Manchester Metropolitan University

e-space

Manchester Metropolitan University's Research Repository

Please cite the Published Version

Nahiduzzaman, Md, Robiul Islam, Md, Omaer Faruq Goni, Md, Shamim Anower, Md, Ahsan, Mominul, Haider, Julfikar and Kowalski, Marcin (2023) Diabetic Retinopathy Identification Using Parallel Convolutional Neural Network Based Feature Extractor and ELM Classifier. Expert Systems with Applications, 217. p. 119557. ISSN 0957-4174

DOI: https://doi.org/10.1016/j.eswa.2023.119557

Publisher: Elsevier

Version: Published Version

Downloaded from: https://e-space.mmu.ac.uk/631197/

Usage rights: CC BY-NC-ND Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Additional Information: This is an Open Access article published in Expert Systems with Applications, by Elsevier.

Data Access Statement: The authors do not have permission to share data.

Enquiries:
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

Expert Systems With Applications 217 (2023) 119557

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Diabetic retinopathy identification using parallel convolutional neural network based feature extractor and ELM classifier

Md. Nahiduzzaman [a,f], Md. Robiul Islam [a], Md. Omaer Faruq Goni [a], Md. Shamim Anower [b], Mominul Ahsan [c], Julfikar Haider [d], Marcin Kowalski [e,*]

[a] Department of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[b] Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[c] Department of Computer Science, University of York, Deramore Lane, Heslington, York YO10 5GH, UK
[d] Department of Engineering, Manchester Metropolitan University, Chester St, Manchester M1 5GD, UK
[e] Institute of Optoelectronics, Military University of Technology, Gen. S. Kaliskiego 2, 00-908 Warsaw, Poland
[f] Department of Electrical Engineering, Qatar University, Doha 2713, Qatar

ARTICLE INFO

Keywords:
Contrast limited adaptive histogram equalization (CLAHE)
Diabetic retinopathy (DR)
Parallel convolutional neural network (PCNN)
Extreme LEARNING MACHine (ELM)

ABSTRACT

Diabetic retinopathy (DR) is an incurable retinal condition caused by excessive blood sugar that, if left untreated, can result in even blindness. A novel automated technique for DR detection has been proposed in this paper. To accentuate the lesions, the fundus images (FIs) were preprocessed using Contrast Limited Adaptive Histogram Equalization (CLAHE). A parallel convolutional neural network (PCNN) was employed for feature extraction and then the extreme learning machine (ELM) technique was utilized for the DR classification. In comparison to the similar CNN structure, the PCNN design uses fewer parameters and layers, which minimizes the time required to extract distinctive features. The effectiveness of the technique was evaluated on two datasets (Kaggle DR 2015 competition (Dataset 1; 34,984 FIs) and APTOS 2019 (3,662 FIs)), and the results are promising. For the two datasets mentioned, the proposed technique attained accuracies of 91.78 % and 97.27 % respectively. However, one of the study's subsidiary discoveries was that the proposed framework demonstrated stability for both larger and smaller datasets, as well as for balanced and imbalanced datasets. Furthermore, in terms of classifier performance metrics, model parameters and layers, and prediction time, the suggested approach outscored existing state-of-the-art models, which would add significant benefit for the medical practitioners in accurately identifying the DR.

1. Introduction

Diabetic retinopathy (DR) is a chronic retinal disease that is regarded as the sixth most common cause of blindness worldwide. It's a hidden progressive chronic disease among the diabetic patients. According to the 2013 statistics, 382 million people are affected by diabetes-related retinal disease, and by 2025, it is projected to exceed 592 million (Pandey & Sharma, 2018). DR shows no clear early sign of appearance; as the condition degrades, complete blindness is basically the obvious end result. Regular screening can help to identify the DR at an early stage, which can help in arresting any further damage through appropriate medication. Fundus images (FIs) with high resolution are utilized for detecting the teensy lesions and grading the severity level. Non-proliferative DR (NPDR) and proliferative DR (PDR) are the two primary forms of the DR. Again, NPDR can be classified with four severity levels: No DR, Mild stage, Moderate stage, and Severe stage (Mumtaz et al., 2018). Fig. 1 reveals some common symptoms of the DR (Mumtaz et al., 2018). The small dark reddish dot-like lesion is visible near the blood vessel's terminal point, called a microaneurysm (MA). Hypertension and blockage of the retinal veins cause retinal hemorrhage (HM), another DR consequence. Small HMs might look a lot similar to the MAs at times. Exudates are yellow flicks that filter out the injured capillaries and are made up of lipids and protein residues.

In its later phases, the DR is difficult to treat. There are only a few microaneurysms that appear in the Mild NPDR. In contrast, multiple MAs, hemorrhages, and venous beading occur in the moderate NPDR, leading patients' capacity to transfer blood to the retina to be compromised. Severe NPDR is defined by the appearance of more than 20 intra-

* Corresponding author.
E-mail addresses: md.ahsan2@mail.dcu.ie (M. Ahsan), J.Haider@mmu.ac.uk (J. Haider), marcin.kowalski@wat.edu.pl (M. Kowalski).

https://doi.org/10.1016/j.eswa.2023.119557
Received 15 February 2022; Received in revised form 2 December 2022; Accepted 12 January 2023
Available online 13 January 2023
0957-4174/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

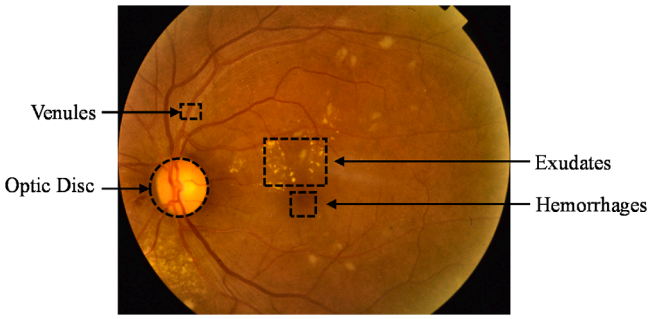

Fig. 1. Fundus image with various lesions for DR classification.

retinal hemorrhages in each of the four quadrants, visible venous beading in two or more quadrants, and substantial intraretinal microvascular abnormality (IRMA) in one or more quadrants. New blood vessels are formed in the PDR stage, along with the aforementioned anomalies (Chudzik et al., 2018).

DR is diagnosed using fundus images. Expert ophthalmologists find existing lesions on the images based on which they grade the DR level and suggest appropriate treatment accordingly. As the lesions are small and often having an overlapping boundaries between the consecutive DR grades, even the expert ophthalmologists cannot provide consistent diagnosis for the same fundus images and it is also a time-consuming process. Therefore, an urgent need for a computer-aided system has been realized by the research community.

Various computer-aided systems have been proposed so far for the DR screening. Ophthalmologists grade the severity level by screening the lesions present in FIs and providing treatment based on the level. Some lesion segmentation techniques were developed to copy this style to mark out these tiny lesions and assist the ophthalmologists in correct diagnosis. Image processing techniques were frequently used for segmenting lesions of FIs. Using image processing techniques, Mumtaz et al. (2018), showed the automatic identification of one of the red lesions, i.e., hemorrhage, which is one of the most recognizable symptoms of retinal disorders among diabetic patients. Akram et al. (2014), detected the MA from small patches extracted from the FIs while PCA was used for dimensionality reduction. Rahim et al. (2016), used fuzzy C-means (FCM) image processing techniques to provide a novel automated diagnosis of the DR and maculopathy in eye fundus pictures. Kar and Maity (2017), developed a four-part lesion detection technique that included extraction of vessels and removal of the optic disc, preprocessing, detection of candidate lesion, and post-processing. The dark lesions were separated from the weakly lit retinal backgrounds using curvelet-based edge enhancement, while the contrast between the bright lesions and the background was improved using a well-designed wideband bandpass filter. Subsequently, the mutual information of the maximum matched filter response and the maximum Laplacian of Gaussian response was maximized together. Finally, morphology-based post-processing was used to exclude the candidate pixels that were incorrectly identified. Umapathy et al. (2019), extracted texture features using the image processing and classified by Decision Tree (DT) classifier. For the second method the authors utilized the transfer learning method. As the complex features were extracted using the image processing technique, the accuracy was not so high. For this, deep learning models were also proposed for the lesion segmentation. For the segmentation of microaneurysms, Chudzik et al. (2018), presented a patch-based Convolutional Neural Network (CNN) with batch normalization layers and a dice loss function Pixel-wise exudate detection with a deep CNN was proposed by Yu et al. (2017). Gondal et al. (2017), presented a weakly-supervised CNN model that highlighted denoting regions of the retinal images. The authors obtained high classification and sensitivity scores. The Mask-RCNN model was proposed to segment small lesions (MA and exudates) by Shenavarmasouleh and Arabnia (2007). The authors utilized the transfer learning (TL) approach to reuse the pre-trained model ResNet101′s weights and achieved an mAP score of 45 %. Besides segmentation, image-level classification is also popular for the DR grading. The whole image is classified into its classification grades based on unique features in the image-level classification.

Several of the studies utilized traditional machine learning (ML) methods such as DT, support vector machine (SVM), Random forests (RF), logistic regression (LR), and Gaussian Naïve Bayes (GNB). For using traditional ML-based classification, features were extracted using image processing techniques later deployed to develop the models. For example, Lachure et al. (2015), used morphological image processing like erosion, dilation, opening, closing, etc., to segment MAs and exudates. Later the features were fed to the SVM and k-nearest neighbors (KNN) classifiers for grading the FIs. Asha and Karpagavalli (2015), detected retinal exudates using machine learning techniques where the FIs were segmented using the fuzzy C means algorithm, then exudates features were detected from the Luv color space. The classifiers utilized included NB, Multilayer Perceptron (MLP), and Extreme Learning Machine (ELM), with ELM providing the best results. ML techniques for automatically identifying and categorizing the DR from the retina images were studied by Honnungar et al. (2016). The proposed method entailed image preprocessing (Contrast Limited Adaptive Histogram Equalization, CLAHE), feature extraction using the bag of visual words model, and image classification into distinct DR phases using a multiclass classifier (logistic regression, SVM, and RF). Raman et al. applied CLAHE to enhance the images, then Sobel operator and contour with circular hough transformation for optic disk segmentation, morphological operation for blood vessel segmentation, regions growing for exudates segmentation, and a mixture model for microaneurysm segmentation (Raman et al., 2016). Finally, an artificial neural network (ANN) was used as a classifier. Carrera et al. (2017), utilized image processing to isolate blood vessels, microaneurysms, and hard exudates for extracting features, which were later deployed to the SVM classifier. They obtained a sensitivity of 95 % and an accuracy of 94 %. Somasundaram and Ali (2017), developed a ML bagging ensemble classifier (ML-BEC) and extracted *t*-distribution Stochastic Neighbor Embedding (*t*-SNE) features. Ramani et al. (2017), proposed a two-level classification for the DR grading. Ensemble of Best First Trees (BFTs) was used, whereas misclassified instances were removed and deployed to second level ensemble classifiers with J48 Graft Trees. Using Local Ternary Pattern (LTP) and Local Energy-based Shape Histogram, Chetoui et al. (2018), identified texture characteristics (LESH). For classification, SVM was used with various kernel functions. For feature representation, a histogram binning method was utilized. They demonstrated that using SVM with an RBF kernel, LESH is the best method, with an accuracy of 90 %. ML approaches for segmentation and categorization of the DR were presented by Ali et al. (2020). They proposed a new regional-growing paradigm based on clustering. They used four types of characteristics for texture analysis: histogram (H), wavelet (W), co-occurrence matrix (COM), and run-length matrix (RLM). The authors utilized data fusion to create hybrid-feature datasets to increase classification accuracy. To obtain 13 optimal features, they used Fisher, correlation-based feature selection, mutual information, and probability of error plus average correlation. Finally, five classifiers were used: SMO (sequential minimum optimization), Lg (logistic), MLP (multilayer perceptron), and SLg (simple logistic). Gayathri et al. (2021), designed a multipath convolutional neural network (M−CNN) for extracting global and local features from fundus images. Then SVM, RF, and J48 classifiers were used for the final DR grade prediction. The M−CNN network obtained the best result with the J48 classifier. Mahmoud et al. (2021), introduced a hybrid inductive ML algorithm (HIMLA) for automatic DR detection.

Color FIs were normalized and a convolutional encoder-decoder was used for segmenting blood vessels. A multiple instance learning technique was utilized for feature extraction and classification. Reddy et al. (2020), experimented with an ensemble learning method with

Adaboost, RF, DT, KNN, and Logistic Regression. The authors used the grid search technique for hyperparameter tuning. Odeh et al. (2016), proposed an ensemble method using RF for robust and powerful learning, NN for improving precision, and SVM for accurate, time-saving prediction. For feature selection, the authors used info gain attribute evaluation and wrapper subset evaluation algorithms.

One problem with the traditional ML is that the complex features need to be extracted first. This manual feature extraction using image processing sometimes fail to capture all the complex features necessary for an accurate classification. Here comes the deep learning (DL) approach, which is used for imaging in a wide range of applications nowadays. DL models were also deployed in the DR identification with significant success through accurate extraction of the complex feature using the convolution layers. A 4 × 4 kernel-based CNN architecture with some preprocessing and augmentation methods was proposed by Islam et al. (2018), for detecting the DR where the authors employed L2 regularizer and dropout to eliminate overfitting and achieved 98 % sensitivity and 94 % specificity with a kappa score of 85 %. Zhou et al. (2018), proposed a multitasking deep learning model for the DR grading. Because of the interrelationship among the DR stages, the authors followed the multitasking approach that predicted the labels with both the classification and regression and got a kappa score of 84 %. A Siamese-like architecture was also proposed for the DR detection by Zeng et al. (2019). The model used binocular fundus images as input and was trained with a transfer learning strategy. An attention-based DL model, BiRA-Net was proposed by Zhao et al. (2019). Islam et al. (2020), proposed a VGG16 based transfer learning approach with a color preprocessing version. The authors used stratified K-fold cross-validation to reduce the overfitting problem. For a smaller Kaggle dataset, Samanta et al. (2020), suggested transfer learning-based DenseNet and attained a kappa score of 0.8836 on the validation set. On the Messidor-1 and APTOS datasets, Gangwar and Ravi (2021), used a pre-trained model, Inception-ResNet-v2, and built a custom layer on top, achieving an accuracy of 72.33 % and 82.18 % respectively. Islam et al. (2021), developed a customized VGG19 model and down sampling technique for DR detection. Majumder and Kehtarnavaz (2021), proposed a multitasking deep learning model to detect the five grades of the DR composed of one regression model, one classification model, and one regression model for inter-dependency. For the APTOS and EyePACS datasets, they achieved a kappa score of 90 % and 88 %. Also, an integrated shallow network was proposed by Chen et al. (2020).

Though various models have been developed, still further improvement is required particularly in the case of multiclass classification. Several ML models were employed in some research, but in this case, the classification performance was not satisfactory despite the model complexity being lower than the existing DL models. Researchers used different transfer learning (TL) models to achieve higher classification performance to overcome these shortcomings. However, the TL models have a vast number of parameters, layers and consume a lot of time for training. Therefore, this study proposes a framework that makes a trade-off between the ML and DL models, increasing classification performance and reducing the vast number of parameters and layers, which reduces the processing time. In this study, the FIs were preprocessed using CLAHE to highlight the lesions of DR. A lightweight parallel CNN model has been developed to extract the most discriminant features, which are standardized using a standard scaler. Finally, a single-layer ML algorithm model named ELM has been used for classification of the DR. The proposed framework brings its novelty through a smaller number of parameters, layers, and comparatively lower processing time. The proposed framework also offers versatile capabilities in any domain, for instance, small or large datasets, balanced or imbalanced datasets, and low-resolution FIs.

2. Dataset description

In this study, two prevalent datasets were used: Kaggle DR 2015 competition (Dataset 1) and APTOS, 2019 respectively provided by EyePACS and Aravind Eye Hospital via Kaggle (California Healthcare Foundation, 2019; APTOS, 2019). The datasets contained five grades of the DR to detect with 34,984 FIs in Dataset 1 and 3,662 images in APTOS, 2019. 80 % of the data was used for training, and the rest was for testing. During image extraction from the Kaggle DR 2015 dataset, some FIs were lost. As both the datasets were collected from Kaggle competition, their corresponding test images were kept in private. Hence, only the trained data was used for the DR classification. The trained dataset then further split into both training and testing set for carried out the classification task. Table 1 shows the number of FIs per class for both datasets. Representative samples from each class are demonstrated in Fig. 2.

3. Proposed framework

An adequate framework was proposed in this study for severity grading of the DR. The benefits of ML and DL algorithms were merged to develop a robust framework with a trade-off between the model's processing performance and classification performance. Fig. 3 exhibits the proposed framework to detect DR from the FIs. First, the FIs were preprocessed using CLAHE to highlight the lesions more clearly, then normalized and finally reshaped. Afterward, a lightweight CNN model was developed to extract the most discriminant features from the processed FIs. The extracted features were standardized to be fed into the ELM algorithm, which to classify the severity level of the DR. In the subsequent sections, all components of the framework have been explained comprehensively.

3.1. Pre-processing

Image preprocessing is crucial for medical image analysis because the classification performance varies depending on how well the image has been preprocessed. CLAHE reveals a favorable result for enhancing image quality in the case of medical image preprocessing (Nahiduzzaman et al., 2021a,b). Since the datasets contained different quality of images, hence for improving the quality of low contrast images while focusing on the lesions of FIs, CLAHE was utilized. The intensification in CLAHE was controlled by clipping the histogram at a user-defined value called the clip limit. The clipping level determined the amount of distortion in the histogram should be eliminated and this defined the limit of contrast adjustment. In this study, the tile size was (4 × 4), and the clip limit was 2.0 while using the color version of the CLAHE. After applying CLAHE, the FIs have been normalized dividing by 255 to make each image range between 0 and 1, which also reduced the complexity of the model. Since the datasets contained diverse FIs, making the FIs with the same size was an essential step to follow. Hence, the FIs were resized to (124 × 124) to fit into the CNN model. Fig. 4 shows the effect of CLAHE in the FIs.

3.2. Features extraction using parallel convolutional layers

One of the main focuses of this study was to design a CNN that reduced both parameters and layers, which eventually shortened the processing time while extracting the most prominent features. The

Table 1
The number of FIs per class for Dataset-1 and APTOS, 2019.

Level	Dataset-1 (Image Ratio)	APTOS, 2019 (Image Ratio)
No DR	25,707 (0.73)	1,805 (0.49)
Mild DR	2,435 (0.07)	370 (0.10)
Moderate DR	5,268 (0.15)	999 (0.27)
Severe DR	869 (0.025)	193 (0.05)
PDR	705 (0.02)	295 (0.08)
Total	34,984	3,662

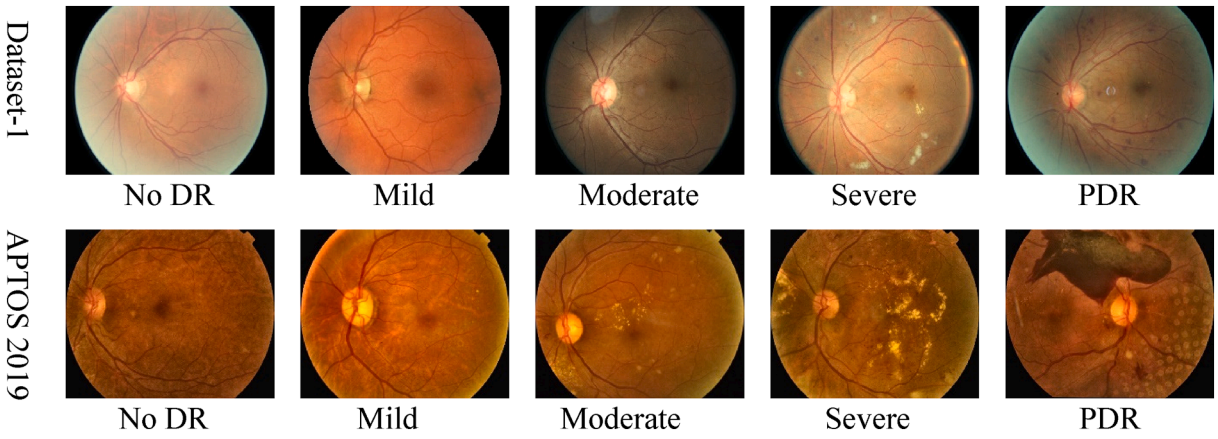

Fig. 2. Samples of No DR, Mild, Moderate, Severe, and PDR from Dataset-1 and APTOS, 2019.

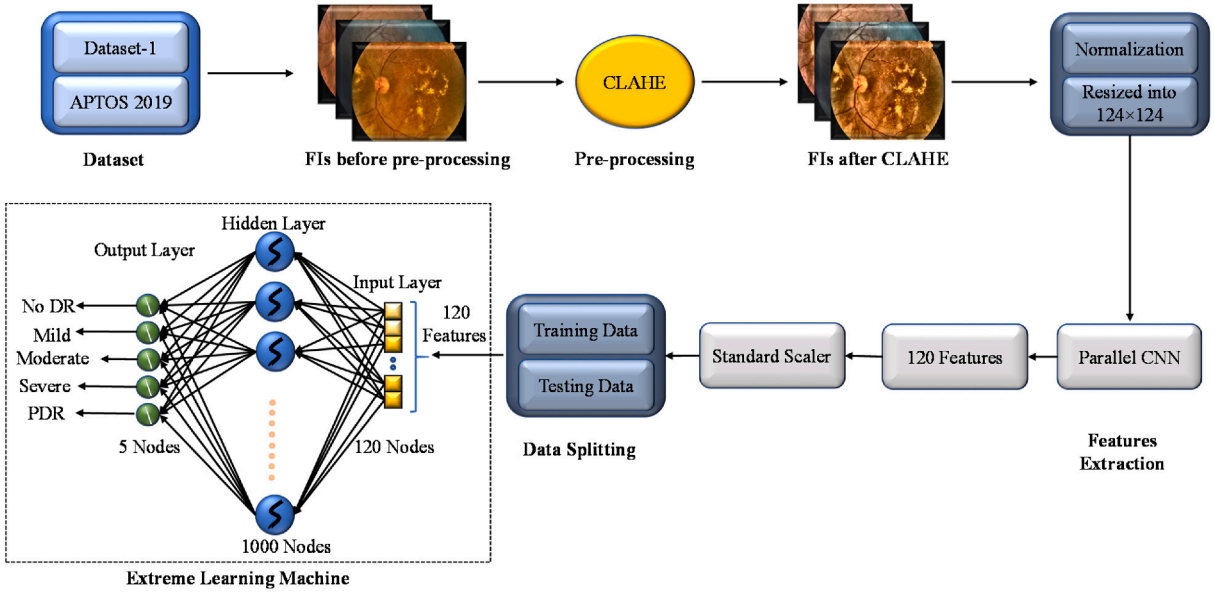

Fig. 3. A proposed framework to detect the five levels of DR.

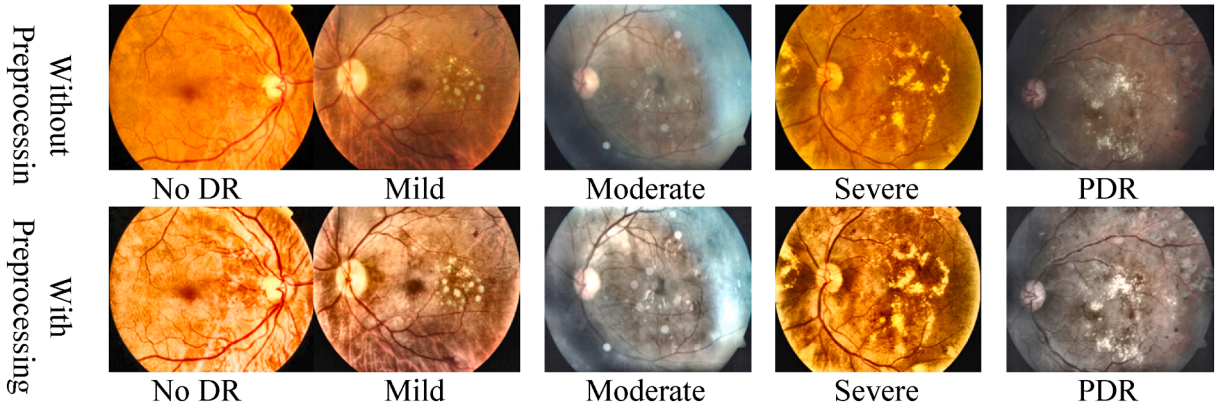

Fig. 4. Five levels of FIs without preprocessing and preprocessing with CLAHE.

notable features assisted the ELM model in accurately detecting the levels of DR. Basically, in CNN, the convolutional layer (CL) was positioned sequentially for obtaining the best features. For instance, selecting a small number of CL layers might result in the loss of some discriminant features, whereas a large number of CL layers might lead to overfitting the model. Hence, the number of CL layers needed to be chosen adequately to extract the most relevant features. In this study, six CL layers were selected to extract the prominent features while reducing

overfitting. The lightweight parallax CNN has been shown in Fig. 5.

In the lightweight parallel CNN, four CLs were placed in parallel, which resulted in lowering the parameters and processing time. Since the four CLs were run in parallel, which could be considered as a single CL but performed just like four CLs. The size of each CL was 64. The kernel sizes of the first, second, third, and fourth CLs were 9×9, 7×7, 5×5, and 3×3, respectively and the activation function was ReLU. In this study, the padding size was kept the same in the first four CLs to check the border element. As sometimes the border element might hold important information in the FIs which were checked using the same padding. Afterwards, the result of these parallel CLs were concatenated and fed into the sequential CNN. The sizes of the last two CLs were 32 and 16, respectively, with a kernel size of 3×3. The padding size in the rest of the CLs was kept "valid". Each CL was followed by batch normalization, activation, and a max-pooling layer. Max-pooling with 2×2 filters was used to extract the most important regions of the FIs by obtaining the highest value in each region at the CLs. There were two fully connected (FC) layers, and the features were extracted from the last FC layer. Two dropouts were used with a 0.5 probability: one after the last CL and another after the first FC layer. Dropout was used to reduce overfitting and speed up the training process by randomly skipping 50 % of all nodes. For extracting the features, the CNN model was run for 50 epochs with a batch size 64 while considering the learning rate of 0.001 with the ADAM optimizer and handling the loss using sparse categorical cross-entropy. A total of 120 features were selected from the last FC layer by using a trial-and-error process. The summary of the CNN model is shown in Table 2.

Table 2
Summary of proposed lightweight CNN for feature extraction.

Layer (Type)	Output Shape	Parameters
model (Functional)	(None, 124, 124, 256)	31, 744
conv5 (Conv2D)	(None, 122, 122, 32)	73, 760
bn1 (BatchNormalization)	(None, 122, 122, 32)	128
av5 (Activation)	(None, 122, 122, 32)	0
mp1 (MaxPooling2D)	(None, 61, 61, 32)	0
conv6 (Conv2D)	(None, 59, 59, 16)	4, 624
bn2 (BatchNormalization)	(None, 59, 59, 16)	64
av2 (Activation)	(None, 59, 59, 16)	0
mp2 (MaxPooling2D)	(None, 29, 29, 16)	0
dp1 (Dropout)	(None, 29, 29, 16)	0
ft (Flatten)	(None, 13456)	0
dense (Dense)	(None, 250)	3, 364, 250
bn4 (BatchNormalization)	(None, 250)	1, 000
av4 (Activation)	(None, 250)	0
dp2 (Dropout)	(None, 250)	0
Feature Extraction (Dense)	(None, 120)	30, 120
Total Parameters	3, 506, 775	
Trainable Parameters	3, 505, 939	
Non-trainable Parameters	836	

3.3. Extreme learning machine

Before fitting the features into ELM, features were standardized by subtracting the mean and scaling to mean–variance. The standard scaler was employed to regularize the extracted features, which improved the classification performance of the models (), (Nahiduzzaman et al., 2019). The standard score for the sample x has been calculated using Eq. (1) (Farrell and Saloner, 1985).

$$y = \frac{x - \underline{x}}{\sigma} \tag{1}$$

where $\underline{x}$ is the mean of the samples and σ is the standard deviation of the samples.

Huang et al. (2006), proposed ELM, a forward feed network-based neural network. The standardized 120 features were classified using a single hidden layer. The number of nodes in the hidden layer for Dataset-1 and APTOS, 2019 were 1000 and 200, respectively, which were selected by trial-and-error method. The number of nodes in the input and output layers of the ELM model for both datasets were 120 and 5, respectively, whereas the ReLU was used as an activation function. Due to the absence of backpropagation, the training time was a thousand times faster than the typical NN, resulting in better generalization power and higher classification performance (Huang et al. (2006); Nahiduzzaman et al., 2021a,b). The parameters from the input to the hidden layer were calculated randomly, whereas the parameters from the hidden layer to the output layer were calculated using pseudoinverse. For extracting features using lightweight CNN, the entire trainable parameters for the DR classification are 3, 505,939. For classification using Dataset-1 and APTOS, 2019, the complete parameters of the ELM were 125,500, and 25,000, resulting in total trainable parameters of 3, 630, 939, and 3,530, 939, respectively.

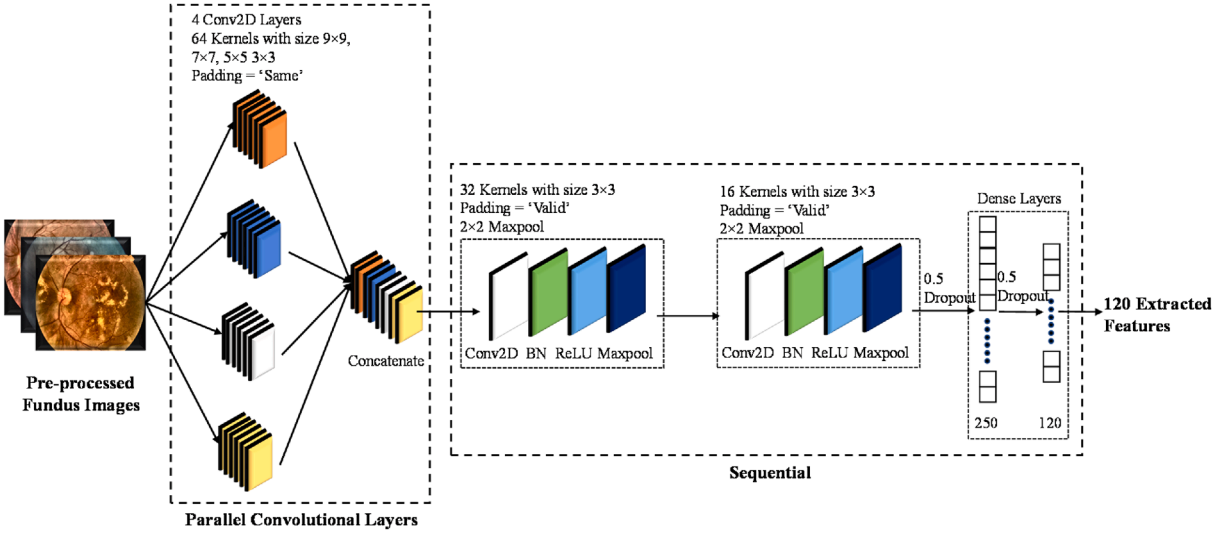

Fig. 5. The lightweight parallel CNN to extract the features from FIs.

Algorithm 1: Extreme Learning Machine

$$X_{(n,m)} = \begin{bmatrix} x_{(1,1)} & x_{(1,2)} & \cdots & x_{(1,m)} \\ x_{(2,1)} & x_{(2,2)} & \cdots & x_{(1,m)} \\ x_{(3,1)} & x_{(3,2)} & \cdots & x_{(1,m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(n,1)} & x_{(n,2)} & \cdots & x_{(n,m)} \end{bmatrix} Y_{(n,t)} = \begin{bmatrix} y_{(1,1)} & y_{(1,2)} & \cdots & y_{(1,t)} \\ y_{(2,1)} & y_{(2,2)} & \cdots & y_{(1,t)} \\ y_{(3,1)} & y_{(3,2)} & \cdots & y_{(1,t)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{(n,1)} & y_{(n,2)} & \cdots & y_{(n,t)} \end{bmatrix}$$

1: Randomly generates the input weight $W_{(m,N)}$ and bias $B_{(1,N)}$ matrix.

$$W(m,N) = \begin{bmatrix} w_{(1,1)} & w_{(1,2)} & \cdots & w_{(1,N)} \\ w_{(2,1)} & w_{(2,2)} & \cdots & w_{(1,N)} \\ w_{(3,1)} & w_{(3,2)} & \cdots & w_{(1,N)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{(m,1)} & w_{(m,2)} & \cdots & w_{(m,N)} \end{bmatrix}$$

$$B(1,N) = \begin{bmatrix} b_{(1,1)} & b_{(1,2)} & \cdots & b_{(1,N)} \end{bmatrix}$$

2: Determine the output $H_{(n,N)}$ of the hidden layer.

$$H_{(n,N)} = G(X_{(n,m)} \cdot W_{(m,N)} + B_{(1,N)})$$

$$H(n,N) = \begin{bmatrix} h_{(1,1)} & h_{(1,2)} & \cdots & h_{(1,N)} \\ h_{(2,1)} & h_{(2,2)} & \cdots & h_{(1,N)} \\ h_{(3,1)} & h_{(3,2)} & \cdots & h_{(1,N)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{(n,1)} & h_{(n,2)} & \cdots & h_{(n,N)} \end{bmatrix}$$

3: Determine the output weight matrix $\beta_{(N,t)}$

$$\beta_{(N,t)} = H^{\dagger}_{(N,n)} \cdot T_{(n,t)}$$

4: Make prediction using $\beta_{(N,t)}$

4. Result and discussion

Several performance metrics, such as accuracy, precision, recall, f1-score, and Area Under the Curve (AUC) curve, were used to evaluate the performance of the proposed framework. Equations (2) through Equation (6) can be used to define the metrics (Powers, 2010).

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{2}$$

$$Precision = \frac{T_P}{T_P + F_P} \tag{3}$$

$$Recall = \frac{T_P}{T_N + F_P} \tag{4}$$

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{5}$$

$$AUC = \frac{1}{2}\left(\frac{T_P}{T_P + F_N} + \frac{T_N}{T_N + F_P}\right) \tag{6}$$

where true positives, true negatives, false positives, and false negatives are symbolized as T_P, T_N, F_P and F_N, respectively. True positives indicated that the normal patients were correctly detected as normal, true negatives indicated that the DR affected patients were correctly identified as DR whereas false positives indicated that the normal patients were wrongly detected as DR and false negatives indicated that the DR patients were wrongly detected as normal.

PyCharm Community Edition (2021.2.3) software was used to run all of the codes, which were written in the python programming language. Keras was used to build the CNN model, with TensorFlow as the backend. The ELM models were trained and tested on a PC with a 64-bit Windows 10 Pro operating system, an Intel (R) Core (TM) i9-11900 CPU @ 2.50 GHz, 32 GB of RAM, and an NVIDIA GeForce, RTX 3090 24 GB GPU.

In this section, the different types of performance were investigated to show the robustness of the proposed framework. A lightweight customized CNN has extracted 120 prominent features from the preprocessed FIs. These prominent features were further preprocessed and fitted into the ELM model to classify different levels of DR. In abridgement, the feature deriving capability was incorporated with the ELM. The proposed combination was examined with two datasets.

4.1. Results of Dataset-1

The ELM model was trained using 27,978 FIs, whereas the numbers of No DR, Mild, Moderate, Severe, and PDR FIs were 20566, 1948, 4214, 695, and 564 respectively. The training process required only one iteration as there was no backpropagation in the ELM. Therefore, the ELM training process was faster than the traditional neural network (NN) and the DL models. Another point that needs to be noted was that to classify the DR levels correctly, a number of iterations needs to be carried out to train the NN and DL models. However, in this study, the proposed ELM achieved a promising result for only one epoch for both the datasets. After completing the training, 6,997 FIs (No DR: 5141, Mild: 487, Moderate: 1054, Severe: 174, and PDR: 141) were employed for assessing the classification performance of the ELM model. The CM obtained by the ELM for Dataset-1 is shown in Fig. 6. Clearly, in the case moderate level, misclassified number of images were much higher than the other levels.

The average precision, recall, f1-score, and accuracy of the ELM for dataset-1 were 0.91, 0.83, 0.87, and 91.78 %, respectively, as shown in Table 3. Furthermore, to demonstrate the superior performance of ELM in this study, five well-known ML algorithms such as SVM, GNB, RF, DT and LR were also employed to obtain the classification results as presented in Tables 3–5 and Fig. 7. The best classification results were obtained from SVM among these five models. The average precision, recall, f1-score, and accuracy of the SVM were 0.58, 0.44, 0.49, 75.83 % respectively which were also quite lower than ELM. In fact, SVM produced good results during the binary classification whereas NN models showed good results for multiclass classifications (Nahiduzzaman et al., 2019). As ELM is like traditional NN except the back-propagation algorithm and for that reason ELM is faster and the rate of learning and generalization are more effective. This provides promising results in the case of multiclass classifications (Afza et al., 2021; Alenezi et al., 2023).

The average AUC of the ELM for the Dataset-1 was 95.08 %, whereas the class-wise AUCs of the ELM are demonstrated in Fig. 7. It was observed that each class contributed almost equally to the final classification result (AUC values for all classes higher than 92 %). It could be concluded that though the class distribution was imbalanced, the proposed framework showed its consistency in detecting every class of DR.

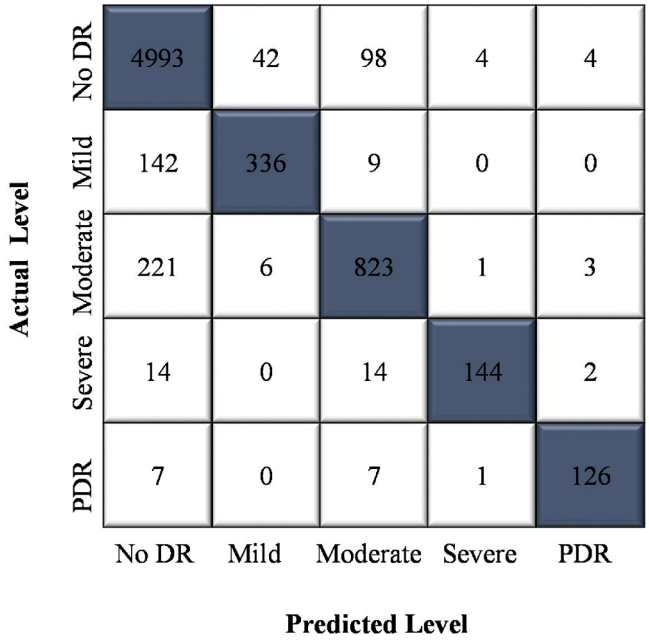

Fig. 6. Confusion Matrix (CM) of ELM for Dataset-1.

Table 3
Classification performance comparison by Precision for Dataset-1.

DR Level	Precision					
	ELM	SVM	GNB	RF	DT	LR
No DR	0.93	0.82	0.86	0.82	0.82	0.82
Mild	0.87	0.42	0.24	0.42	0.28	0.41
Moderate	0.87	0.48	0.40	0.47	0.44	0.47
Severe	0.95	0.56	0.47	0.55	0.46	0.52
PDR	0.94	0.61	0.56	0.64	0.61	0.65
Average	0.91	0.58	0.50	0.58	0.52	0.57

Table 4
Classification performance comparison by F-1Precision for Dataset-1.

DR Level	F1-Score					
	ELM	SVM	GNB	RF	DT	LR
No DR	0.95	0.86	0.82	0.86	0.85	0.86
Mild	0.77	0.31	0.28	0.30	0.25	0.31
Moderate	0.82	0.43	0.45	0.43	0.41	0.43
Severe	0.89	0.39	0.40	0.39	0.36	0.39
PDR	0.91	0.46	0.46	0.47	0.44	0.49
Average	0.87	0.49	0.48	0.49	0.46	0.49

Table 5
Classification performance comparison by Recall for Dataset-1.

DR Level	Recall					
	ELM	SVM	GNB	RF	DT	LR
No DR	0.97	0.91	0.78	0.90	0.87	0.90
Mild	0.70	0.24	0.33	0.24	0.23	0.25
Moderate	0.78	0.39	0.52	0.40	0.39	0.39
Severe	0.84	0.30	0.35	0.30	0.29	0.31
PDR	0.89	0.37	0.39	0.37	0.34	0.39
Average	0.83	0.44	0.48	0.44	0.43	0.45

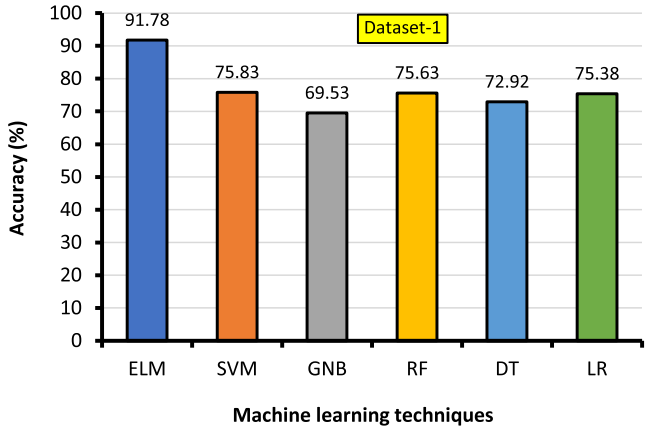

Fig. 7. Accuracies of employed ML techniques for Dataset-1.

4.2. Results of APTOS, 2019 Dataset

In the previous section, the proposed framework revealed promising results for the Dataset-1, which contained a total of 34,984 FIs. Since the DL models worked well for larger datasets, the proposed framework validated this by showing favorable classification performance. In this study, it was also checked whether the proposed framework could achieve promising classification performance with a small dataset. Hence, a small dataset, APTOS, 2019 was used that contained FIs, almost ten times less than the Dataset-1.

Among the total 3,662 FIs, 2,929 FIs were used for training the ELM and other five ML models, whereas the numbers of No DR, Mild, Moderate, Severe, and PDR were 1444, 296, 799, 154, and 236, respectively. For evaluating the ELM classification performance, a CM was developed using 733 FIs (No DR: 361, Mild: 74, Moderate: 200, Severe: 39, and PDR: 59). The level-wise precision, f1-score and recall shown in Tables 6–8 demonstrated that the ELM model performed well in the case of the imbalance or smaller dataset. The best accuracy (97.27 %) was achieved by ELM model for the APTOS, 2019 dataset with a recall of 95

Table 6
Classification performance comparison by Precision for APTOS, 2019 dataset.

DR Level	Precision					
	ELM	SVM	GNB	RF	DT	LR
No DR	1.0	0.96	0.96	0.97	0.96	0.97
Mild	0.99	0.74	0.74	0.79	0.74	0.75
Moderate	0.94	0.8	0.8	0.79	0.79	0.8
Severe	0.9	0.75	0.75	0.71	0.69	0.71
PDR	0.96	0.73	0.73	0.68	0.57	0.72
Average	0.96	0.8	0.8	0.79	0.75	0.79

Table 7
Classification performance comparison by F-1Precision for APTOS, 2019 dataset.

DR Level	F1-Score					
	ELM	SVM	GNB	RF	DT	LR
No DR	0.99	0.97	0.97	0.98	0.97	0.98
Mild	0.97	0.72	0.72	0.74	0.7	0.73
Moderate	0.96	0.84	0.84	0.83	0.81	0.83
Severe	0.92	0.68	0.68	0.63	0.59	0.63
PDR	0.92	0.62	0.62	0.6	0.58	0.63
Average	0.95	0.77	0.77	0.76	0.73	0.76

Table 8
Classification performance comparison by Recall for APTOS, 2019 dataset.

DR Level	Recall					
	ELM	SVM	GNB	RF	DT	LR
No DR	0.99	0.98	0.98	0.99	0.97	0.98
Mild	0.96	0.7	0.7	0.7	0.66	0.7
Moderate	0.97	0.88	0.88	0.87	0.83	0.88
Severe	0.95	0.62	0.62	0.56	0.51	0.56
PDR	0.88	0.54	0.54	0.54	0.59	0.56
Average	0.95	0.74	0.74	0.73	0.72	0.74

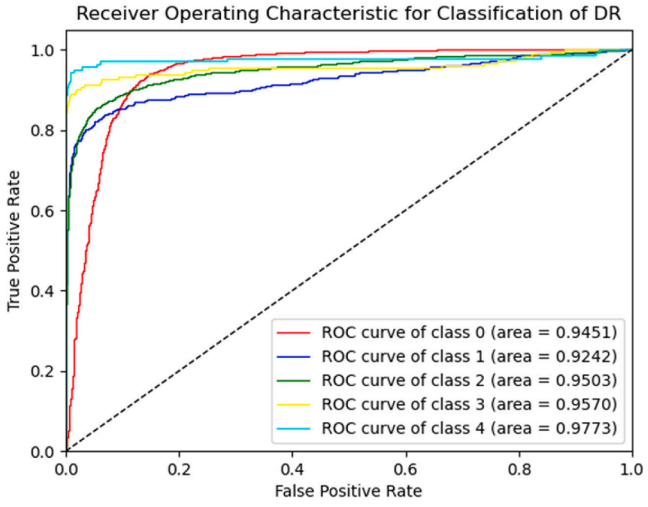

Fig. 8. Receiver Operating Characteristic (ROC) curve of ELM for Dataset-1.

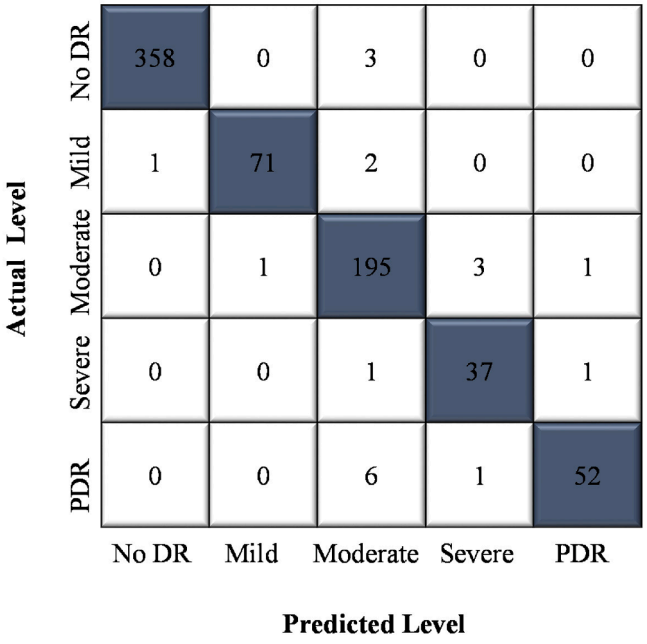

Fig. 9. Confusion matrix of ELM for APTOS, 2019 dataset.

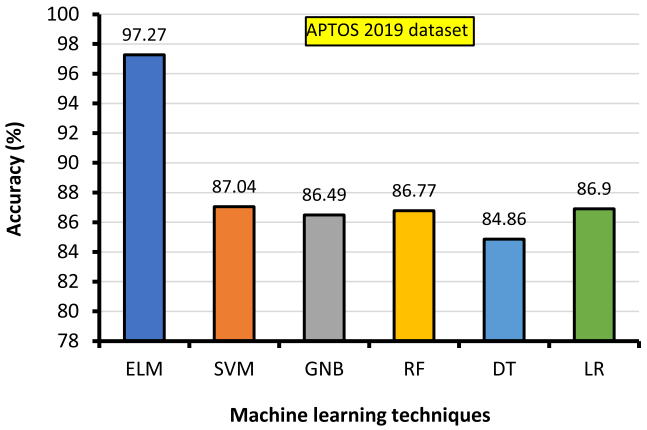

Fig. 10. Accuracies of employed ML techniques for Dataset-1.

% and a precision of 96 % (Fig. 10). Whereas the best accuracy obtained by SVM (87.04 %) among other models was almost 10 % lower than the ELM model. In the case of medical image analysis, the recall must be maximized i.e., the affected patient should be identified accurately.

The class-wise ROC is shown in Fig. 9 to assess the ELM's ability to distinguish between the DR levels. The estimated ROC of the ELM model for the APTOS, 2019 dataset was 98.87 %. The ROC of each class was quite good even if the dataset was unbalanced, demonstrating the model's robustness (See Fig. 11).

A graphical illustration is shown in Fig. 10 to make the results more legible and comparable between the two datasets. The suggested framework is compatible in any setting, such as smaller (APTOS, 2019) or larger (Dataset 1) datasets. This was accomplished by employing CLAHE to highlight the lesions. Hence, it is easy for the parallel CNN model to extract the most discriminating features and the ELM based on deep learning mechanism can accurately detected the DR levels. The framework was also straightforward to use as it performed well even when dealing with an unbalanced dataset, which is common with real-world medical data (See Fig. 12).

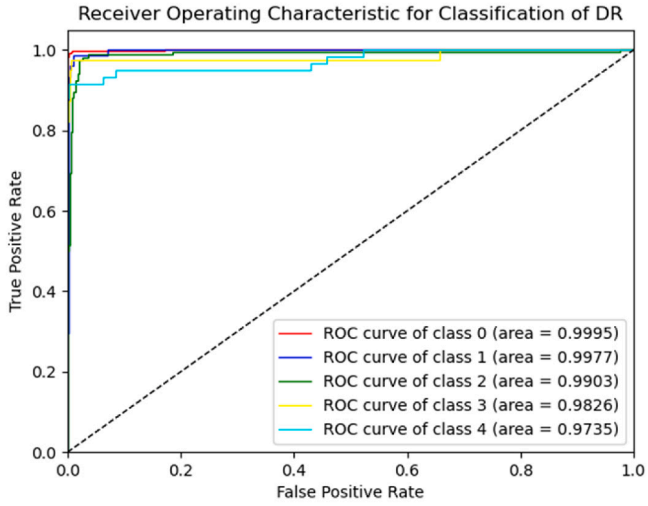

Fig. 11. ROC matrix of ELM for APTOS, 2019 dataset.

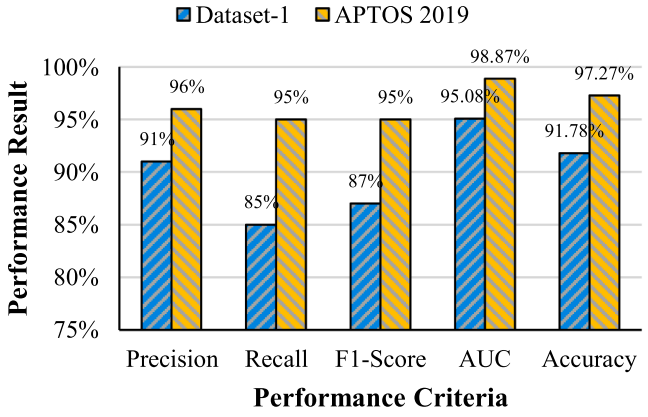

Fig. 12. Graphical illustration of the classification performance of proposed framework.

4.3. Comparison with previous works

Tables 9 and 10 show the classification performance compared with previous state-of-the-art (SOTA) models for both the datasets. For Dataset-1, the proposed framework (PF) was compared with two studies. Pratt et al. processed the FIs using color normalization, and developed a CNN with 10 CLs and two FC layers (Pratt et al., 2016). The number of filters in 10 CLs were 32, 32, 64, 64, 128, 128, 256, 256, 512, and 512, respectively, and both the FC layers had 1,024 nodes. Apart from these, they used 5,000 FIs (the shape of the FIs were 512 × 512) for testing and achieved an overall accuracy and sensitivity (recall) of 75 % and 30 %, respectively. In contrast, Qummar et al. (2019), used five TL models: Resnet50, Inceptionv3, Xception, Dense121, and Dense169 for classifying DR from the FIs. In addition, they ensemble these five TL models for final prediction. They also resized the FIs into 512 × 512 and achieved an accuracy, recall, precision, and f1-score of 80.8 %, 51.5 %, 63.85 %, and 53.74 %, respectively, while testing the model on 5,608 FIs and performing up and down sampling. The proposed PCNN-ELM has only 8 CLs with 3.6 million parameters, which was quite fewer than the other two works. Again, the contrast of the FIs were enhanced using CLAHE, and for that reason, the lesion was highlighted as shown in Fig. 4. Finally, the FIs were resized into 124 × 124 and the framework has been tested using 5608 FIs and achieved an accuracy of 91.88 %, which is 10 % higher than the previous study, and a recall of 83 %, which is almost 30 % higher than the previous study. The prior two research were significantly affected by the imbalanced dataset. No DR

Table 9
Class-wise classification performance of the proposed framework (PF) compared with the previous studies for the Dataset-1.

Level/ Ref. No.	Precision			Recall			F1-Score			AUC		
	(Pratt et al., 2016)	(Qummar et al., 2019)	PF	(Pratt et al., 2016)	(Qummar et al., 2019)	PF	(Pratt et al., 2016)	(Qummar et al., 2019)	PF	(Pratt et al., 2016)	(Qummar et al., 2019)	PF
No DR	0.78	0.84	0.93	0.95	0.97	0.97	0.85	0.90	0.95	–	0.85	0.94
Mild	0.00	0.51	0.89	0.00	0.80	0.68	0.00	0.15	0.78	–	0.71	0.92
Moderate	0.23	0.65	0.87	0.23	0.41	0.78	0.29	0.50	0.82	–	0.85	0.95
Severe	0.78	0.48	0.92	0.78	0.51	0.83	0.10	0.49	0.88	–	0.96	0.96
PDR	0.44	0.69	0.93	0.44	0.56	0.88	0.37	0.62	0.90	–	0.97	0.97

Table 10
Classification performance compared with SOTA models for the APTOS, 2019.

Ref. No.	Precision (%)	Recall (%)	Accuracy (%)	AUC (%)
(Dondeti et al., 2020)	76.00	77.00	77.90	–
(Bodapati et al., 2020)	80.00	81.00	81.70	–
(Liu et al., 2020)	91.37	–	86.34	–
(Kassani et al., 2019)	87.00	88.24	83.09	91.80
(Bodapati et al., 2021)	82.00	83.00	82.54	79.00
(Sikder et al., 2021)	94.34	92.69	94.20	–
(Alyoubi et al., 2021)	89.00	–	89.00	97.90
Proposed Framework	96.00	95.00	97.27	98.87

level highly dominated the final classification result and showed a preliminary result in the case of other classes as seen in Table 9. On the contrary, each class almost equally contributed to the final classification result, which validated the handling capability of the unbalanced dataset of the proposed framework. Pratt et al. (2016), showed that their proposed methodology required 0.04 s to classify-one FI. In contrast, the proposed framework required only 0.0009987 s to test the total of 5608 FIs, whereas 2 μs were required for classifying one FI. These two studies reshaped the FIs by 512 × 512, whereas this study used 124 × 124 but still ensuring a promising result with a calculated AUC of 92.10 % (Qummar et al. achieved an average AUC of 86.8 %) that showed the robustness of the proposed model.

Several researchers used the APTOS, 2019 dataset to detect the levels of DR. Table 10 shows the average classification performance of the SOTA models for the APTOS, 2019 dataset as class wise results were not available. Sikder et al. achieved the highest classification accuracy of 94.20 %, and the highest AUC of 97.90 % was achieved by Alyoubi et al. (2021), from the SOTA models (Sikder et al., 2021). In contrast, the proposed framework outperformed all the SOTA models with an accuracy and an AUC of 97.27 % and 98.87 %, respectively. Table 11 shows the comparison of the proposed framework's performance with the previous works. From the table, most of the SOTA models employed transfer learning (TL) models to extract the features and classify the DR from the FIs. The TL models have many layers and parameters; for instance, the VGG16 model has almost 138.3 million parameters, and DenseNet-169 has 169 layers, which are too many. They also required high-resolution FIs (512 × 512, 380 × 380, 224 × 224, etc.) to distinguish the DR levels correctly.

Table 11
Simplicity of the proposed framework compared with SOTA models.

Model Name [Ref. No.]	No. of Layers	No. of Parameters (million)
ResNet 50 (Qummar et al., 2019; Kassani et al., 2019)	50	25.6
Inception-V3 (Qummar et al., 2019; Kassani et al., 2019)	48	23.8
Xception (Qummar et al., 2019; Bodapati et al., 2020; Liu et al., 2020; Kassani et al., 2019; Bodapati et al., 2021)	71	22.9
Dense 121 (Qummar et al., 2019)	121	8.0
Dense 169 (Qummar et al., 2019)	169	14.3
VGG16 (Bodapati et al., 2020; Bodapati et al., 2021)	16	138.3
NasNet-Large (Bodapati et al., 2020; Liu et al., 2020)	–	88.9
Inception Resnet V2 (Bodapati et al., 2020; Liu et al., 2020)	164	55.8
EfficientNetB4 (Liu et al., 2020)	–	19.4
EfficientNetB5 (Liu et al., 2020)	–	30.5
CNN512 (Alyoubi et al., 2021)	9	8.2
Proposed Framework	8	3.6

In contrast, the proposed framework only employed six CLs, where four of them were run in parallel, which was considered a single CL. Hence, there were total eight layers, including four CLs, two FC layers, and three from the ELM. The total parameters of the proposed framework are almost 3.6 million, including both the parallel CNN and ELM model parameters that validated the lightweight capability of the CNN. This framework required an image size of 124 × 124, which was another objective of this study to detect DR levels using low-resolution FIs. Table 11 shows that the proposed framework has the lowest number of parameters and layers, which could be the main reason for shorter processing time.

From the above comparison, it was concluded that the proposed framework could classify the levels of DR accurately with lower parameters, layers, low-resolution FIs, and relatively shorter time. It was also revealed that the framework is capable of adapting to any dataset environment, small or large, balanced or imbalanced and that classifying a FI requires only 2 μs seconds, allowing for real-time patient feedback.

In fact, the diabetic retinopathy datasets used in this study were highly imbalanced particularly for the multiclass classification due to the unavailability of the PDR images. Since in real life consideration, the number of PDR patients are not many, and most datasets contain small portion of PDR images with respect to the other classes. Since the dataset was fairly imbalanced, some researchers used data augmentation and other techniques (adding weight to the poorly detected class, up sampling, down sampling) to improve the classification performance (Pratt et al., 2016; Dondeti. al., 2020). However, the results were not better than the findings obtained by the proposed framework. Using data augmentation, more data can be produced, but it needs additional time for processing. Most studies, while using these two imbalanced datasets, reported results without any data balancing (Nahiduzzaman et al., 2021a,b; Pratt et al., 2016; Dondeti. al., 2020; Bodapati et al., 2020; Bodapati et al., 2021). Apart from these, the results presented for Dataset-1 in Table 4 to Table 6, it was observed that the precision, recall and f1-score of PDR images were 0.94, 0.91 and 0.89 respectively which were quite satisfactory. Though there were a smaller number of PDR images, but the classification results were similar to the normal images (precision, f1-score and recall of No DR images are 0.93, 0.97 and 0.95 respectively). Again, similar observations were made for the APTOS, 2019 dataset. In addition, from Fig. 8, it was found that the class wise ROC of PDR was 97.73 % whereases for No DR it was 94.51 % in the case of Dataset-1. Therefore, it can be concluded that without implementing any data augmentation technique, the proposed CNN-ELM model detected the DR accurately without producing any biased results due to

the imbalanced datasets. The proposed framework achieved a promising outcome based on the performance metrics considered in this study and eliminated additional time required for data augmentation.

5. Conclusion

This study proposed a novel framework to enable fast and accurate detection of the levels of DR from the FIs, which can aid diabetic patients in preventing or delaying vision loss. CLAHE was adopted to make the lesson clear so that a CNN model can easily extract the most discriminating features. 120 features were extracted using a lightweight parallel CNN to reduce processing time and complexity. Finally, these features were standardized and fit into the ELM model to adequately distinguish the different levels of the DR. The proposed framework exhibited a promising result in the cases of 34,984 (Dataset-1) and 3,662 (APTOS, 2019) FI datasets with not only higher classification performance but also lowering the parameters, layers, and processing time significantly. The framework also outperformed the existing SOTA models for both the datasets. The proposed model can accurately detect the severity degree of the DR earlier on, hence reducing vision loss of the patients and saving valuable time of the medical practitioners.

CRediT authorship contribution statement

Md. Nahiduzzaman: Data curation, Conceptualization, Investigation, Methodology, Validation, Formal analysis, Writing – original draft. **Md. Robiul Islam:** Conceptualization, Investigation, Methodology, Validation, Formal analysis, Data curation, Writing – original draft. **Md. Omaer Faruq Goni:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Md. Shamim Anower:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – review & editing, Supervision. **Mominul Ahsan:** Methodology, Visualization, Conceptualization, Formal analysis, Writing – review & editing, Supervision. **Julfikar Haider:** Visualization, Formal analysis, Conceptualization, Methodology, Validation, Writing – review & editing, Supervision. **Marcin Kowalski:** Conceptualization, Methodology, Formal analysis, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Afza, F., Sharif, M., Khan, M. A., Tariq, U., Yong, H. S., & Cha, J. (2021). Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine. *Sensors, 22*(3), 799.

Akram, M. U., Khalid, S., Tariq, A., Khan, S. A., & Azam, F. (2014). Detection and classification of retinal lesions for grading of diabetic retinopathy. *Computers in Biology and Medicine, 45*, 161–171.

Ali, A., Qadri, S., Mashwani, W. K., Kumam, W., Kumam, P., Naeem, S., … Anam, S. (2020). Machine learning based automated segmentation and hybrid feature analysis for diabetic retinopathy classification using fundus image. *Entropy, 22*(5), 567.

Alenezi, F., Armghan, A., & Polat, K. (2023). Wavelet transform based deep residual neural network and ReLU based Extreme Learning Machine for skin lesion classification. *Expert Systems with Applications, 213*, Article 119064.

Alyoubi, W. L., Abulkhair, M. F., & Shalash, W. M. (2021). Diabetic retinopathy fundus image classification and lesions localization system using deep learning. *Sensors, 21*(11), 3704.

Asha, P. and Karpagavalli, S. "Diabetic retinal exudates detection using extreme learning machine," in Emerging ICT for Bridging the Future- Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2. Springer, 2015, pp. 573–578.

Bodapati, J. D., Naralasetti, V., Shareef, S. N., Hakak, S., Bilal, M., Maddikunta, P. K. R., & Jo, O. (2020). Blended multi-modal deep convnet features for diabetic retinopathy severity prediction. *Electronics, 9*(6), 914.

Bodapati, J. D., Shaik, N. S., & Naralasetti, V. (2021). Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification. *Journal of Ambient Intelligence and Humanized Computing*, 1–15.

Chudzik, P., Majumdar, S., Calivá, F., Al-Diri, B., & Hunter, A. (2018). Microaneurysm detection using fully convolutional neural networks. *Computer methods and programs in biomedicine, 158*, 185–192.

Chetoui, M., Akhloufi, M. A., & Kardouchi, M. (2018). Diabetic retinopathy detection using machine learning and texture features. In *In 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)* (pp. 1–4). IEEE.

Chen, W., Yang, B., Li, J., and Wang, J., "An approach to detecting diabetic retinopathy based on integrated shallow convolutional neural networks," IEEE Access, vol. 8, pp. 178 552–178 562, 2020.

Carrera, E. V., González, A. and Carrera, R. "Automated detection of diabetic retinopathy using SVM," in 2017 IEEE XXIV international conference on electronics, electrical engineering and computing (INTERCON). IEEE, 2017, pp. 1–4.

Dondeti, V., Bodapati, J. D., Shareef, S. N., & Veeranjaneyulu, N. (2020). Deep convolution features in non-linear embedding space for fundus image classification. *Rev. d'Intelligence Artif., 34*(3), 307–313.

[dataset 1] California Healthcare Foundation, "Diabetic retinopathy detection," https://www.kaggle.com/c/diabetic-retinopathy-detection/data, 2015, [accessed on 1-February-2022].

[dataset 2] Asia Pacific Tele-Ophthalmology Society (APTOS), "Aptos 2019 blindness detection," https://www.kaggle.com/c/aptos2019-blindness-detection/data, 2019, [Accessed: 1-February- 2022].

Farrell, J., & Saloner, G. (1985). Standardization, compatibility, and innovation. *the RAND Journal of Economics*, 70–83.

Gangwar, A. K., & Ravi, V. (2021). Diabetic retinopathy detection using transfer learning and deep learning. In *Evolution in Computational Intelligence* (pp. 679–689). Springer.

Gayathri, S., Gopi, V. P., & Palanisamy, P. (2021). Diabetic retinopathy classification based on multipath cnn and machine learning classifiers. *Physical and Engineering Sciences in Medicine*, 1–15.

Gondal, W. M., Köhler, J. M., Grzeszick, R., Fink, G. A. and Hirsch, M. "Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images," in 2017 IEEE international conference on image processeing (ICIP). IEEE, 2017, pp. 2069–2073.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1–3), 489–501.

Honnungar, S., Mehra, S. and Joseph, S. "Diabetic retinopathy identification and severity classification," Fall 2016, 2016.

Islam, M. R., M. A. M. Hasan, and Sayeed, A. "Transfer learning based diabetic retinopathy detection with a novel preprocessed layer," in 2020 IEEE Region 10 Symposium (TENSYMP). IEEE, 2020, pp. 888–891.

Islam, M. R., Hasan, M. N., and Nahiduzzaman, M., "Severity grading of diabetic retinopathy using deep convolutional neural network." International Journal of Innovative Science and Research Technology, vol. 6 no. 1, pp. 1395–1401.

Islam, S. M. S., Hasan, M. M. and Abdullah, S. "Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images," arXiv preprint arXiv: 1812.10595, 2018.

Kassani, S. H., Kassani, P. H., Khazaeinezhad, R., Wesolowski, M. J., Schneider, K. A. and Deters, R. "Diabetic retinopathy classification using a modified xception architecture," in 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2019, pp. 1–6.

Kar, S. S., & Maity, S. P. (2017). Automatic detection of retinal lesions for screening of diabetic retinopathy. *IEEE Transactions on Biomedical Engineering, 65*(3), 608–618.

Liu, H., Yue, K., Cheng, S., Pan, C., Sun, J., & Li, W. (2020). Hybrid model structure for diabetic retinopathy classification. *Journal of Healthcare Engineering, 2020*.

Lachure, J., Deorankar, A., Lachure, S., Gupta, S. and Jadhav, R. "Diabetic retinopathy using morphological operations and machine learning," in 2015 IEEE international advance computing conference (IACC). IEEE, 2015, pp. 617–622.

Majumder, S. and Kehtarnavaz, N., "Multitasking deep learning model for detection of five stages of diabetic retinopathy," arXiv preprint arXiv:2103.04207, 2021.

Mahmoud, M. H., Alamery, S., Fouad, H., Altinawi, A., & Youssef, A. E. (2021). An automatic detection system of diabetic retinopathy using a hybrid inductive machine learning algorithm. *Personal and Ubiquitous Computing*, 1–15.

Mumtaz, R., Hussain, M., Sarwar, S., Khan, K., Mumtaz, S., & Mumtaz, M. (2018). Automatic detection of retinal hemorrhages by exploiting image processing techniques for screening retinal diseases in diabetic patients. *International Journal of Diabetes in Developing Countries, 38*(1), 80–87.

Nahiduzzaman, M., Islam, M. R., Islam, S. R., Goni, M. O. F. Anower, M. S., and Kwak, K. S. "Hybrid cnn-svd based prominent feature extraction and selection for grading diabetic retinopathy using extreme learning machine algorithm," IEEE Access, vol. 9, pp. 152 261–152 274, 2021.

Nahiduzzaman, M., Nayeem, M. J., Ahmed, M. T. and Zaman, M. S. U. "Prediction of heart disease using multi-layer perceptron neural network and support vector machine," in 2019 4th International conference on electrical information and communication technology (EICT). IEEE, 2019, pp. 1–6.

Nahiduzzaman, M., Goni, M. O. F., Anower, M. S., Islam, M. R., Ahsan, M., Haider, J., Gurusamy, S., Hassan, R. and Islam, M. R., "A novel method for multivariant pneumonia classification based on hybrid CNN-PCA based feature extraction using extreme learning machine with CXR images," IEEE Access, vol. 9, pp. 147 512–147 526, 2021.

Odeh, I., Alkasassbeh, M. and Alauthman, M., "Diabetic retinopathy detection using ensemble machine learning," arXiv preprint arXiv:2106.12545, 2021.

Pandey, S. K., & Sharma, V. (2018). World diabetes day 2018: Battling the emerging epidemic of diabetic retinopathy. *Indian Journal of Ophthalmology, 66*(11), 1652.

Powers, D. M. "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," arXiv preprint arXiv:2010.16061, 2020.

Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., & Zheng, Y. (2016). Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science, 90*, 200–205.

Qummar, S., Khan, F. G., Shah, S., Khan, A., Shamshirband, S., Rehman, Z. U., Khan, I. A. and Jadoon, W. "A deep learning ensemble approach for diabetic retinopathy detection," IEEE Access, vol. 7, pp. 150 530– 150 539, 2019.

Rahim, S. S., Palade, V., Shuttleworth, J., & Jayne, C. (2016). Automatic screening and classification of diabetic retinopathy and maculopathy using fuzzy image processing. *Brain informatics, 3*(4), 249–267.

Raman, V., Then, P., & Sumari, P. (2016). Proposed retinal abnormality detection and classification approach: Computer aided detection for diabetic retinopathy by machine learning approaches. In *In 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN)* (pp. 636–641). IEEE.

Ramani, R. G., & Lakshmi, B. (2017). Automatic diabetic retinopathy detection through ensemble classification techniques automated diabetic retinopathy classification. In *In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (pp. 1–4). IEEE.

Reddy, G. T., Bhattacharya, S., Ramakrishnan, S. S., Chowdhary, C. L., Hakak, S., Kaluri, R. and Reddy, M. P. K. "An ensemble-based machine learning model for diabetic retinopathy classification," in 2020 international conference on emerging trends in information technology and engineering (ic-ETITE). IEEE, 2020, pp. 1–6.

Samanta, A., Saha, A., Satapathy, S. C., Fernandes, S. L., & Zhang, Y.-D. (2020). Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset. *Pattern Recognition Letters, 135*, 293–298.

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.

Shenavarmasouleh, F. and Arabnia, H. R. "Drdr: Automatic masking of exudates and microaneurysms caused by diabetic retinopathy using mask r-cnn and transfer learning," arXiv preprint arXiv:2007.02026, 2020.

Somasundaram, S., & Ali, P. (2017). A machine learning ensemble classifier for early prediction of diabetic retinopathy. *Journal of Medical Systems, 41*(12), 1–12.

Umapathy, A., Sreenivasan, A., Nairy, D. S., Natarajan, S. and Rao, B. N., "Image processing, textural feature extraction and transfer learning based detection of diabetic retinopathy," in Proceedings of the 2019 9th International Conference on Bioscience, Biochemistry and Bioinformatics, 2019, pp. 17–21.

Yu, S., Xiao, D., and Kanagasingam, Y. "Exudate detection for diabetic retinopathy with convolutional neural networks," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 1744–1747.

Zhou, K., Gu, Z., Liu, W., Luo, W., Cheng, J., Gao, S. and Liu, J. "Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018, pp. 2724–2727.

Zeng, X., Chen, H., Luo, Y. and Ye, W. "Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network," IEEE Access, vol. 7, pp. 30 744–30 753, 2019.

Zhao, Z., Zhang, K., Hao, X., Tian, J., Chua, M. C. H., Chen, L. and Xu, X. "Bira-net: Bilinear attention net for diabetic retinopathy grading," in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 1385–1389.