output size and learnable parameters of the other subsequent convolutional layers are also obtained using these equations.

The pooling layer is the subsequent layer to the convolutions in the network architecture. The output size of this layer is computed from the previous convolutional layer's output, i.e., $h_2 \times w_2 \times c_2 = 147 \times 147 \times 64$, filter size of pooling *is* $f \times f = 3 \times 3$, and stride (S)$=2$ given by Eq. (11).

$$\text{For } h_p = \frac{h_2 - f}{S} + 1, w_P = \frac{w_2 - f}{S} + 1, d_p = c_2$$

$$\overset{yields}{\rightarrow} h_p = \left(\frac{147 - 3}{2} + 1\right), w_p = \left(\frac{147 - 3}{2} + 1\right), d_p = 64$$

$$\qquad (11)$$

Output size of Pooling Layer $= h_p \times w_p \times d_p = 73 \times 73 \times 64$

No trainable parameters exist for the pooling layer as it calculated using a fixed function which requires no training.

Preceding the stack of convolutional and pooling layers, there is a fully connected layer at the end of the network which contains a vector of same size as that of the number of neurons present in it. The trainable parameters present in this layer are calculated using Eq. (12).

The output size of the preceding layer is $o = 1 \times 1$ for $K = 2048$ number of filters, and the fully connected layer consists of ($n_{FC} = 4$) neurons. Thus, the weight (w_{FC}) and bias (b_{FC}) values are calculated as follows:

$$w_{FC} = o^2 \times K \times n_{FC} = 1^2 \times 768 \times 2048 \text{ and } b_{FC} = n_{FC} = 4$$

$$\overset{yields}{\rightarrow} 1,572,864 + 4 = 1,572,868 \text{ parameters} \qquad (12)$$

Correspondingly, the output sizes and the parameters of further layers of InceptionResnet-V2 CNN model are calculated. Table 1 gives the layered network design of InceptionResnet-V2 network presenting the indexing of layer including layer type, activations, learnables, and trainable parameters.

The initial layers of InceptionResnet-V2 model are accountable for evaluating the edge properties of fundus image; on the other hand, final layers are countable for evaluating the classification attributes of images for retinopathy grading. With the increase in number of activation layers, the dimension of the

Table 1 Network architecture of InceptionResnet-V2 network

Type	Activations	Learnables			Parameters
		Size of filter	Stride	Number of filters/pooling	Trainable
Input data	$(299 \times 299 \times 3)$	—			—
Convolution layer	$(149 \times 149 \times 32)$	(3×3)	2	32	896
Convolution layer	$(147 \times 147 \times 32)$	(3×3)	1	32	9,248
Convolution layer	$(147 \times 147 \times 64)$	(3×3)	1	64	18,496
Pooling layer	$(73 \times 73 \times 64)$	(3×3)	2	Max pool	—
Convolution layer	$(73 \times 73 \times 80)$	(1×1)	1	80	5,200
Convolution layer	$(71 \times 71 \times 192)$	(3×3)	1	192	138,432
Pooling layer	$(35 \times 35 \times 192)$	(3×3)	2	Max pool	—
$5 \times$ InceptionResnet module A	$(35 \times 35 \times 256)$	—			—-
Reduction A	$(17 \times 17 \times 768)$	—			—-
$10 \times$ InceptionResnet module B	$(17 \times 17 \times 768)$	—			—-
Reduction B	$(8 \times 8 \times 2048)$	—			—-
$5 \times$ InceptionResnet module C	$(8 \times 8 \times 2048)$	—			—-
Pooling layer	$(1 \times 1 \times 2048)$	(8×8)	8	Avg pool	—-
Softmax output layer	4	—			1,572,868

image diminishes in terms of length and breadth, but the death of image increases until the layers output flattens.

In the initial phase, all the fundus images are provided to the InceptionResnet-V2 framework. The performance of the framework is observed for specific indices such as network accuracy, time complexity, and cross-entropy loss function considering epoch size as 20 and iteration count 200. Table 2 presents the performance outcomes obtained from the InceptionResnet-V2 framework for the considered parameters.

It depicts the comparison of statistics observed for Inception-Resnet-V2 framework which presents 82.14% accuracy for the maximum epoch size of 20 at the iteration of 200. The cross-entropy cost function is observed to provide the minimum value of 0.65 for the same parameter consideration. The time complexity taken by the CNN framework for the experimentation is observed as 38 min and 50 s. The accuracy of the network improves while taking 200 iterations, and it is noticed that it improves from 60% for first epoch to 82.14% at the 20th epoch of the network. After this point, the accuracy does not improve by increasing the epoch size or the iteration count. The higher accuracy values are not obtained for this experimentation as the complete fundus image is considered, and the features of small lesion are not extracted using CNN framework which further reduces the performance of the network. In order to overcome this restriction of the detection of small lesion and to improve network performance, the authors in this work have presented a

novel approach which is Quadrant-based Ensemble Inception-Resnet-V2 framework.

QEIRV-2

In this paper, the authors presented a QEIRV-2 framework which is designed utilizing four InceptionResnet-V2 frameworks. The models I, II, III, and IV shown in Fig. 4 are joined together for obtaining greater number of free parameters in order to achieve superior accuracy. The input to these four models is provided utilizing crops of fundus image at four various quadrants. The proposed CNN framework is encouraged from the approach of transfer learning that permits the knowledge transfer from one domain to another.

Transfer learning can be achieved implementing various approaches like data mapping from different domains at one common platform, and another is to use various templates and assigning them with individual weights. This article is an attempt where the initialization of model is carried out utilizing the InceptionResnet-V2 CNN framework for addressing the issue of DR diagnosis.

The input layer of the model assumes input size as 299×299 for the direct implementation, and therefore, the resizing of the fundus images up to that smaller size results in miss identification of tiny sized microaneurysms and exudates, which were easily visible in the original sized fundus images. To combat this issue, the fundus images are resized to pixel size of 600×600 which are further cropped in four quadrants of size 300×300. The cropped quadrants are further subjected to 4 Inception-Resnet-V2 frameworks and thereafter the processing; the outcomes of fully connected layers are concatenated for four different models and transferred to the softmax output layer. This alteration in the experimental phase of the original Inception framework presents improved capability by providing huge number of network parameters. Figure 3 depicts the strategy of cropping of fundus image into four quadrants.

The input subjected to the network layer are pre-processed using constant enhancement and OD localization steps followed by data augmentation forming the quadrant crops of input size 299×299. Figure 4 depicts the architectural layout of InceptionResnet-V2 CNN framework utilized in this work.

The proposed Quadrant-based Ensemble InceptionResnet-V2 model is effective in recognizing very small and tiny DR lesions present at quadrants of the retinal fundus image. Another benefit of using the proposed quadrant-based arrangement for DR classification lies in the expansion of training samples considered for the experimentation. Data augmentation strategies and partitioning every fundus into four quadrant crops overcome the data deficiency issue for neural network application. So, the drawback of availability of limited annotated dataset is addressed in this work through quadrant cropping and data augmentation. The complete process of the proposed methodology is detailed in Algorithm 1.

Table 2 Results of InceptionResnet-V2 model considering complete Fundus image as input

Epoch	Iterations	Time complexity	Accuracy value	Cross-entropy loss
1	1	52 s	60.00%	0.721
2	20	1 min 19 s	51.32%	0.793
3	30	2 min 47 s	53.21%	0.762
4	40	4 min 31 s	55.67%	0.771
5	50	6 min 15 s	62.17%	0.762
6	60	8 min 43 s	55.67%	0.732
7	70	11 min 47 s	62.17%	0.824
8	80	13 min 59 s	65.21%	0.765
9	90	15 min 09 s	62.17%	0.691
10	100	19 min 38 s	65.21%	0.682
11	110	21 min 56 s	55.67%	0.710
12	120	24 min 48 s	60.00%	0.652
13	130	25 min 52 s	62.17%	0.681
14	140	27 min 16 s	65.21%	0.684
15	150	30 min 26 s	71.57%	0.695
16	160	33 min 06 s	75.33%	0.672
17	170	35 min 22 s	75.33%	0.681
18	180	36 min 56 s	80.00%	0.672
19	190	37 min 05 s	82.14%	0.669
20	200	38 min 50 s	82.14%	0.665

min minutes, *s* seconds

Algorithm 1: Proposed Methodology for Quadrant based Ensemble Inception Resnet V-2 model

Input: Fundus Images acquired from the dataset.

Output: Diabetic Retinopathy grades distinguishing between different DR severities.

Begin

Step 1: Acquire the retinal fundus images from the image datastore and annotate these fundus images into different DR grades.

Step 2: Apply background elimination to attain the background segmented image.

Step 2.1: Pre-processing of fundus images

- Color normalization to model the fundus image *(I(x,y))* into corresponding R,G, B channels.
- Median filtering to discriminate the isolated noise from the normalized image.
- CLAHE on the denoised image to attain the pre-processed output fundus image ($I_{pre\text{-}process}$).

Step 2.2: Optical Disc (OD) segmentation

- Apply Hue, Saturation and Value (HSV) model conversion on the pre-processed image ($I_{pre\text{-}process}$).
- The largest circular region (I_{OD}) is identified using Morphological closing operation and contour filling.
- The OD segmentation ($I_{ODsegment}$) is obtained after OD removal.

$$I_{ODsegment} = I_{pre-processed} - I_{OD}$$

Step 4: Split all the OD segmented images ($I_{ODsegmented}$) into four quadrants: I_{QI}, I_{QII}, I_{QIII} and I_{QIV}.

Step 5: Apply data augmentation: Flipping (horizontal and vertical), 90-180 degrees random rotation and random zooming that ranges in between [0.85, 1.15].

Step 6: The augmented retinal fundus image crops are fed to four InceptionResnet-V2 CNN networks.

Step 7: Apply depth concatenation at the fully connected layers of all QEIRV-2 models.

Step 8: Access the network performance using network accuracy, cross-entropy loss function and time complexity.

Step 9: Acquire the classification output for different DR classes using benchmark MESSIDOR dataset.

Step 10: Perform validation using latest IDRiD dataset to improve the generalizability of the proposed model.

End

The Inception Resnet-V2 model is retrained using the transfer learning approach over the retinal fundus image MESSIDOR dataset while persisting the knowledge of its original training development. The learned knowledge of the previous training process is applied for the DR classification using comparatively smaller dataset and the network performance is observed for different cases of experimentation.

Fig. 3 Fundus image cropping strategy
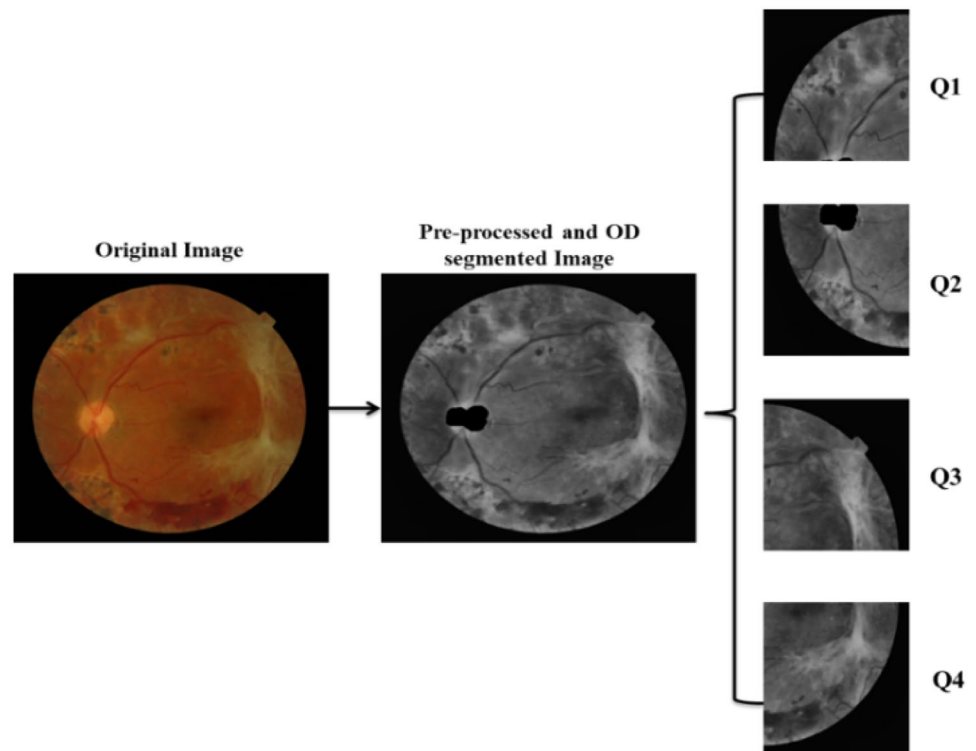
Experimental Results and Discussions

The experimentation was performed utilizing MATLAB2018b environment on a computer system framework equipped with Intel Core i5 processor, 8 GB RAM, and NVIDIA GeForce 4 GB GPU. The retinal images from MESSIDOR dataset are divided into training and testing set using the proportion of 3:1 to prevent poor randomization. The outcomes were investigated for the epoch size of 1 to 20, and the iteration count was expanded from 1 to 200 [18, 19]. The complete setting of parameters for the experimental analysis is depicted in Table 3.

The hyper-parameter setting for the experimentation consists of the batch size, base learning rate, dropout rate, epoch sizes, and iteration count. Keeping the values of all the hyper-parameters constant, the performance of the QEIRV-2 model is analyzed using evaluation parameters like network accuracy, cross-entropy cost function, and time complexity during the model handling [21, 41].

For the performance assessment of the proposed methodology, two unique modules are considered involving the image cropping and data augmentation strategies associated with the proposed model. The variation in experimentation of both the modules is seen in term of using the OD localization and contrast enhancement steps.

Module 1: QEIRV-2 CNN model without OD localization and contrast enhancement.

Module 2: QEIRV-2 CNN model with OD localization and contrast enhancement.

In module 1, the retinal fundus images are provided directly to the QEIRV-2 CNN model after data augmentation and image cropping without undergoing the OD localization and contrast enhancement steps. However, in module 2, initially, the OD localization and contrast enhancement steps are applied on the retinal fundus images and then data augmentation and imaged cropping are performed subsequently on the OD segmented and contrast enhanced retinal fundus images prior feeding them to the QEIRV-2 convolutional neural network. The results of the proposed model are assessed for the two unique modules in terms of various evaluation parameters.

Experimental Assessment of Module 1

In module 1 of experimentation, the input retinal images are cropped into four quadrant crops and data augmentations steps are utilized to increase the training examples. Quadrant-based Ensemble InceptionResnet-V2 model has been utilized in this work, and network execution is assessed in terms of accuracy, cross-entropy loss, and time complexity. The outcomes of the proposed QEIRV-2 CNN model are assessed for module 1 scenario which is tabulated in Table 4.
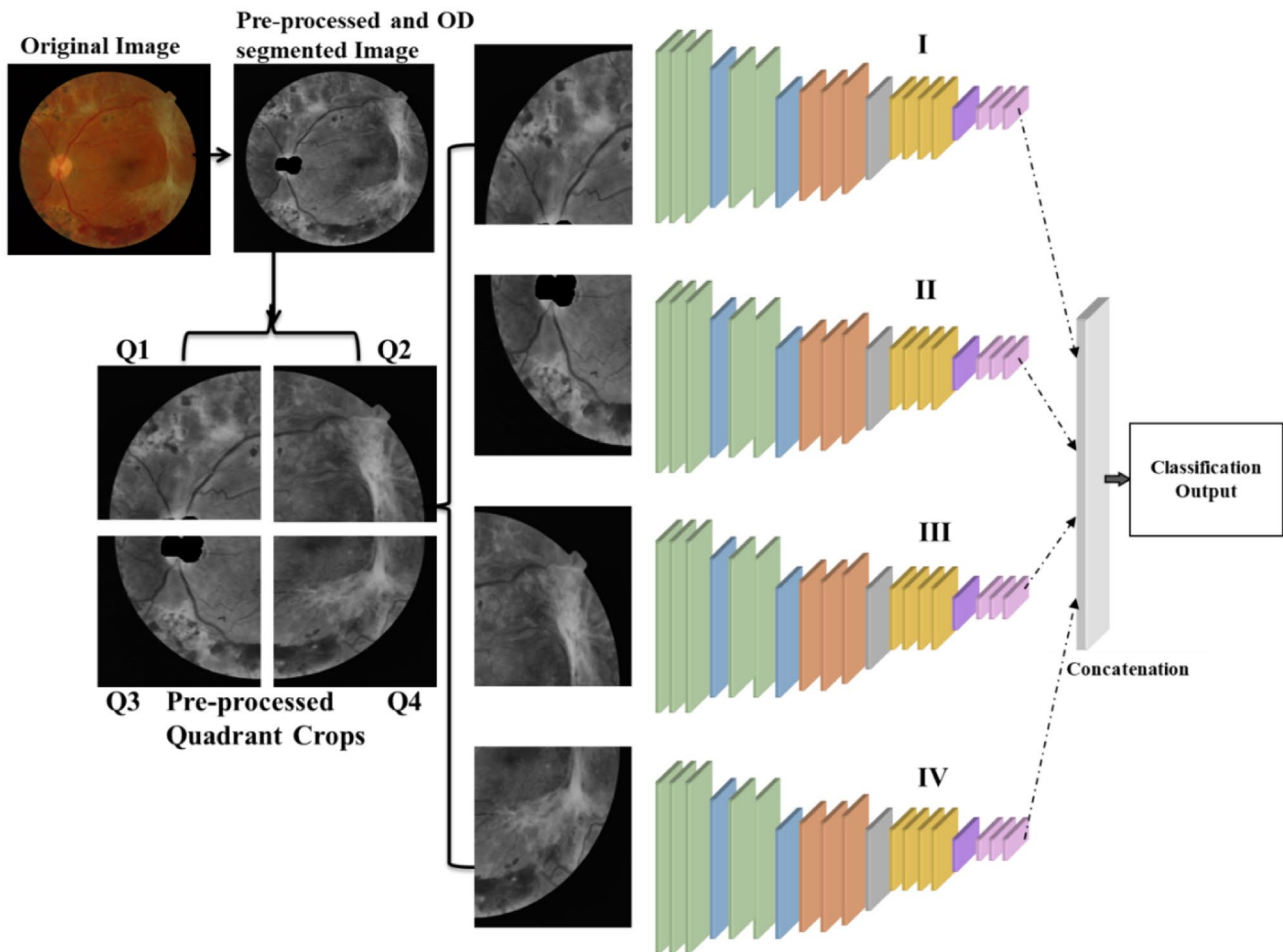
Fig. 4 Architectural layout of proposed Quadrant-based Ensemble InceptionResnet-V2 CNN model

The tabular representation depicts the performance indices of Quadrant-based Ensemble InceptionResnet-V2 CNN model while excluding the contrast enhancement and OD localization stages and including image cropping and data augmentation steps. For module 1 scenario, the proposed QEIRV-2 model gives 85.68% of accuracy rate at the 20th epoch and 200th iteration count while considering a fixed set of hyper-parameters. The cost analysis of the proposed model is done in terms of cross entropy loss and time complexity. From the experimentation, it is seen that the cross-entropy loss diminishes with the increase

Table 3 Hyper-parameter setting for experimental analysis

Hyper-parameters	Values allocated
Batch size	15
Epoch	20
Iterations	200
Base learning rate	0.0001
Dropout rate	0.5

in epoch size and the minimum value 0.401 of this cost function is attained at 20th epoch. The time complexity of the entire network execution is 49 min and 17 s. It is observed that the accuracy value improves from 40% at the 1st epoch count to 85.68% at the 20th epoch while considering 200 iterations. Authors have also tried for further evaluation while refining the epoch size beyond 20, but this leads to accuracy saturation. This quadrant ensemble methodology produces better network performance in terms of accuracy and cross-entropy cost function while maintaining a tradeoff with the time complexity. In order to make the framework more effective and robust, image processing pipeline is considered in the subsequent module 2 including the contrast enhancement and optical disc segmentation steps.

Experimental Assessment of Module 2

The impact of fundus image contrast enhancement and OD localization steps is checked on CNN network

Table 4 Results of proposed QEIRV-2 model for module 1

Epoch	Iterations	Time complexity	Accuracy value	Cross-entropy loss
1	1	52 s	40.00%	0.789
2	20	2 min 37 s	51.78%	0.698
3	30	4 min 56 s	55.49%	0.678
4	40	6 min 44 s	61.82%	0.667
5	50	8 min 52 s	54.94%	0.712
6	60	10 min 48 s	65.48%	0.609
7	70	13 min 53 s	62.95%	0.632
8	80	15 min 49 s	65.93%	0.614
9	90	18 min 29 s	65.50%	0.622
10	100	20 min 47 s	61.84%	0.675
11	110	23 min 28 s	73.37%	0.537
12	120	25 min 32 s	53.38%	0.667
13	130	27 min 47 s	65.68%	0.587
14	140	29 min 16 s	66.87%	0.562
15	150	32 min 47 s	65.58%	0.561
16	160	35 min 26 s	82.05%	0.493
17	170	37 min 52 s	82.05%	0.468
18	180	42 min 52 s	85.58%	0.458
19	190	45 min 48 s	85.68%	0.421
20	200	49 min 17 s	85.68%	0.401

min minutes, *s* seconds

Table 5 Results proposed QEIRV-2 model for module 2

Epoch	Iterations	Time complexity	Accuracy value	Cross-entropy loss
1	1	55 s	46.67%	0.705
2	20	3 mins18 s	50.00%	0.702
3	30	5 min 05 s	52.78%	0.672
4	40	7 min 27 s	61.11%	0.647
5	50	9 min 15 s	69.44%	0.641
6	60	10 min 26 s	70.83%	0.620
7	70	12 min 48 s	73.61%	0.580
8	80	15 min 08 s	70.83%	0.598
9	90	16 min 53 s	69.44%	0.613
10	100	19 min 38 s	72.22%	0.577
11	110	21 min 58 s	75.00%	0.546
12	120	23 min 47 s	77.78%	0.490
13	130	26 min 54 s	75.00%	0.592
14	140	29 min 19 s	76.39%	0.503
15	150	32 min 17 s	80.00%	0.486
16	160	35 min 50 s	80.00%	0.474
17	170	38 min 48 s	86.67%	0.584
18	180	41 min 07 s	86.67%	0.313
19	190	42 min 51 s	93.33%	0.359
20	200	44 min 22 s	93.33%	0.325

min minutes, *s* seconds

execution. The retinal fundus images are provided as the input to the proposed CNN architecture after applying the pre-processing steps. The same setting of hyper-parameters is considered as depicted in Table 3. InceptionResnet-V2 variant of baseline Inception model is tested experimentally using image pre-processing, OD segmentation, quadrant cropping, and data augmentation steps to analyze the variation in results obtained.

The assessment results shown in Table 5 present the performance statistics of proposed InceptionResnet-V2 model for module 2 scenario. The maximum accuracy of 93.33% is achieved for 20th epoch and 200th iteration. The time complexity of 44 min 22 s was observed for the training of network. The cross-entropy loss value also reduces at every epoch, and a much-reduced cross-entropy loss of 0.325 is observed at 20th epoch. The network accuracy performance improved from 46.67% at epoch 1 to 93.33% at 20th epoch considering 200 iterations for module 2 scenario. The cost function also reduced from 0.705 to 0.325 using this scenario while maintaining a tradeoff with the time complexity. Accuracy performance and cost function do not improve beyond this level even at increasing the epoch size.

The assessment of both the modules reveals that the involvement of the pre-processing pipeline (contrast enhancement and OD localization) improves the network performance by 8.89% in terms of accuracy, and the cost function is minimized by 18.95% in adjustment with the time complexity. The proposed Quadrant-based Ensemble InceptionResnet-V2 model yields improved DR grading performance with reduced cost function while utilizing the novel pipeline of image pre-processing and OD segmentation.

Comparative Analysis of Conventional Inception-V3 and Proposed QEIRV-2 Model

The performance assessment of proposed Quadrant-based Ensemble Inception-V3 (QEIRV-3) models is done with the conventional Inception-V3 CNN model in terms of accuracy, time complexity, and cross-entropy cost function, in order to validate the viability of image pre-processing pipeline. The examination was done considering the hyper-parameter setting of 20 epoch size, 200 iteration count, batch size of 15, dropout rate of 0.5, and base learning rate of 0.0001 as constant.

Figure 5 indicates that our proposed strategy has provided a noteworthy performance improvement comparative to the classical Inception-V3 network architecture. The accuracy improvement of 13.58% is seen with a critical reduction from
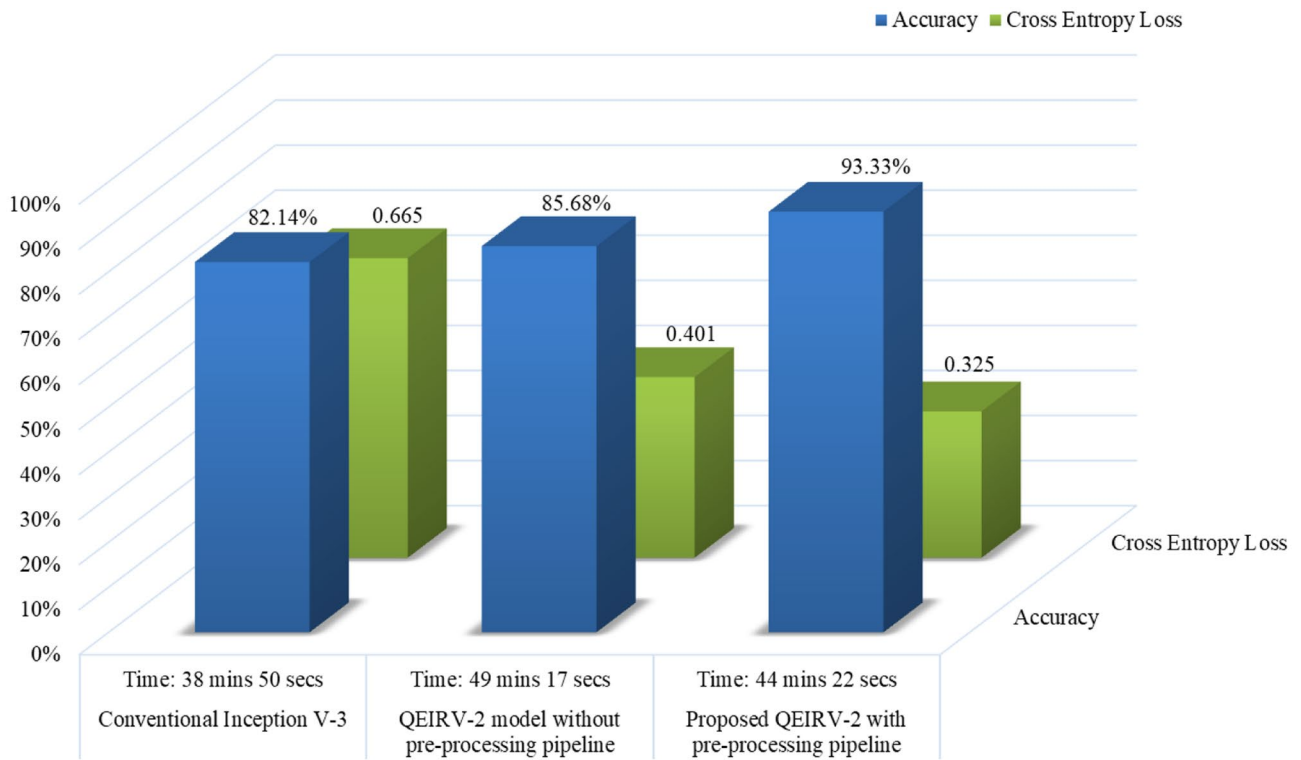
Fig. 5 Comparative analysis of conventional Inception-V3 and proposed QEIRV-2 models

0.665 to 0.325 in cross-entropy cost function comparative to the classical approach while using the proposed QEIRV-2 model. However, time complexity is a compromise in this case as the execution time for Inception-V3 model is 38 min 50 s while module 1 and module 2 of the QEIRV-2 take 49 min 17 s and 44 min 22 s, respectively.
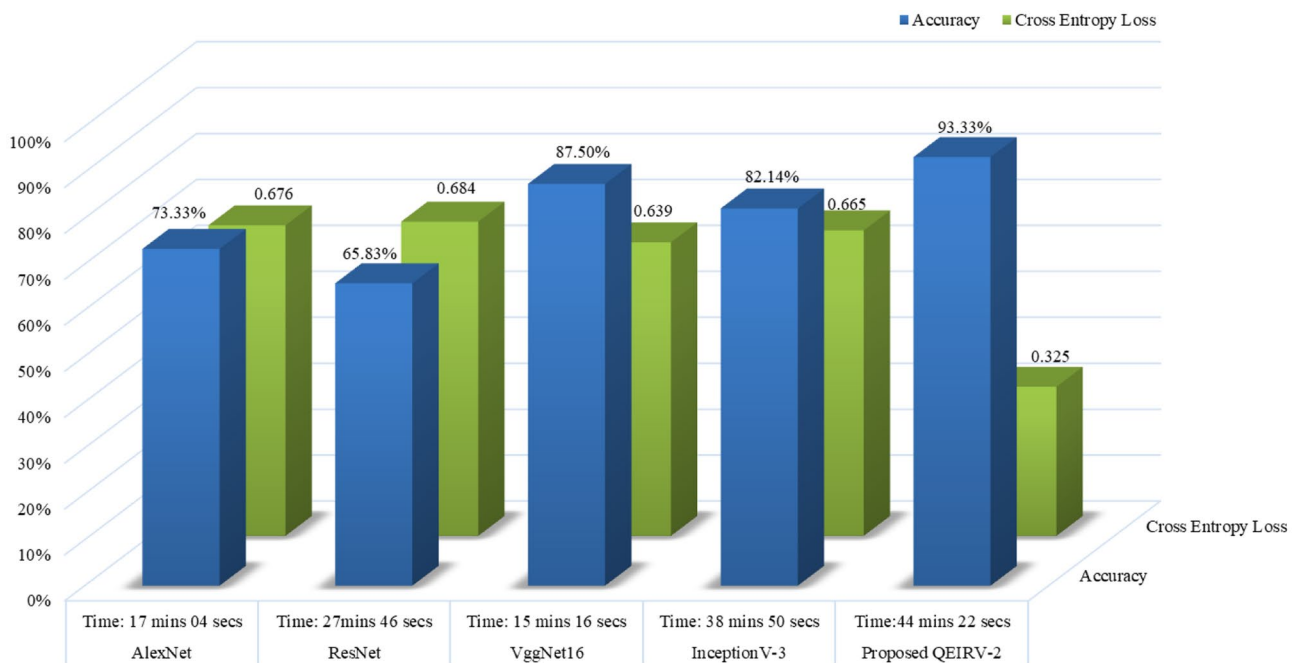

Fig. 6 Comparison of the proposed QEIRV-2 model with different CNN models for MESSIDOR dataset

Table 6 Results proposed QEIRV-2 model for module 2

Epoch	Iterations	Time complexity	Accuracy value	Cross-entropy loss
1	1	59 s	48.98%	0.715
2	20	03 min 45 s	51.05%	0.709
3	30	05 min 38 s	52.69%	0.697
4	40	07 min 47 s	60.13%	0.654
5	50	9 min 39 s	65.67%	0.648
6	60	10 min 49 s	70.67%	0.623
7	70	12 min 52 s	73.86%	0.583
8	80	15 min 27 s	70.67%	0.591
9	90	16 min 45 s	69.85%	0.603
10	100	19 min 53 s	73.86%	0.562
11	110	21 min 40 s	75.08%	0.552
12	120	23 min 36 s	77.39%	0.495
13	130	26 min 51 s	75.08%	0.561
14	140	29 min 29 s	76.45%	0.534
15	150	32 min 37 s	81.78%	0.472
16	160	35 min 54 s	81.78%	0.492
17	170	38 min 42 s	87.81%	0.432
18	180	41 min 45 s	87.81%	0.339
19	190	43 min 58 s	92.38%	0.338
20	200	45 min 59 s	92.38%	0.338

min minutes, *s* seconds

Comparative Analysis of Various Mainstream CNN Models for Benchmark Dataset

The similar arrangement of hyper-parameter setting as well as image-preprocessing is done, and the performance of various mainstream standard CNN models like AlexNet, ResNet, and VggNet is assessed considering benchmark MESSIDOR fundus image dataset. The comparison is drawn to assess the practicability of our proposed QEIRV-2 model and to establish its robustness among the other CNN counterparts. Figure 6 represents the comparison of different CNN models with the Inception network derivatives.

The comparison is evaluated for various pre-trained CNN models in terms of network accuracy, time complexity, and cross-entropy cost function. The analysis reveals that the AlextNet architecture involves a smaller number of layers and therefore takes the least computational season of 17 min 04 s as far as the execution time is concerned. However, the performance accuracy accomplished utilizing this network is 73.33% which is less when compared with different networks. The proposed QEIRV-2 yields the best network outcomes in terms of both the accuracy as well as cost parameters while involving a trade-off with the time complexity. The

time complexity however is not a big issue as the latest system is trained on GPUs; thus, the main concern is performance which is improving by our proposed strategy.

Performance Validation with the Latest IDRiD Dataset

The generalization of the proposed QEIRV-2 model is validated by the utilization of latest IDRiD dataset. To get an oversight of classification performance in terms of performance parameters, the proposed methodology has been applied on the images acquired from IDRiD dataset and the observations are made for QEIRV-2 model considering same hyper-parameter settings.

The performance outcomes obtained from Table 6 reveals that the IDRiD dataset also provides the equivalent performance for the proposed QEIRV-2 approach. An accuracy value of 92.38% is achieved while maintaining a reduced cross-entropy loss of 0.338 with the time complexity of 45 min and 59 s. The QEIRV-2 model has made the preeminent attempts to address the DR severity problem and generalizes the grading capability of the proposed method attaining better performance. The performance of IDRiD dataset follows the same trend as obtained for the MESSIDOR dataset which justifies that the proposed models yield the uniform outcomes regardless of the dataset utilized, which in turn establishes its generalization abilities.

Comparative Analysis of Proposed QEIRV-2 Model with Other Existing Methods

Inspite of the fact that the DR research is mostly centered around utilizing the machine learning approaches till date and a very little of the advancement is experienced in the CNN-based DR severity grading techniques. Regardless of this restriction, a comparative investigation of other existing state-of-the-art DR severity grading approaches has been made relative to our proposed QEIRV-2 model.

Table 7 portrays the outcomes of various CNN-based DR grading and classification methods using the

Table 7 Comparison of CNN-based existing methods with the proposed QEIRV-2 model using MESSIDOR dataset

Author	Year	Accuracy value
Wang et al. [26]	2017	91.1%
Lam et al. [23]	2018	74.5%
Johari et al. [25]	2018	88.3%
Chen et al. [27]	2018	90.5%
Goncalves et al. [28]	2019	81.6%
Li et al. [29]	2019	92.6%
Saranya and Prabakaran [30]	2020	90.9%
Proposed QEIRV-2 model	2020	93.3%

generalized MESSIDOR dataset. The comparison of our proposed method is done with the latest approach proposed by Saranya and Prabakaran [30], and it is observed that the proposed QEIRV-2 model surpasses the existing state-of-the-art method without the requirement of feature explicit detection of DR lesions. The highest percentage accuracy improvement of 25.23% is observed from Lam et al. [23] by utilizing our proposed strategy. The primary concern of future perspective for this work will be the attainability and robustness of the proposed QEIRV-2 model in real-time scenario for dealing with the DR grading issue.

Conclusion

The interest for computerized DR diagnosis framework has increased because of enormous diabetic populace and expectedness of diabetic retinopathy cases among them. The state-of-the-art review suggests that a great deal of accomplishment was seen in numerous DR-related areas like DR lesion recognition and blood vessel segmentation; however, the outcomes attained may vary from the clinical practices in real-world scenario. In this article, a computerized DR grading framework has been proposed utilizing the deep learning architecture. The proposed Quadrant-based Ensemble InceptionResnet-V2 model incorporates a pre-processing pipeline to improve the effectiveness of DR grading and evaluation. The exploratory outcomes acquired exhibit the ability of the proposed QEIRV-2 model for proficient DR determination. The proposed model yields the finest accuracy performance of 93.33% with a significantly reduced cross-entropy loss function to 0.325. The accuracy improvement of 13.58% is observed while comparing the QEIRV-2 model with the classical Inception-V3 CNN model. The comparative analysis with the other mainstream CNN models reveals the outperformance of the proposed method in terms of all the performance parameters. The validation of the QEIRV-2 is also done utilizing the latest IDRiD dataset which establishes the viability of the proposed model. An accuracy improvement of 25.23% was accomplished when compared with the state-of-the-art DR grading and classification approaches using the same generalized MESSIDOR dataset, justifying the diagnostic capability of the proposed model. In the future, the proposed framework will be utilized for giving demonstrative help to the ophthalmologists by providing a second opinion for DR grading problems, thereby, addressing the real time scenario.

Compliance with Ethical Standards

References

1. Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, Ferris FL, Klein R, "Retinopathy in diabetes." Diabetes care 27, 1:s84-s87,2004.
2. Yen GG, Leong WF, "A sorting system for hierarchical grading of diabetic fundus images: A preliminary study." IEEE Trans Inf Technol Biomed 12, 1:118-130,2008.
3. Xiao D, Bhuiyan A, Frost S, Vignarajan J, Tay-Kearney ML, Kanagasingam Y, "Major automatic diabetic retinopathy screening systems and related core algorithms: a review. " Mach Vis Appl 30, 3:423-446,2019.
4. Kandel I, and Castelli M, "Transfer Learning with Convolutional Neural Networks for Diabetic Retinopathy Image Classification. A Review. " Appl Sci 10:6,2020.
5. Deng J, Dong W, Socher R, Li LJ, Li K, and Fei-Fei L, "ImageNet: A large-scale hierarchical image database. " in IEEE Conference on Computer Vision and Pattern Recognition, 248–255,2009.
6. Somasundaram SK, and Alli P, "A machine learning ensemble classifier for early prediction of diabetic retinopathy." J Med Syst 41,12:201,2017.
7. Srivastava R, Duan L, Wong DWK, Liu J, and Wong TY, Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels. Comput Methods Prog Biomed 138:83–91,2017.
8. Seoud L, Chelbi J, and Cheriet F, Automatic grading of diabetic retinopathy on a public database. 2015.
9. Sankar M, Batri K, and Parvathi R, "Earliest diabetic retinopathy classification using deep convolution neural networks. pdf." Int J Adv Eng Technol, 2016.
10. Seoud L, Hurtut T, Chelbi J, Cheriet F, and Langlois JP, Red lesion detection using dynamic shape features for diabetic retinopathy screening. IEEE Trans Med Imaging 35, 4:1116–1126,2015.
11. Pires R, Jelinek HF, Wainer J, Goldenstein S, Valle E, and Rocha A, "Assessing the need for referral in automatic diabetic retinopathy detection." IEEE Trans Biomed Eng 60, 12:3391-3398,2013.
12. Antal B, and Hajdu A, "An ensemble-based system for microaneurysm detection and diabetic retinopathy grading." IEEE Trans Biomed Eng 59, 6:1720,2012.
13. Mansour RF, Evolutionary Computing Enriched Computer-Aided Diagnosis System for Diabetic Retinopathy: A Survey. IEEE Rev Biomed Eng 10:334–349,2017.
14. Dai L, Fang R, Li H, Hou X, Sheng B, Wu Q, and Jia W, "Clinical Report Guided Retinal Microaneurysm Detection With Multi-Sieving Deep Learning." IEEE Trans Med Imaging 37, 5:1149-1161,2018.
15. Rahim SS, Jayne C, Palade V, and Shuttleworth J. Automatic detection of microaneurysms in colour fundus images for diabetic retinopathy screening. Neural Comput Applic 27, 5:1149–1164,2016.
16. Mohammed ZF, and Abdulla AA. Thresholding-based White Blood Cells Segmentation from Microscopic Blood Images. UHD J Sci Tech 4, 1:9–17,2020.
17. Abbas Q, Fondon I, Sarmiento A, Jiménez S, and Alemany P, "Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features." Med Biol Eng Comput 55, 11:1959-1974,2017.
18. Wang Z, and Yang J, Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. arXiv preprint arXiv 1703.10757, 2017.

19. Yu FL, Sun J, Li A, Cheng J, Wan C, and Liu J, Image quality classification for DR screening using deep learning. In 2017 39th Conf Proc IEEE Eng Med Biol Soc (EMBC), 664–667. IEEE, 2017.

20. Gao Z, Li J, Guo J, Chen Y, Yi Z, and Zhong J, "Diagnosis of Diabetic Retinopathy Using Deep Neural Networks." IEEE Access 7:3360-3370,2018.

21. Mateen M, Wen J, Song S, and Huang Z, "Fundus Image Classification Using VGG-19 Architecture with PCA and SVD." Symmetry 11, 1:1,2019.

22. Li X, Pang T, Xiong B, Liu W, Liang P, and Wang T, Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 1–11. IEEE, 2017.

23. Lam C, Yi D, Guo M, and Lindsey T, Automated detection of diabetic retinopathy using deep learning. AMIA Summits on Translational Science Proceedings 147, 2018.

24. Perdomo O, Otalora S, Rodr´ıguez F, Arevalo J, and Gonz´alez FA, A novel machine learning model based on exudate localization to detect diabetic macular edema, 2016.

25. Hazim JM, et al: "Early Detection of Diabetic Retinopathy by Using Deep Learning Neural Network", International Journal of Engineering & Technology 7, 411:1997- 2004,2018.

26. Wang Z, Yin Y, Shi J, Fang W, Li H, and Wang X, Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 267–275. Springer, Cham, 2017.

27. Chen YW, Wu TY, Wong WH, Lee CY, Diabetic retinopathy detection based on deep convolutional neural networks. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 1030–1034,2018.

28. Gonçalves J, Conceiçao T, Soares F, Inter-observer Reliability in Computer-aided Diagnosis of Diabetic Retinopathy. In HEALTHINF. 481–491,2019.

29. Li X, Hu X, Yu L, Zhu L, Fu CW, Heng PA, "CANet: Cross-Disease Attention Network for Joint Diabetic Retinopathy and Diabetic Macular Edema Grading. " IEEE Trans Med Imaging 39. 5:1483-93,2019.

30. Saranya P, Prabakaran S, Automatic detection of non-proliferative diabetic retinopathy in retinal fundus images using convolution neural network. J Ambient Intell Humaniz Comput, 2020.

31. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, and Van Ginneken B, Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imaging 23, 4:501–509,2004.

32. Hoover AD, Kouznetsova V, and Goldbaum M, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans Med Imaging 19, 3:203–210,2000.

33. Kauppi T, Kalesnykiene V, Kamarainen JK, Lensu L, Sorri I, Raninen A, Voutilainen R, Uusitalo H, Kälviäinen H, and Pietilä J, "The diaretdb1 diabetic retinopathy database and evaluation protocol." In BMVC 1:1-10,2007.

34. Niemeijer M, Van Ginneken B, Cree MJ, Mizutani A, Quellec G, Sánchez CI, Zhang B, et al: Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans Med Imaging 29, 1:185–195,2009.

35. Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, and Charton B. "Feedback on a publicly distributed image database: the Messidor database. " Image Analysis & Stereology 33, 3:231-234,2014.

36. Porwal P, Pachade S, Kamble R, Kokare M, Deshmukh G, Sahasrabuddhe V, and Meriaudeau F, Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. Data 3, 3:25,2018.

37. Bhardwaj C, Jain S, and Sood M, Appraisal of Pre-processing Techniques for Automated Detection of Diabetic Retinopathy. In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), 734–739. IEEE,2018.

38. Janney BJ, Divakaran S, Abraham S, Uma SG, Meera G, Shankar G, Detection and classification of exudates in retinal image using image processing techniques. J Chem Pharm Sci 8(3):541–546. 2015.

39. Akyol K, Şen B, and Bayır Ş, Automatic detection of optic disc in retinal image by using keypoint detection, texture analysis, and visual dictionary techniques. Comput Math Methods Med 2016, 2016.

40. Bhardwaj C, Jain S, Sood M, Automated Optical Disc Segmentation and Blood Vessel Extraction for Fundus Images Using Ophthalmic Image Processing, In International Conference on Advanced Informatics for Computing Research, 182–194. Springer, Singapore, 2018.

41. Bajwa MN, Malik MI, Siddiqui SA, Dengel A, Shafait F, Neumeier W, Ahmed Sheraz, Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning, BMC Med Inform Decis. Mak 19, 1:136, 2019.

42. LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE, Hubbard WE, Jackel LD, Handwritten digit recognition with a back-propagation network. In Advances in neural information processing systems, 396–404. 1990.

43. Xu K, Feng D, Mi H, "Deep convolutional neural network-based automated detection of diabetic retinopathy using fundus image," Molecules 22, 12:2054,2017.

44. Mohammadian S, Karsaz A, Roshan YM, 2017, November. Comparative Study of Fine-Tuning of Pre-Trained Convolutional Neural Networks for Diabetic Retinopathy Screening, In 2017 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME), 1–6, IEEE, 2017.

45. LeCun Y, Bottou L, Bengio Y, Haffner P, "Gradient-based learning applied to document recognition." Proc IEEE 86, 11:2278-2324,1998.

46. Krizhevsky A, Sutskever I, Hinton GE, Imagenet classification with deep convolutional neural networks. In Adv Neural Inf Proces. Syst, 1097–1105,2012.

47. Simonyan K, Zisserman A, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv 1409.1556, 2014.

48. He K, Zhang X, Ren S, Sun J, Deep residual learning for image recognition. In Proc IEEE Conf Comput Vis Pattern Recognit, 770–778,2016.

49. Szegedy C, Ioffe S, Vanhoucke V, Alemi A, Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv 1602.07261, 2016.