

Diabetic retinopathy detection through deep learning techniques: A review

Wejdan L. Alyoubi ^{*}, Wafaa M. Shalash, Maysoon F. Abulkhair

Information Technology Department, University of King Abdul Aziz, Jeddah, Saudi Arabia



ARTICLE INFO

Index Terms:

Computer-aided diagnosis
Deep learning
Diabetic retinopathy
Diabetic retinopathy stages
Retinal fundus images

ABSTRACT

Diabetic Retinopathy (DR) is a common complication of diabetes mellitus, which causes lesions on the retina that affect vision. If it is not detected early, it can lead to blindness. Unfortunately, DR is not a reversible process, and treatment only sustains vision. DR early detection and treatment can significantly reduce the risk of vision loss. The manual diagnosis process of DR retina fundus images by ophthalmologists is time-, effort-, and cost-consuming and prone to misdiagnosis unlike computer-aided diagnosis systems. Recently, deep learning has become one of the most common techniques that has achieved better performance in many areas, especially in medical image analysis and classification. Convolutional neural networks are more widely used as a deep learning method in medical image analysis and they are highly effective. For this article, the recent state-of-the-art methods of DR color fundus images detection and classification using deep learning techniques have been reviewed and analyzed. Furthermore, the DR available datasets for the color fundus retina have been reviewed. Difference challenging issues that require more investigation are also discussed.

1. Introduction

In the healthcare field, the treatment of diseases is more effective when detected at an early stage. Diabetes is a disease that increases the amount of glucose in the blood caused by a lack of insulin [1]. It affects 425 million adults worldwide [2]. Diabetes affects the retina, heart, nerves, and kidneys [1,2].

Diabetic Retinopathy (DR) is a complication of diabetes that causes the blood vessels of the retina to swell and to leak fluids and blood [3]. DR can lead to a loss of vision if it is in an advanced stage. Worldwide, DR causes 2.6% of blindness [4]. The possibility of DR presence increases for diabetes patients who suffer from the disease for a long period. Retina regular screening is essential for diabetes patients to diagnose and to treat DR at an early stage to avoid the risk of blindness [5]. DR is detected by the appearance of different types of lesions on a retina image. These lesions are microaneurysms (MA), haemorrhages (HM), soft and hard exudates (EX) [1,6,7].

- *Microaneurysms* (MA) is the earliest sign of DR that appears as small red round dots on the retina due to the weakness of the vessel's walls. The size is less than 125 µm and there are sharp margins. Michael et al. [8] classified MA into six types, as shown in Fig. 1. The types of MA were seen with AOSLO reflectance and conventional fluorescein imaging.

- *Haemorrhages* (HM) appear as larger spots on the retina, where its size is greater than 125 µm with an irregular margin. There are two types of HM, which are flame (superficial HM) and blot (deeper HM), as shown in Fig. 2.

- *Hard exudates* appear as bright-yellow spots on the retina caused by leakage of plasma. They have sharp margins and can be found in the retina's outer layers.

- *Soft exudates* (also called cotton wool) appear as white spots on the retina caused by the swelling of the nerve fiber. The shape is oval or round.

Red lesions are MA and HM, while bright lesions are soft and hard exudates (EX). There are five stages of DR depending on the presence of these lesions, namely, no DR, mild DR, moderate DR, severe DR and proliferative DR, which are briefly described in Table 1. A sample of DR stages images is provided in Fig. 3.

The automated methods for DR detection are cost and time saving and are more efficient than a manual diagnosis [10]. A manual diagnosis is prone to misdiagnosis and requires more effort than automatic methods. This paper reviews the recent DR automated methods that use deep learning to detect and to classify DR. The current work covered 33 papers which used deep learning techniques to classify DR images. This paper is organized as follows: Section 2 briefly explains deep learning techniques, while Section 3 presents the various fundus retina datasets. Section 4 presents the performance measures while Section 5 reviews

* Corresponding author.

E-mail address: walyoubi0016@stu.kau.edu.sa (W.L. Alyoubi).

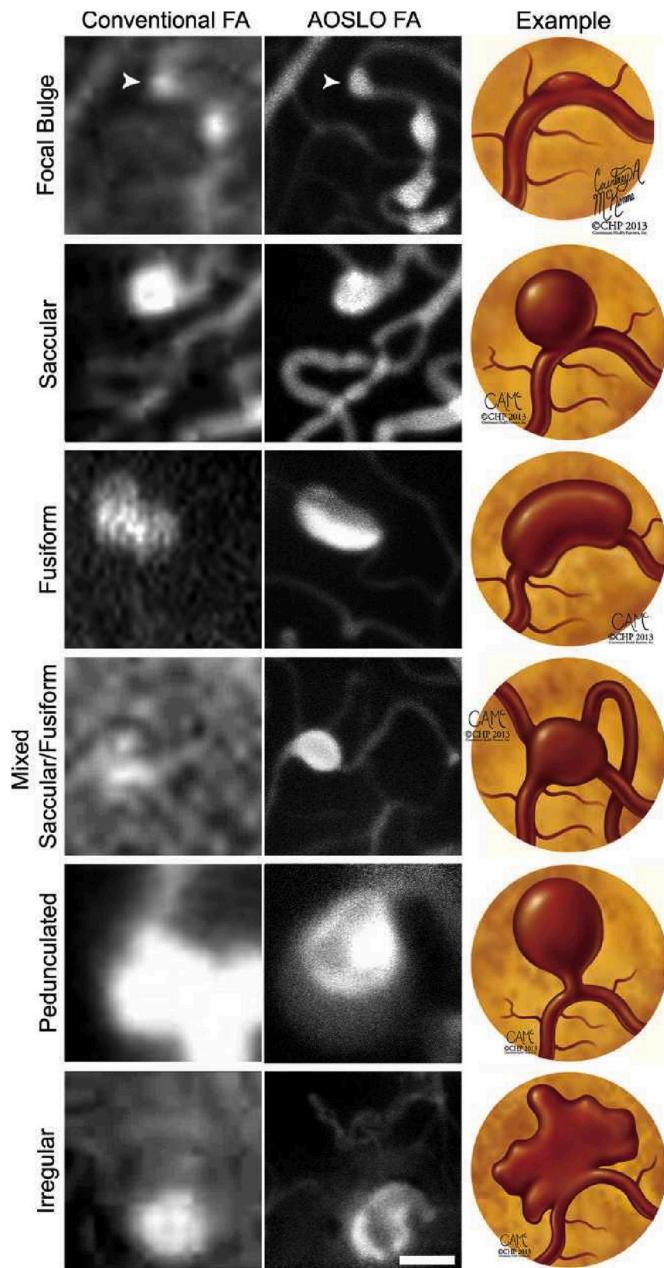


Fig. 1. The different types of MA [8].

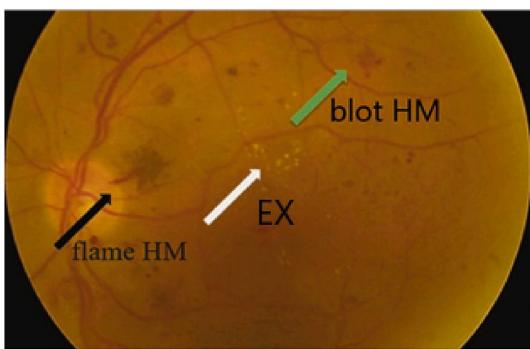


Fig. 2. The different types of HM [9].

different image preprocessing methods used with fundus images.

Table 1
Levels of DR with its associative lesions [13].

DR Severity Level	Lesions
No DR	Absent of lesions
Mild non-proliferative DR	MA only
Moderate non-proliferative DR	More than just MA but less than severe DR
Severe non-proliferative DR	Any of the following: <ul style="list-style-type: none"> more than 20 intraretinal HM in each of 4 quadrants definite venous beading in 2+quadrants Prominent intraretinal microvascular abnormalities in 1+ quadrant no signs of proliferative DR
Proliferative DR	One or more of the following: vitreous/pre-retinal HM, neovascularization

Section 6 describes different DR automated classification methods while a discussion section is presented in Section 7. A summary is provided in Section 8.

2. Deep learning

Deep learning (DL) is a branch of machine learning techniques that involves hierarchical layers of non-linear processing stages for unsupervised features learning as well as for classifying patterns [11]. DL is one computer-aided medical diagnosis method [12]. DL applications to medical image analysis include the classification, segmentation, detection, retrieval, and registration of the images.

Recently, DL has been widely used in DR detection and classification. It can successfully learn the features of input data even when many heterogeneous sources integrated [14]. There are many DL-based methods such as restricted Boltzmann Machines, convolutional neural networks (CNNs), auto encoder, and sparse coding [15]. The performance of these methods increases when the number of training data increase [16] due to the increase in the learned features unlike machine learning methods. Also, DL methods did not require hand-crafted feature extraction. Table 2 summarizes these differences between DL and machine learning methods.

CNNs are more widely used more than the other methods in medical image analysis [17], and it is highly effective [15].

There are three main layers in the CNN architecture, which are convolution layers (CONV), pooling layers, and fully connected layers (FC). The number of layers, size, and the number of filters of the CNN vary according to the author's vision. Each layer in CNN architecture plays a specific role. In the CONV layers, different filters convolve an image to extract the features. Typically, pooling layer follows the CONV layer to reduce the dimensions of feature maps. There are many strategies for pooling but average pooling and max pooling are adopted most [15]. A FC layers are a compact feature to describe the whole input image. SoftMax activation function is the most used classification function. There are different available pretrained CNN architectures on ImageNet dataset such as AlexNet [19], Inception-v3 [20] and ResNet [21]. Some studies like [22,23] transfer learning these pretrained architectures to speed up training while other studies build their own CNN from scratch for classification. The transfer learning strategies of pretrained models include finetuning last FC layer or finetuning multiple layers or training all layers of pretrained model.

Generally, the process used to detect and to classify DR images using DL begins by collecting the dataset and by applying the needed pre-process to improve and enhance the images. Then, this is fed to the DL method to extract the features and to classify the images, as shown in Fig. 4. These steps are explained in the following sections.

3. Retina dataset

There are many publicly available datasets for the retina to detect DR

and to detect the vessels. These datasets are often used to train, validate and test the systems and also to compare a system's performance against other systems. Fundus color images and optical coherence tomography (OCT) are types of retinal imaging. OCT images are 2 and 3- dimensional images of the retina taken using low-coherence light and they provide considerable information about retina structure and thickness, while fundus images are 2-dimensional images of the retina taken using reflected light [24]. OCT retinal images have been introduced in past few years. There is a diversity of publicly available fundus image datasets that are commonly used. Fundus image datasets are as follows:

- **DIARETDB1** [25]: It contains 89 publicly available retina fundus images with the size of 1500×1152 pixels acquired at the 50-degree field of view (FOV). It includes 84 DR images and five normal images annotated by four medical experts.
- **Kaggle** [26]: It contains 88,702 high-resolution images with various resolutions, ranging from 433×289 pixels to 5184×3456 pixels, collected from different cameras. All images are classified into five DR stages. Only training images ground truths are publicly available. Kaggle contains many images with poor quality and incorrect labeling [23,27].
- **E-ophtha** [28]: This publicly available dataset includes E-ophtha EX and E-ophtha MA. E-ophtha EX includes 47 images with EX and 35 normal images. E-ophtha MA contains 148 images with MA and 233 normal images.
- **DDR** [23]: This publicly available dataset contains 13,673 fundus images acquired at a 45-degree FOV annotated to five DR stages. There are 757 images from the dataset annotated to DR lesions.
- **DRIVE** [29]: This publicly available dataset is used for blood vessel segmentation. It contains 40 images acquired at a 45-degree FOV. The images have a size of 565×584 pixels. Among them, there are seven mild DR images, and the remaining include images of a normal retina.
- **HRF** [30]: These publicly available images provided for blood vessel segmentation. It contains 45 images with a size of 3504×2336 pixels. There are 15 DR images, 15 healthy images and 15 glaucomatous images.
- **Messidor** [31]: This publicly available dataset contains 1200 fundus color images acquired at a 45-degree FOV annotated to four DR stages.
- **Messidor-2** [31]: This publicly available dataset contains 1748 images acquired at a 45-degree FOV.

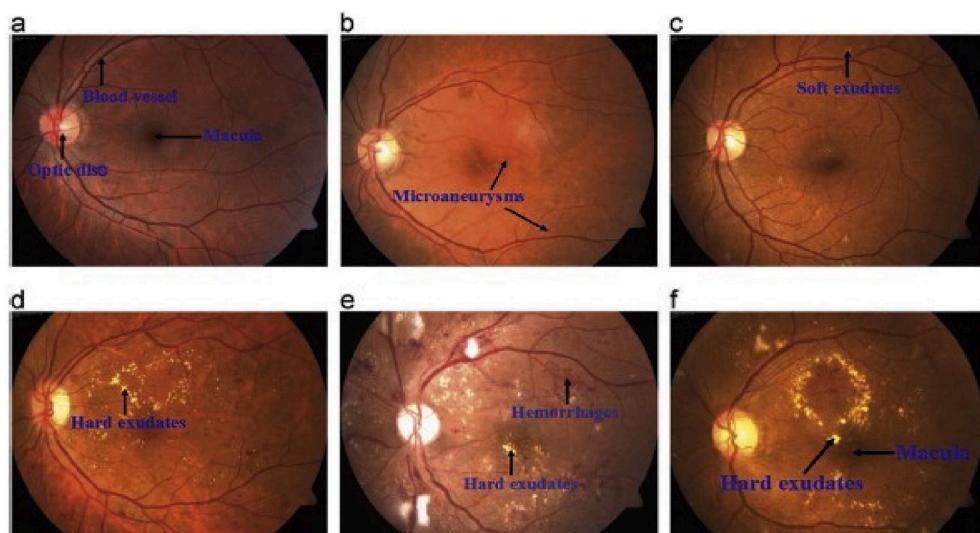


Fig. 3. The DR stages: (a) normal retinal (b) Mild DR, (c) Moderate DR, (d) Severe DR, (e) Proliferative DR,(f) Macular edema [18].

Table 2

The differences between DL and machine learning methods.

	DL	Machine learning
Hand-crafted feature extraction	Not required	Required
Training data	Required large data	Not required large data

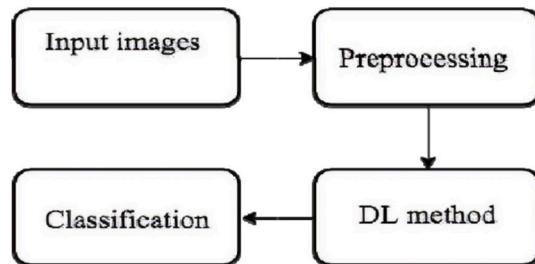


Fig. 4. The process of classifying the DR images using DL.

- **STARE** [32]: This publicly available dataset is used for blood vessel segmentation. It contains 20 images acquired at a 35-degree FOV. The images have a size of 700×605 pixels. Among them, there are 10 normal images.
- **CHASE DB1** [33]: This publicly available dataset provided for blood vessel segmentation. It contains 28 images with a size of 1280×960 pixels and acquired at a 30-degree FOV.
- **Indian Diabetic Retinopathy Image dataset (IDRiD)** [34]: This publicly available dataset contains 516 fundus images acquired at a 50-degree FOV annotated to five DR stages.
- **ROC** [35]: It contains 100 publicly available retina images acquired at the 45-degree FOV. Its size ranging from 768×576 to 1389×1383 pixels. The images annotated to detect MA. Only training ground truths are available.
- **DR2** [36]: It contains 435 publicly available retina images with 857×569 pixels. It provides referral annotations for images. There are 98 images were graded as referral.

The study of [37] used DIARETDB1 datasets to detect DR lesions. The study of [38] used DIARETDB1 and E-ophtha to detect red lesion while the study of [39] used these datasets to detect MA. In Ref. [40] DIARETDB1 was used to detect EX. The Kaggle dataset was used in the studies of [22,37,41–45] to classify DR stages. DRIVE, HRF, STARE and

CHASE DB1 were used in the work of [46] to segment the blood vessels, while in Ref. [47] DRIVE dataset was used. The results of these studies are discussed in section 5. Table 3 compares these datasets. Most of the studies processed the datasets before using them for DL methods. The next sections discuss the performance measures and preprocessed methods.

4. Performance measures

There are many performance measurements that applied to DL methods to measure their classification performance. The commonly used measurements in DL are accuracy, sensitivity, specificity and area under the ROC curve (AUC). Sensitivity is the percentage of abnormal images that classified as abnormal, and specificity is the percentage of normal images that classified as normal [65]. AUC is a graph created by plotting sensitivity against specificity. Accuracy is the percentage of images that are classified correctly. The following is the equations of each measurement.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \quad (3)$$

True positive (TP) is the number of disease images that classified as disease. True negative (TN) is the number of normal images that classified as normal while false positive (FP) is the number of normal images that classified as disease. False negative (FN) is the number of disease images that classified as normal. The percentage of performance measures used in the studies, that involved in the current work, shown in Fig. 5.

5. IMAGE PREPROCESSING

Image preprocessing is a necessary step to remove the noise from images, to enhance image features and to ensure the consistency of images [43]. The following paragraph discusses the most common preprocessing techniques that have been used recently in researches.

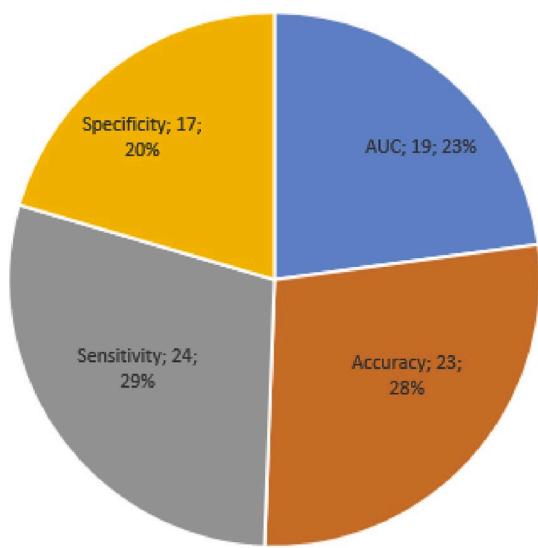


Fig. 5. The percentage of performance measures used in the studies.

Many researchers resized the images to a fixed resolution to be suitable for the used network, as done in Refs. [37,41]. Cropped images were applied to remove the extra regions of the image, while data normalization was used to normalize the images into a similar distribution, as in Ref. [45]. In some works, such as [38], only the green channel of images was extracted due to its high contrast [46], the images were converted into grayscale, such as in Ref. [43].

Noise removal methods include a median filter, Gaussian filter, and NonLocal Means Denoising methods, such as in the works of [38,43,45], respectively. Data augmentation techniques were performed when some image classes were imbalance or to increase the dataset size, such as in Ref. [38,45]. Data augmentation technique include translation, rotation, shearing, flipping, contrast scaling and rescaling. A morphological method was used, such as in Ref. [39], for contrast enhancement. The canny edge method was used for feature extraction in the study of [40].

Table 3
Details of DR datasets.

Dataset	Numbers of images	Normal image	Mild DR	Moderate and severe non-proliferative DR	Proliferative DR	Training Sets	Test Sets	Image Size
DiaretDB1	89 images	27 images	7 images	28 images	27 images	28 images	61 images	1500 × 1152 pixels
Kaggle	88,702 images	–	–	–	–	35,126 images	53,576 images	Different image resolution
DRIVE	40 images	33 images	7 images	–	–	20 images	20 images	565 × 584 pixels
E-ophtha	In e-ophtha EX 82 images and e-ophtha MA 381 images	35 images in e-ophtha EX, 233 images in e-ophtha MA	–	–	–	–	–	Different image resolution
HRF	45 images	15 images	15 images	–	–	–	–	3504 × 2336 pixels
DDR	13,673 images	6266 images	630 images	4713 images	913 images	6835 images	4105 images	Different image resolution
Messidor	1200 images	–	–	–	–	–	–	Different image resolution
Messidor-2	1748 images	–	–	–	–	–	–	Different image resolution
STARE	20 images	10 images	–	–	–	–	–	700 × 605 pixels
CHASE DB1	28 images	–	–	–	–	–	–	1280 × 960 pixels
IDRiD	516 images	–	–	–	–	413 images	103 images	4288 × 2848 pixels
ROC	100 images	–	–	–	–	50 images	50 images	Different image resolution
DR2	435 images	–	–	–	–	–	–	857 × 569 pixels

After preprocessing the images, the images are ready to be used as an input for the DL, which is explained in the next section.

6. Diabetic retinopathy screening systems

Several researches have attempted to automate DR lesions detection and classification using DL. These methods can be categorized according to the classification method used as binary classification, multi-level classification, lesion-based classification, and vessels-based classification. Table 4 summarizes these methods.

6.1. Binary classification

This section summarizes the studies conducted to classify the DR dataset into two classes only. K. Xu et al. [41] automatically classified the images of the Kaggle [26] dataset into normal images or DR images using a CNN. They used 1000 images from the dataset. Data augmentation and resizing to 224*224*3 were performed before feeding the images to the CNN. Data augmentation was used to increase the dataset images by applying several transformations, such as rescaling, rotation, flipping, shearing and translation. The CNN architecture included eight CONV layers, four max-pooling layers and two FC layers. The SoftMax function was applied at the last layer of CNN for classification. This method had an accuracy of 94.5%.

In the study performed by G. Quellec et al. [37], each image was classified as referable DR (refer to moderate stage or more) or non-referable DR (No DR or mild stage) by training three CNNs. The images were taken from three datasets, namely, Kaggle (88,702 image) [26], DiaretDB1 (89 image) [25] and private E-ophtha (107,799 image) [28]. During the preprocessing stage, the images were resized, cropped to 448 × 448 pixels, normalized, and eroded the FOV by 5%. A large Gaussian filter was used and the augmented data were applied. The used CNNs architectures were pretrained AlexNet [19] and the two networks of o_O solution [48]. MA, HM, soft and hard EX were detected by the CNNs. This study had an area under the ROC curve of 0.954 in Kaggle and 0.949 in E-ophtha.

M. T. Esfahan et al. [22] used a known CNN, which is ResNet34 [49] in their study to classify DR images of the Kaggle dataset [26] into normal or DR image. ResNet34 is one available pretrained CNN architecture on ImageNet database. They applied a set of image preprocessing techniques to improve the quality of images. The image preprocessing included the Gaussian filter, weighted addition and image normalization. The image number was 35000 images and its size was 512 × 512 pixels. They reported an accuracy of 85% and a sensitivity of 86%.

R. Pires et al. [50] built their own CNN architecture to determine whether an image was referable DR. The proposed CNN contains 16 layers, which is similar to pretrained VGG-16 [51] and o_O team [48]. Two-fold cross-validation and multi-image resolution were used during training. The CNN of the 512 × 512 image input was trained after initializing the weights by the trained CNN on a smaller image resolution. The drop-out and L2 regularization techniques were applied to the CNN to reduce overfitting. The CNN was trained on the Kaggle dataset [26] and was tested by the Messidor-2 [31] and DR2 dataset. The classes of the training dataset were balanced using data augmentation. The work achieved an area under the ROC curve of 98.2% when testing the Messidor-2.

The study of H. Jiang et al. [52] integrated three pretrained CNN models, namely, Inception V3 [20], Inception-Resnet-V2 [53] and Resnet152 [21] to classify their own dataset as referable DR or non-referable DR. In CNNs training, Adam optimizer was used to update their weights. These models were integrated using the Adaboost algorithm. The dataset of 30,244 images was resized to 520 × 520 pixels, enhanced and augmented before being fed to the CNNs. The work obtained an accuracy of 88.21% and area under the curve (AUC) of 0.946.

Y. Liu et al. [54] built a weighted paths CNN (WP-CNN) to detect referable DR images. They collected over 60,000 images labeled as

referable or non-referable DR and augmented them many times to balance the classes. These images were resized to 299 × 299 pixels and were normalized before being fed to the CNN. The WP-CNN includes many CONV layers with different kernel sizes in different weighted paths that merged by taking the average. The WP-CNN of 105 layers had a better accuracy than pretrained ResNet [21], SeNet [55] and DenseNet [56] architectures with 94.23% in their dataset and 90.84% in the STARE dataset.

G. Zago et al. [57] detected DR red lesions and DR images based on augmented 65*65 patches using two CNN models. The CNNs used were pretrained VGG16 [51] and a custom CNN, which contains five CONV, five max-polling layers and a FC layer. These models were trained on the DIARETDB1 [25] dataset and tested on the DDR [23], IDRiD [34], Messidor-2, Messidor [58], Kaggle [26], and DIARETDB0 [59] datasets to classify patches into red lesions or non-red lesions. After that, the image with DR or non-DR were classified based on a lesion probability map of test images. The results of this work achieved the best sensitivity of 0.94 and an AUC of 0.912 for the Messidor dataset.

Unfortunately, the researchers who classified DR images into two classes did not consider the five DR stages. The DR stages are important to determine the exact stage of DR to treat the retina with the suitable process and to prevent the deterioration and blindness.

6.2. Multi-level classification

This section reviews the studies in which the DR dataset was classified into many classes. The work by V. Gulshan et al. [60] introduced a method to detect DR and diabetic macular edema (DME) using CNN model. They used Messidor-2 [31] and eyepacs-1 datasets which contain 1748 images and 9963 images, respectively to test the model. These images were first normalized, and the diameter was resized to 299 pixels wide before feeding them to the CNN. They trained 10 CNNs with the pretrained Inception-v3 [20] architecture with a various number of images, and the final result was computed by a linear average function. The images were classified into referable diabetic macular edema, moderate or worse DR, severe or worse DR, or fully gradable. They obtained a specificity of 93% in both datasets and 96.1% and 97.5% in sensitivity for the Messidor-2 and eyepacs-1 datasets, respectively; however, they did not explicitly detect non-DR or the five DR stage images.

M. Abramoff et al. [61] integrated a CNN with an IDX-DR device to detect and to classify DR images. They applied data augmentation to the Messidor-2 [31] dataset, which contains 1748 images. Their various CNNs were integrated using a Random Forest classifier to detect DR lesions as well as retina normal anatomy. The images in this work were classified as no DR, referable DR, or vision threatening DR. They reported an area under the curve of 0.980, a sensitivity of 96.8%, and a specificity of 87.0%. Unfortunately, they considered images of the mild DR stage as no DR, and the five DR stages were not considered.

H. Pratt et al. [42] proposed a method based on a CNN to classify images from the Kaggle dataset [26] into five DR stages. In the preprocessing stage, color normalization and image resizing to 512 × 512 pixels were performed. Their custom CNN architecture contained 10 CONV layers, eight max-pooling layers, and three FC layers. The Soft-Max function was used as a classifier for 80,000 test images. L2 regularization and dropout methods was used in CNN to reduce overfitting. Their results had a specificity of 95%, an accuracy of 75% and a sensitivity of 30%. Unfortunately, CNN does not detect the lesions in the images, and only one dataset was used to evaluate their CNN.

S. Dutta et al. [43] detected and classified DR images from the Kaggle dataset [26] into five DR stages. They investigated the performance of three networks, the back propagation neural network (BNN), the deep neural network (DNN), and the CNN, using 2000 images. The images were resized to 300 × 300 pixels and converted into grayscale, and the statistical features were extracted from the RGB images. Furthermore, a set of filters were applied, namely, edge detection, median filter,

Table 4

The methods used for DR detection/classification.

Ref.	DL method	Lesion detection	Dataset (Dataset size)	Performance measure			
				AUC	Accuracy	Sensitivity	Specificity
[60]	CNN (Inception-v3)	No	Messidor-2 (1748) and EyePACS-1 (9963)	–	–	96.1%	93.9%
[61]	CNN	yes	Messidor-2 (1748)	0.980	–	97.5%	93.4%
[42]	CNN	No	Kaggle (80,000)	–	75%	96.8%	87.0%
[41]	CNN	No	Kaggle (1000)	–	94.5%	30%	95%
[37]	CNN	yes	Kaggle (88,702), DiaretDB1 (89) and E-ophtha (107,799).	0.954	–	–	–
				0.949			
[38]	CNN	Red lesion only	DIARETDB1 (89), E-Ophtha (381) and MESSIDOR (1200)	CPM = 0.4874 for DIARETDB1 and CPM = 0.3683 for e-ophtha	0.4883	0.3680	–
[22]	CNN-ResNet34	No	Kaggle (35000)	–	85%	86%	–
[43]	DNN, CNN (VGGNET architecture), BNN	No	Kaggle (2000)	–	BNN = 42% DNN = 86.3% CNN = 78.3%	–	–
[44]	CNN (InceptionNet V3, AlexNet and VGG16)	No	Kaggle (166)	–	AlexNet = 37.43%, VGG16 = 50.03%, and InceptionNet V3 = 63.23%	–	–
[45]	CNN (AlexNet, VggNet, GoogleNet and ResNet)	No	Kaggle (35,126)	The higher is VggNet-s (0.9786)	The higher is VggNet-s (95.68%)	VggNet-16 achieved higher result (90.78%).	The higher is VggNet-s (97.43%)
[39]	CNN	MA only	E- Ophtha (381), ROC (100) and DIARETDB1 (89)	0.562 0.193 0.392	–	–	–
[40]	CNN	EX Only	DiaretDB0 (130), DiaretDB1 (89), and DrimDB (125).	–	99.17 98.53 99.18	100 99.2 100	98.41 97.97 98.44
[46]	Fully CNN	No	STARE (20), HRF (45), DRIVE (40) and CHASE DB1 (28).	0.9801 0.9701 0.9787 0.9752	0.9628 0.9608 0.9634 0.9664	0.8090 0.7762 0.7941 0.7571	0.9770 0.9760 0.9870 0.9823
[47]	Fully CNN	No	STARE (20), DRIVE (40) and CHASE_DB1 (28)	0.9905 0.9821 0.9855	0.9694 0.9576 0.9653	0.8315 0.8039 0.7779	0.9858 0.9804 0.9864
[63]	CNN (AlexNet, VggNet16, custom CNN)	No	MESSIDOR (1200)	–	98.15%	98.94%	97.87%
[65]	CNN (ResNet50, InceptionV3, InceptionResNetV2, Xception and DenseNets)	No	Their Own dataset (13767)	–	96.5%	98.1%	98.9%
[50]	CNN	No	Messidor-2 (1748), Kaggle (88,702) and DR2 (520)	98.2% – 98%	–	–	–
[68]	CNN (AlexNet)	No	Kaggle (22,700) and IDRiD (516)	–	90.07%	–	–
[52]	CNN (Inception V3, Inception-Resnet-V2 and Resnet152)	No	Their Own dataset (30244)	0.946	88.21%	85.57%	90.85%
[54]	CNN (WP-CNN, ResNet, SeNet and DenseNet)	No	Their Own dataset (60,000), and STARE (131)	0.9823 0.951	94.23% 90.84%	90.94%	95.7–4%
[74]	CNN (improved LeNet, U-net)	Red lesion only	DIARETDB1 (89)	CPM = 0.4823	48.71%	–	–
[57]	CNN (VGG16, custom CNN)	Red lesion only	DIARETDB1 (89), DIARETDB0 (130), Kaggle (15,919), Messidor (1200), Messidor-2 (874), IDRiD (103) and DDR (4105)	– 0.786 0.764 0.912 – 0.818 0.848	– 0.821 0.911 0.94 – 0.841 0.891	–	–
[23]	CNN (GoogLeNet, ResNet-18, DenseNet-121, VGG-16 and SE-BN-Inception)	No	DDR (13,673)	–	0.8284	–	–
[69]	CNN (modified Alexnet)	No	Messidor (1190)	–	96.35	92.35	97.45
[78]	CNN	No	HRF (45) and DRIVE (40)	0.894	93.94%	–	–
[81]	CNN (ResNet-101)	No	DRIVE (40)	0.9732	0.951	0.793	0.974
[75]	CNN	EX only	E-ophtha (82) and HEI-MED (169)	0.9644 0.9323	– 0.8990	0.9477 0.9447	–
[77]	Deep residual network	EX only	E-ophtha (82) and HEI-MED (169)	0.9647 0.9709	– 0.9227	0.9227 0.9255	–
[82]	CNN	No	DRIVE (40) and STARE (20)				

(continued on next page)

Table 4 (continued)

Ref.	DL method	Lesion detection	Dataset (Dataset size)	Performance measure			
				AUC	Accuracy	Sensitivity	Specificity
[83]	CNN	No	DRIVE (40), STARE (20) and CHASE (28).	0.9822	0.9685	0.7439	0.99
				0.9868	0.9735	0.8196	0.9871
				98.30%	95.82%	79.96% 79.63%	98.13%
				98.75%	96.72%	80.03%	98.63%
[84]	CNN	No	DRIVE (40) and CHASE_DB1 (28).	0.9844	0.9688%		98.80%
				0.9560	0.9580	0.8639	0.9665
[70]	CNN	Red lesion only	Their dataset (9194) and Messidor (1200)	0.9577	0.9601	0.8778	0.9680
				–	92.95	99.39%	99.93%
[72]	CNN (ResNet50)	No	Messidor (1200) and IDRiD (516)	0.972	–	92.59%	96.20%
				96.3%	92.6%	92%	–
				–	65.1%	–	

morphological processing, and binary conversion, before being fed into the networks. Pretrained VGG16 [51] was used as the CNN architecture, which includes 16 CONV layers and 4 max pooling layers and three FC layers while the DNN includes three FC layers. Their results shown that the DNN outperforms the CNN and the BNN. Unfortunately, few images were used for networks training, and thus the networks could not learn more features. Also, only one dataset was used to evaluate their study.

X. Wang et al. [44] studied the performance of the three available pretrained architectures of CNN, VGG16 [51], AlexNet [19] and InceptionNet V3 [20], to detect the five DR stages in the Kaggle [26] dataset. The images were resized to 224×224 pixels for VGG16, 227×227 pixels for AlexNet, and 299×299 pixels for InceptionNet V3 at the preprocessing stage. The dataset only contains 166 images. They reported an average accuracy of 50.03% in VGG16, 37.43% in AlexNet and 63.23% in InceptionNet V3; however, they trained the networks with limited number of images, which could prevent the CNN from learning more features and the images required more preprocessing functions to improve them. Also, only one dataset was used to evaluate their study.

The performance of four available pretrained architectures of the CNN was investigated in Ref. [45]: AlexNet [19], ResNet [21], GoogleNet [62] and VggNet [51]. These architectures were trained to detect the five DR stages from the Kaggle [26] dataset, which contains 35,126 images. Transfer learning these CNNs was done by fine tuning the last FC layer and hyperparameter. During the preprocessing stage, the images were augmented, cropped, normalized and the NonLocal Means Denoising function was applied. This study achieved an accuracy of 95.68%, AUC of 0.9786 and a specificity of 97.43% for VggNet-s, which had a higher accuracy, specificity, and an AUC than the other architectures. The use of more than one dataset makes a system more reliable and able to generalize [83]. Unfortunately, the study only included one dataset and their method does not detect the DR lesions.

Mobeen-ur-Rehman et al. [63] detected the DR levels of the MES-SIDOR dataset [31] using their custom CNN architecture and pretrained models, including AlexNet [19], VGG-16 [51] and SqueezeNet [64]. This dataset contains 1200 images classified into four DR stages. The images were cropped, resized to a fixed size, which was 244×244 pixel, and enhanced by applying the histogram equalization (HE) method at the pre-processing stage. The custom CNN includes five layers: two CONV layers, two max-pooling layers, and three FC layers. They reported the best accuracy of 98.15%, specificity of 97.87% and sensitivity of 98.94% by their custom CNN. Unfortunately, only one dataset was used to evaluate their CNN and does not detect the DR lesions.

W. Zhang et al. [65] proposed a system to detect the DR of their own dataset. The dataset includes 13,767 images, which are grouped into four classes. These images were cropped, resized to the required size of each network, and improved by applying HE and adaptive HE. In addition, the size of the training images was enlarged by data augmentation, and the contrast was improved by a contrast stretching algorithm that is used for dark images. They finetuned pretrained CNN

architectures: ResNet50 [66], InceptionV3 [20], InceptionResNetV2 [53], Xception [67], and DenseNets [56] to detect the DR. Their approach involved training the added new FC layers on top of these CNNs. After that, they finetuned some layers of the CNNs to retrain it. Lastly, the strong models were integrated. This approach achieved an accuracy of 96.5%, a specificity of 98.9% and a sensitivity of 98.1%. Unfortunately, CNNs do not detect the lesions in the images and only one private dataset was used to evaluate their method.

B. Harangi et al. [68] integrated the available pretrained AlexNet [19] and the hand-crafted features to classify the five DR stages. The CNN was trained by the Kaggle dataset [26] and tested by the IDRiD [34]. The obtained accuracy for this study was 90.07%. Unfortunately, the work does not detect the lesions in the images and only one dataset was used to test their method.

T. Li et al. [23] detected DR stages in their dataset (DDR) by finetuning the GoogLeNet [62], ResNet-18 [21], DenseNet-121 [56], VGG-16 [51], and SE-BN-Inception [55] available pretrained networks. Their dataset includes 13,673 fundus images. During preprocessing, the images were cropped, resized to 224×224 pixels, augmented and resampled to balance the classes. The SE-BN-Inception network obtained the best accuracy at 0.8284. Unfortunately, the work does not detect the lesions in the images and only one dataset was used to test their method.

T. Shanthi and R. Sabeeenian [69] detected the DR stages of the Messidor dataset [31] using a pretrained architecture Alexnet [19]. The images were resized, and the green channel was extracted before being fed into the CNN. This CNN achieved an accuracy of 96.35. Unfortunately, the work does not detect the lesions in the images and only one dataset and architecture were used to test their method.

J. Wang et al. [70] modified a R-FCN method [71] to detect DR stages in their private dataset and the public Messidor dataset [58]. Moreover, they detected MA and HM in their dataset. They modified the R-FCN by adding a feature pyramid network and also adding five region proposal networks rather than one to the method. The lesion images were augmented for training. The obtained sensitivity for detecting DR stages were 99.39% and 92.59% in their dataset and the Messidor dataset, respectively. They reported a PASCAL-VOC AP of 92.15 in lesion detection. Unfortunately, the study only evaluated the method on one public dataset and only detected HM and MA without detecting EX.

X. Li et al. [72] classified the public Messidor [58] dataset into referable or non-referable images and classified the public IDRiD dataset [34] into five DR stages and three DME stages by using the ResNet50 [21] and four attention modules. The features extracted by ResNet50 used as the inputs for the first two attention modules to select one disease features. The first two attention modules contain average pooling layers, max-pooling layers, multiplication layers, concatenation layer, CONV layer and FC layers while the next two attention modules contain FC and multiplication layers. Data augmentation, normalization and resizing were performed before feeding the images to the CNN. This work achieved a sensitivity of 92%, an AUC of 96.3% and an accuracy of 92.6% for the Messidor dataset and an accuracy of 65.1% for the IDRiD.

Unfortunately, the study does not detect the lesions in the images.

6.3. Lesion-based classification

This section summarizes the works performed to detect and to classify certain types of DR lesions. For example, J. Orlando et al. [38] detected only red lesions in DR images by incorporating DL methods with domain knowledge for feature learning. Then, the images were classified by applying the Random Forest method. The images of the MESSIDOR [58], E-ophtha [73] and DIARETDB1 [25] datasets were processed by extracting the green band and expanding the FOV, and applying a Gaussian filter, r-polynomial transformation, thresholding operation and, many morphological closing functions. Next, red lesion patches were resized to 32*32 pixels and were augmented for CNN training. The datasets contain 89 images, 381 images and 1200 images in DIARETDB1, E-ophtha and MESSIDOR, respectively. Their custom CNN contains four CONV layers, three pooling layers and one FC layer. They achieved a Competition Metric (CPM) of 0.4874 and 0.3683 for the DIARETDB1 and the E-ophtha datasets, respectively.

P. Chudzik et al. [39] used custom CNN architecture to detect MA from DR images. Three datasets were used in this study: ROC [35] (100 images), E-ophtha [73] (381 images), and DIARETDB1 [25] (89 images). These datasets were processed by extracting the green plane and then performing cropping, resizing, applying Otsu thresholding to generate a mask, and utilizing a weighted sum and morphological functions. Next, MA patches were extracted, and random transformations were applied. The used CNN includes 18 CONV layers, and each CONV layer is followed by a batch normalization layer, three max-pooling layers, three simple up-sampling layers, and four skip connections between both paths. They reported a ROC score of 0.355.

The system proposed by Refs. [40], detected the exudates from DR images using the custom CNN with Circular Hough Transformation (CHT). They used three public datasets: the DiaretDB0 dataset includes 130 images, the DiaretDB1 dataset contains 89 images and the DrimDB dataset has 125 images. All the datasets were converted into grayscale. Then, Canny edge detection and adaptive histogram equalization functions were applied. Next, the optical disc was detected by CHT and then removed from the images. The 1152*1152 pixels of the images were fed into the custom CNN, which contains three CONV layers, three max pooling layers, and an FC layer that uses SoftMax as a classifier. The accuracies of detecting exudates were 99.17, 98.53, and 99.18 for DiaretDB0, DiaretDB1, and DrimDB, respectively.

Y. Yan et al. [74] detected DR red lesions in the DIARETDB1 [25] dataset by integrating the features of a handcrafted and improved pre-trained LeNet architecture using a Random Forest classifier. The green channel of the images was cropped, and they were enhanced by CLAHE. Also, noise was removed by the Gaussian filter, and a morphological method was used. After that, the blood vessels were segmented from images by applying the U-net CNN architecture. The improved LeNet architecture includes four CONV layers, three max-pooling layers, and one FC layer. This work achieved a sensitivity of 48.71% in red lesions detection.

H. Wang et al. [75] detected hard exudate lesion in the E-ophtha dataset [28] and the HEI-MED dataset [76] by integrating the features of a handcrafted and custom CNN using a Random Forest classifier. These datasets were processed by performing cropping, color normalizing, modifying a camera aperture and detecting the candidates by using morphological construction and dynamic thresholding. After that, patches of size 32*32 are collected and augmented. The custom CNN includes three CONV and three pooling layers and a FC layer to detect the patches features. This work achieved a sensitivity of 0.8990 and 0.9477 and an AUC of 0.9644 and 0.9323 for the E-ophtha and HEI-MED datasets, respectively.

J. Mo et al. [77] detected exudate lesions in the public available E-ophtha [28] and the HEI-MED [76] datasets by segmenting and classifying the exudates using deep residual network. The exudates were

segmented using a fully convolutional residual network which contains up-sampling and down-sampling modules. After that, the exudates were classified using a deep residual network which includes one CONV layer, one max-pooling layer and 5 residual blocks. The down-sampling module includes CONV layer followed by a max pooling layer and 12 residual blocks while the up-sampling module comprises CONV and deconvolutional layers to enlarge the image as the input image. The residual block includes three CONV layers and three batch normalization layers. This work achieved a sensitivity of 0.9227 and 0.9255 and an AUC of 0.9647 and 0.9709 for the E-ophtha and HEI-MED datasets, respectively.

Unfortunately, these studies detected only some DR lesions without considering the five DR stages. Furthermore, they used a limited number of images for DL methods.

6.4. Vessel-based classification

Vessel segmentation is used to diagnosis and to evaluate the progress of retinal diseases, such as glaucoma, DR and hypertension. Many studies have been conducted to investigate vessel segmentation as part of DR detection. DR lesions remain in the image after the vessels have been extracted. Therefore, detecting the remaining lesions in the images lead to detect and classify DR images. The study in Ref. [74] detected the red lesions after vessels were extracted. Some studies on vessel segmentation used DL methods, which is reviewed in this section.

Sunil et al. [78] used a modified CNN of pretrained DEEPLAB-COCO-LARGEFOV [79] to extract the retinal blood vessels from RGB retina images. They extracted 512×512 image patches from the dataset and then fed them to the CNN. After that, they applied a threshold to binarize the images. The CNN includes eight CONV layers and three max-pooling layers. The HRF [30], DRIVE [29] datasets were used to evaluate the method. They reported an accuracy of 93.94% and an area under the ROC of 0.894.

The study conducted by Refs. [46] used fully CNN to segments the blood vessels in RGB retina images. The images of the STARE [32], HRF [30], DRIVE [29] and CHASE DB1 [33] datasets were preprocessed by applying morphological methods, flipped horizontally, adjusted to different intensities, and cropped into patches. Then they were fed to the CNN for segmentation and to condition random field model [80] to consider the non-local correlations during segmentation. After that, the vessel map was rebuilt, and morphological operations were applied. Their CNN contains 16 CONV layers and five dilated CONV layers. The STARE, HRF, DRIVE and CHASE DB1 datasets contain 20, 45, 40, and 28 images, respectively were used. An accuracy of 0.9634, 0.9628, 0.9608 and 0.9664 was achieved for the DRIVE, STARE, HRF and CHASE DB1, respectively.

The work conducted by Refs. [47] included the Stationary Wavelet Transform (SWT) with a fully CNN to extract the vessels from the images. The STARE (20 images) [32], DRIVE (40 images) [29] and CHASE_DB1 (28 images) [33] datasets were preprocessed by extracting the green channel and normalizing images, and SWT was applied. Next, the patches were extracted and augmented. Then, the patches were fed to the CNN which includes CONV layers, max-pooling layers, crop layer, and SoftMax classifier and up-sampling layer that return the feature maps to the previous dimensions. The results of this study reached an AUC of 0.9905, 0.9821 and 0.9855 for the STARE, DRIVE and CHASE_DB1 datasets, respectively.

Cam-Hao et al. [81] extracted retinal vessels from the DRIVE dataset [29]. They selected four feature maps from the pretrained ResNet-101 [21] network and then combined each feature map with its neighbor. After that, the feature map outputs were also combined until one feature map was obtained. Next, each round of the best resolution feature maps was concatenated. They augmented the training images before being fed to the network. They achieved a sensitivity of 0.793, an accuracy of 0.951, a specificity of 0.9741 and an AUC of 0.9732.

Ü. Budak et al. [82] extracted retinal vessels from the DRIVE [29]

and STARE [32] public datasets using custom CNN architecture. The custom CNN includes three blocks of concatenated encoder-decoder and two CONV layers between them. Each block contains eight CONV layers, eight Batch Normalization layers, two max-unpooling layers and two max-pooling layers. They cropped images, extracted and augmented patches before feeding them into the CNN for training. They reported an accuracy of 0.9685, 0.9735 an AUC of 0.9822 and 0.9868 for the DRIVE and STARE datasets, respectively.

Y. Wu et al. [83] used a custom CNN to extract the retinal blood vessels from the DRIVE [29], STARE [32] and CHASE [33] public datasets. They converted RGB images to grayscale, normalized and enhanced them by CLAHE. After that, they extracted and augmented 48×48 patches from the datasets and fed them to the CNN. The CNN includes of two networks that have encoder-decoder structure and skip connections. The encoder-decoder structure has CONV layers, Batch Normalization layers, concatenation layers and dropout layers. They achieved an accuracy of 95.82%, 96.72% and 96.88%, and an AUC of 98.30%, 98.75% and 98.94% for the DRIVE, STARE and CHASE, respectively.

C. Tian et al. [84] extracted retinal vessels from the DRIVE [29] and CHASE_DB1 [33] public datasets using custom multi-path CNN architecture and Gaussian matched filter. They obtained a high-frequency and a low-frequency images from Gaussian filter. Then they constructed CNN path for low-frequency images that composed of CONV layers, down-sampling and up-sampling modules and constructed other CNN path for high-frequency images that composed of two CONV layers and seven encoder-decoder blocks. The encoder-decoder structure has dilated CONV layers, down-sampling and up-sampling modules. The segmentation maps were extracted from the two paths and then merged them for final segmentation results. They achieved an accuracy of 0.9580 and 0.9601 and an AUC of 0.9560 and 0.9577 for the DRIVE and CHASE_DB1, respectively.

Unfortunately, these works only considered segmented vessels and did not detect DR stages or DR lesions.

7. Discussion section

The current study reviewed 33 papers. All of the studies mentioned in the current work manipulated the diabetic retinopathy screening system using deep learning techniques. The need for reliable diabetic retinopathy screening systems became a critical issue recently due to the increase in the number of diabetic patients. Using DL in DR detection and classification overcomes the problem of selecting reliable features for ML; on the other hand, it needs a huge data size for training. Most studies used data augmentation to increase the number of images and overcoming overfitting on training stage. The studies covered on the current work 94% of them used public datasets, 59% of them used a combination of two or more public datasets to overcome the problem of data size and to evaluate the DL methods on many datasets as shown in Fig. 6.

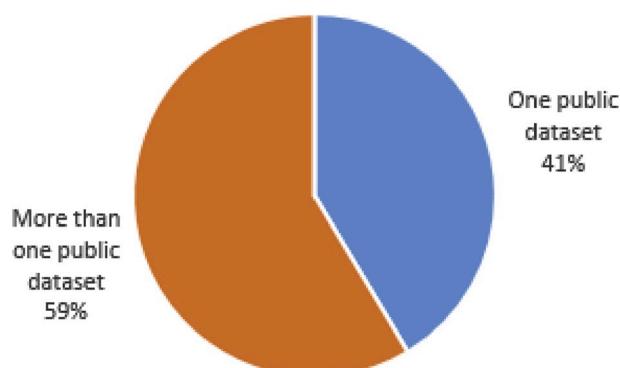


Fig. 6. The percentage of studies that used one or more public datasets.

One of the limitations of the usage of deep learning with medical field faces is the size of the datasets needed to train the DL systems, as DL is required large amount of data. The results of DL systems depend heavily on the size of the training data as much as its quality and balance its classes. So, the current public datasets sizes need to be increased, while the big size one like public Kaggle dataset needs to be refined to eliminates miss labeled and low-quality data.

The covered studies here varied in using DL techniques. They differ in the number of studies that built their own CNN structure, and those who preferred to use the existing structures, such as VGG, ResNet, or AlexNet, with transfer learning is slightly small. Building a new CNN architecture from scratch needs a lot of effort and time consuming while using transfer learning is much easier and speed up the process of structuring and developing new architecture. On the other hand, it is notable that the accuracy of the system which built their own CNN structure is higher than those using the existing structures. This point needs to be focused by the researchers and more studies should be conducted to judge among the two trends.

Most of the studies covered here (73%) only classified the fundus input image to DR non-DR, while 27% classified input to one or more stages as shown in Fig. 7. On the other hands 70% of the current studies didn't detect the affected lesions while, 30% of them detected the affected lesions. Among them, only 6% of the studies succeeded in classifying images and detecting the type of the affected lesion on the retain image as shown in Fig. 8. The existence of a reliable DR screening system capable of detecting different lesions types and DR stages leads to an effective follow up system for DR patients, which averted the danger of losing sight. The gap that needed to be covered is the existence of systems that could determine the five DR stages with high accuracy as well as detecting DR lesions. This point could be considered as the current challenge for researchers for further investigations.

8. Conclusion

Automated screening systems significantly reduce the time required to determine diagnoses, saving effort and costs for ophthalmologists and result in the timely treatment of patients. Automated systems for DR detection play an important role in detecting DR at an early stage. The DR stages are based on the type of lesions that appear on the retina. This article has reviewed the most recent automated systems of diabetic retinopathy detection and classification that used deep learning techniques. The common fundus DR datasets that are publicly available have been described, and deep-learning techniques have been briefly explained. Most researchers have used the CNN for the classification and the detection of the DR images due to its efficiency. This review has also discussed the useful techniques that can be utilized to detect and to classify DR using DL.

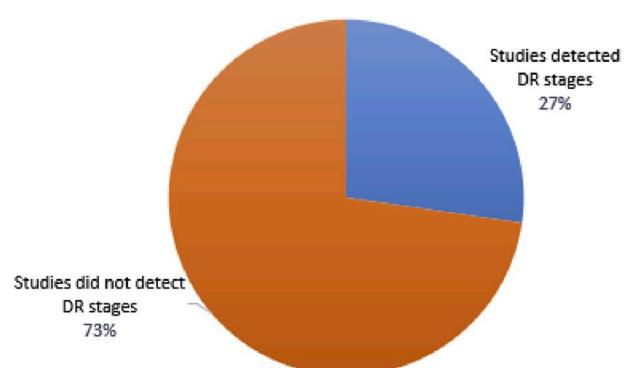


Fig. 7. The percentage of studies that detected DR stages.

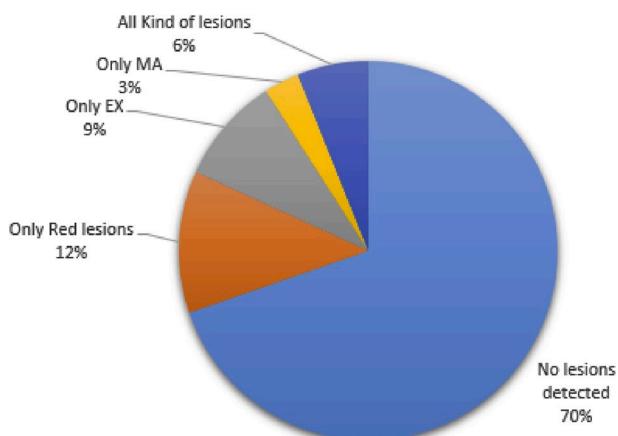


Fig. 8. The percentage of studies that detected DR lesions.

Ethical statement

This work did not receive any grant from funding agencies.

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.

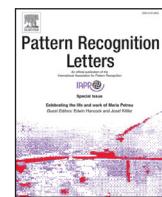
Acknowledgements

We thank King Abdul-Aziz University for encouraging us to do this work.

References

- [1] Taylor R, Batey D. Handbook of retinal screening in diabetes: diagnosis and management. second ed. John Wiley & Sons, Ltd Wiley-Blackwell; 2012.
- [2] International diabetes federation - what is diabetes [Online]. Available, <https://www.idf.org/aboutdiabetes/what-is-diabetes.html>.
- [3] American academy of ophthalmology-what is diabetic retinopathy? [Online]. Available, <https://www.aao.org/eye-health/diseases/what-is-diabetic-retinopathy>.
- [4] Bourne RR, et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Global Health* 2013;1(6):339–49.
- [5] Harper CA, Keeffe JE. Diabetic retinopathy management guidelines. *Expert Rev Ophthalmol* 2012;7(5):417–39.
- [6] E. T. D. R. S. R. GROUP. Grading diabetic retinopathy from stereoscopic color fundus photographs- an extension of the modified Airlie House classification. *Ophthalmology* 1991;98(5):786–806.
- [7] Scanlon PH, Wilkinson CP, Aldington SJ, Matthews DR. A Practical manual of diabetic retinopathy management. first ed. Wiley-Blackwell; 2009.
- [8] Dubow M, et al. Classification of human retinal microaneurysms using adaptive optics scanning light ophthalmoscope fluorescein angiography. *Investig Ophthalmol Vis Sci* 2014;55(3):1299–309.
- [9] Bandello F, Zarbin MA, Lattanzio R, Zucchiatti I. Clinical strategies in the management of diabetic retinopathy. second ed. Springer; 2019.
- [10] Scotland GS, et al. Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy. *Br J Ophthalmol* 2010; 94(6):712–9.
- [11] deng Li. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Trans. Signal Inf Process* 2014;3(2):1–29.
- [12] V Vasilakos A, Tang Y, Yao Y. Neural networks for computer-aided diagnosis in medicine : a review. *Neurocomputing* 2016;216:700–8.
- [13] Wilkinson CP, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Am Acad Ophthalmol* 2003;110(9): 1677–82.
- [14] Chen XW, Lin X. Big data deep learning: challenges and perspectives. *IEEE Access* 2014;2:514–25.
- [15] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: a review. *Neurocomputing* 2016;187:27–48.
- [16] Deng L, Yu D. Deep learning: methods and applications. *Found Trends® Signal Process* 2014;7(3–4):197–387.
- [17] Bakator M, Radosav D. Deep learning and medical diagnosis: a review of literature. *Multimodal Technol Interact* 2018;2(3):47.
- [18] Mookiah M, Acharya UR, Kuang C, Min C, Ng EYK, Laude A. Computer-aided diagnosis of diabetic retinopathy : a review. *Comput Biol Med* 2013;43:2136–55.
- [19] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60(6).
- [20] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *IEEE Conference on computer Vision and pattern recognition*; 2016. p. 2818–26.
- [21] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *IEEE Conference on computer Vision and pattern recognition*; 2016. p. 770–8.
- [22] Esfahani MT, Ghaderi M, Kafiyeh R. Classification of diabetic and normal fundus images using new deep learning method. *Leonardo Electron J Pract Technol* 2018; 17(32):233–48.
- [23] Li T, Gao Y, Wang K, Guo S, Liu H, Kang H. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Inf Sci* 2019;501:511–22.
- [24] Abràmoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. *IEEE Rev Biomed Eng* 2010;3:169–208.
- [25] Kauppi T, et al. The DIARETDB1 diabetic retinopathy database and evaluation protocol. In: *Proceedings of the British machine vision conference 2007*; 2007. p. 1–10.
- [26] Kaggle dataset [Online]. Available, <https://kaggle.com/c/diabetic-retinopathy-detection>.
- [27] Lam C, Yi D, Guo M, Lindsey T. Automated detection of diabetic retinopathy using Deep Learning. *AMIA Joint Summits on Translational Science* 2018;2017:147–55.
- [28] Decencière E, et al. TeleOpta : machine learning and image processing methods for teleophthalmology. *IRBM* 2013;34(2):196–203.
- [29] Staal JJ, Abràmoff MD, Niemeijer M, Viergever MA, van Ginneken B. Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imag* 2004;23 (4):501–9.
- [30] Budai R, Bock A, Maier J, Hornegger J, Michelson G. Robust vessel segmentation in fundus images,” *Int J Biomed Imag*, vol. 2013, 2013.
- [31] Decencière E, et al. Feedback on a publicly distributed image database: the messidor database. *Image Anal Stereol* 2014;33(3):231–4.
- [32] Hoover A, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. *IEEE Trans Med Imag* 2000;19(3):203–10.
- [33] Owen CG, et al. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program. *Investig Ophthalmol Vis Sci* 2009;50(5):2004–10.
- [34] Porwal P, et al. Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. *Data* 2018;3(3).
- [35] ROC dataset [Online]. Available, <http://roc.healthcare.uiowa.edu>.
- [36] DR2 [Online]. Available, https://figshare.com/articles/Advancing_Bag_of_Visual_Words_Representations_for_Lesion_Classification_in_Retinal_Images/953671.
- [37] Quellec G, Charrère K, Boudi Y, Cochener B, Lamard M. Deep image mining for diabetic retinopathy screening. *Med Image Anal* 2017;39:178–93.
- [38] Orlando JI, Prokofyeva E, del Fresno M, Blaschko MB. An ensemble deep learning based approach for red lesion detection in fundus images. *Comput Methods Progr Biomed* 2018;153:115–27.
- [39] Chudzik P, Majumdar S, Calivá F, Al-Diri B, Hunter A. Microaneurysm detection using fully convolutional neural networks. *Comput Methods Progr Biomed* 2018; 158:185–92.
- [40] Adem K. Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks. *Expert Syst Appl* 2018;114: 289–95.
- [41] Xu K, Feng D, Mi H. Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. *Molecules* 2017;22(12): 2054.
- [42] Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. *Procedia Comput Sci* 2016;90:200–5.
- [43] Dutta S, Manideep BC, Basha SM, Caytiles RD, Iyengar NCSN. Classification of diabetic retinopathy images by using deep learning models. *Int J Grid Distr Comput* 2018;11(1):99–106.
- [44] Wang X, Lu Y, Wang Y, Chen WB. Diabetic retinopathy stage classification using convolutional neural networks. In: *International Conference on information Reuse and Integration for data science*; 2018. p. 465–71.
- [45] Wan S, Liang Y, Zhang Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. *Comput Electr Eng* 2018;72:274–82.
- [46] Lu J, Xu Y, Chen M, Luo Y. A coarse-to-fine fully convolutional neural network for fundus vessel segmentation. *Symmetry* 2018;10(11):607.
- [47] Oliveira A, Pereira S, Silva CA. Retinal vessel segmentation based on fully convolutional neural networks. *Expert Syst Appl* 2018;112:229–42.
- [48] Team o-O solution [Online]. Available, <https://www.kaggle.com/c/diabetic-retinopathydetection/discussion/15617>.
- [49] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: *The thirty-first AAAI conference on artificial intelligence*; 2016. p. 4278–84.
- [50] Pires R, Avila S, Wainer J, Valle E, Abramoff MD, Rocha A. A data-driven approach to referable diabetic retinopathy detection. *Artif Intell Med* 2019;96:93–106.
- [51] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *International Conference on learning representations*; 2015.
- [52] Jiang H, Yang K, Gao M, Zhang D, Ma H, Qian W. An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In: *2019 41st Annual International conference of the IEEE engineering in medicine and biology society (EMBC)*; 2019. p. 2045–8.

- [53] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI Conference on artificial intelligence; 2017. p. 4278–84.
- [54] Liu YP, Li Z, Xu C, Li J, Liang R. Referable diabetic retinopathy identification from eye fundus images with weighted path for convolutional neural network. *Artif Intell Med* 2019;99:101694.
- [55] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 7132–41.
- [56] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: The IEEE Conference on computer Vision and pattern recognition (CVPR); 2017. p. 4700–8.
- [57] Zago GT, Andreão RV, Dorizzi B, Teatin Salles EO. Diabetic retinopathy detection using red lesion localization and convolutional neural networks. *Comput Biol Med* 2019;116:103537.
- [58] Messidor dataset [Online]. Available, <http://messidor.crihan.fr>.
- [59] Kauppi T, et al. DIARETDB0 : evaluation database and methodology for diabetic retinopathy algorithms. In: Machine Vision and pattern recognition research group; 2006. p. 1–17.
- [60] Gulshan V, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Am Med Assoc* 2016;316:2402–10.
- [61] Abràmoff MD, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investig Ophthalmol Vis Sci* 2016;57(13):5200–6.
- [62] Szegedy C, et al. Going deeper with convolutions. In: IEEE Conference on computer Vision and pattern recognition (CVPR); 2015. p. 1–9.
- [63] Mobeen-Ur-Rehman, Khan SH, Abbas Z, Danish Rizvi SM. Classification of diabetic retinopathy images based on customised CNN architecture. In: Proceedings - 2019 Amity International Conference on artificial intelligence, AICAI 2019; 2019. p. 244–8.
- [64] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. 2016. p. 1–13. arXiv:1602.07360.
- [65] Zhang W, et al. Automated identification and grading system of diabetic retinopathy using deep neural networks. *Knowl Base Syst* 2019;175:12–25.
- [66] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 2015;37(9):1904–16.
- [67] Chollet F. Xception: deep learning with depthwise separable convolutions. In: The IEEE Conference on computer Vision and pattern recognition (CVPR); 2017. p. 1251–8.
- [68] Harangi B, Toth J, Baran A, Hajdu A. Automatic screening of fundus images using a combination of convolutional neural network and hand-crafted features. In: 2019 41st annual International Conference of the IEEE Engineering in Medicine and biology society (EMBC); 2019. p. 2699–702.
- [69] Shanhi T, Sabenian RS. Modified Alexnet architecture for classification of diabetic retinopathy images. *Comput Electr Eng* 2019;76:56–64.
- [70] Wang J, Luo J, Liu B, Feng R, Lu L, Zou H. Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm. *IET Comput Vis* 2020;14(1):1–8.
- [71] Dai J, Li Y, He K, Sun J. R-FCN: object detection via region-based fully convolutional networks. In: Advances in neural information processing systems; 2016. p. 379–87.
- [72] Li X, Hu X, Yu L, Zhu L, Fu C-W, Heng P-A. CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading. *IEEE Trans Med Imag* 2020;39(5):1483–93.
- [73] E-Ophtha dataset [Online]. Available, <http://www.adcis.net/en/Download-Third-Party/E-Ophtha.html>.
- [74] Yan Y, Gong J, Liu Y. A novel deep learning method for red lesions detection using hybrid feature. In: Proceedings of the 31st Chinese Control and decision conference, CCDC 2019; 2019. p. 2287–92.
- [75] Wang H, et al. Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening. *Comput Methods Progr Biomed* 2020;191:105398.
- [76] Giancardo L, et al. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Med Image Anal* 2012;16(1):216–26.
- [77] Mo J, Zhang L, Feng Y. Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks. *Neurocomputing* 2018;290:161–71.
- [78] Vengalil SK, Sinha N, Kruthiventi SSS, Babu RV. Customizing CNNs for blood vessel segmentation from fundus images. In: International Conference on signal Processing and communications, SPCOM 2016; 2016. p. 1–4.
- [79] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected crfs. 2014. arXiv preprint arXiv:1412.7062.
- [80] Zheng S, et al. Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 1529–37.
- [81] Hua CH, Huynh-The T, Lee S. Retinal vessel segmentation using round-wise features aggregation on bracket-shaped convolutional neural networks. In: Proceedings of the annual International Conference of the IEEE Engineering in Medicine and biology society, EMBS; 2019. p. 36–9.
- [82] Budak Ü, Cömert Z, Çibuk M, Şengür A. DCCMED-Net: densely connected and concatenated multi Encoder-Decoder CNNs for retinal vessel extraction from fundus images. *Med Hypotheses* 2020;134.
- [83] Wu Y, Xia Y, Song Y, Zhang Y, Cai W. NFN + : a novel network followed network for retinal vessel segmentation. *Neural Network* 2020;126:153–62.
- [84] Tian C, Fang T, Fan Y, Wu W. Multi-path convolutional neural network in fundus segmentation of blood vessels. *Biocybern Biomed Eng* 2020;40(2):583–95.



Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset

Abhishek Samanta^a, Aheli Saha^a, Suresh Chandra Satapathy^{a,*}, Steven Lawrence Fernandes^b, Yo-Dong Zhang^c

^a School of Computer Engineering, Kalinga Institute of Industrial Technology (Deemed to Be University), Bhubaneswar, Odisha 751024, India

^b Department of Electronics and Communication Engineering, Sahyadri College of Engineering and Management, Mangaluru 575007, India

^c Department of Informatics, University of Leicester, Leicester LE1 7RH, UK



ARTICLE INFO

Article history:

Received 30 December 2019
Revised 13 April 2020
Accepted 17 April 2020
Available online 12 May 2020

Keywords:

Diabetic Retinopathy
CNN architecture
Colour fundus photography

ABSTRACT

Diabetic Retinopathy is a complication based on patients suffering from type-1 or type-2 diabetes. Early detection is essential as complication can lead to vision problems such as retinal detachment, vitreous hemorrhage and glaucoma. The principal stages of diabetic retinopathy are non-Proliferative diabetic retinopathy and Proliferative diabetic retinopathy. In this paper, we propose a transfer learning based CNN architecture on colour fundus photography that performs relatively well on a much smaller dataset of skewed classes of 3050 training images and 419 validation images in recognizing classes of Diabetic Retinopathy from hard exudates, blood vessels and texture. This model is extremely robust and lightweight, garnering a potential to work considerably well in small real time applications with limited computing power to speed up the screening process. The dataset was trained on Google Colab. We trained our model on 4 classes - I)No DR ii)Mild DR iii)Moderate DR iv)Proliferative DR, and achieved a Cohens Kappa score of 0.8836 on the validation set along with 0.9809 on the training set.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Deep learning has enhanced the purpose of computer vision in identifying and classifying images and are a key tool used to automate tasks in our daily lives. Convolutional networks have been consistently developed for object detection, classification, segmentation. The use of convolutional neural networks (CNNs) on medical images has helped the medical sector immensely due to it's ability to learn representations of data [1].

Diabetic Retinopathy turns out to be a major cause of blindness in the western world, and regular screening of the patients reduces the risk of blindness. There are a number of features pertaining to the recognition of retinopathy in fundus photography, and computer vision based trained classifiers work pretty well in classification. Promising work has been displayed in the detection of retinopathy using k-NN classifiers and vector machines. CNNs have also been used for the classification of Diabetic Retinopathy [13,15], given a big dataset and considerable computing power [18]. They have been instrumental in detecting the features such as haemorrhage and hard exudes that identify retinopathy [8]. Deep architec-

tures of CNN have been instrumental in providing the finesse and high performance to trained models by learning patterns from raw images [2]. Due to the availability of annotated data and evolution of GPUs, CNNs have been increasingly applicable in a number of cases. However, in case of medical datasets, huge amounts of annotated data are not readily available yet as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Thus, transfer learning has not been very useful for medical datasets as most networks have been trained well to recognize objects present in the ImageNet dataset as shown in Fig. 1 and Figs. 5–13.

A major problem faced in training the model on less data is underfitting. Moreover, the presence of skewed classes causes the model to overfit on the largest class, which in turn, decreases the corresponding F1 scores and Cohen's Kappa. Large datasets can often be over-sampled on the lower class, but oversampling on a small dataset will not be of much help against overfitting.

In this paper, we propose a deep learning based CNN method to classify images from a small and skewed dataset of 3050 training images belonging to 4 classes and 419 validation images to achieve a considerably good result. The accuracy metric used by us is Cohen's Kappa.

* Corresponding author.

E-mail addresses: sureshsatapathy@ieee.org, suresh.satapathy@kiit.ac.in (S.C. Satapathy).

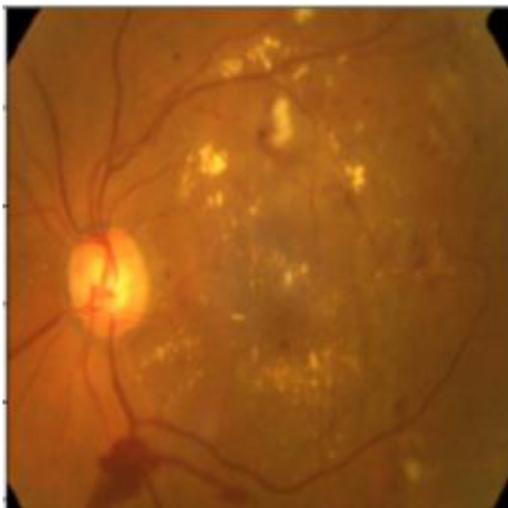


Fig. 1. An fundus image of a retina with Proliferative DR.

2. Related work

There has been considerable work regarding binary classification of Diabetic Retinopathy. Gardner et al used neural networks on 200 images by splitting an image into patches to achieve a sensitivity of 88.4% and specificity of 83.5% for binary classification. His work had been aided by a clinician in classifying the patches prior to using SVM.

Dr.Nayak et al also used neural networks to classify Diabetic Retinopathy Based on 3 classes by recognizing blood vessels and hard exudes from 140 images [3]. His model worked on classifying between normal, non-Proliferative and Proliferative retinopathy. His results were validated by clinical practitioners and showcased an accuracy of 93%, specificity of 100% and sensitivity of 90%.

Most research in the automated detection of Diabetic Retinopathy exceeding 3 classes have been done using SVMs. Acharya et al had worked on the detection of retinopathy using 5 classes [4]. Features extracted from images were used as input into the SVM classifier to capture the contours and variation of shapes. It resulted in an accuracy and specificity of 82% and 88% respectively. The areas of features such as hard exudates, hemorrhages, blood vessels and micro-aneurysms were calculated from the images and further used as input to the SVM classifier. Sensitivity and specificity were observed to be 82% and 86% respectively with an accuracy of 85.9% using this method.

Adarsh et al used image processing for automated Detection of Diabetic Retinopathy by detecting the features associated with retinopathy [5]. Texture features and the area of lesions were used for the construction of the feature vector for the SVM. Accuracies of 94.6% and 96% were obtained on the public image databases of DIARETDB1 and DIARETDB0 respectively. DIARETDB1 had 130 images and DIARETDB0 had 89.

Harry Pratt et al had proposed an approach using CNN to diagnose Diabetic Retinopathy from digital fundus images. 80,000 training images were used along with 5000 validation images belonging to 5 classes to achieve an accuracy and sensitivity of 75% and 95% respectively [6]. The training was hardware intensive with the requirement of NVIDIA K40c. Normalization was used in pre-processing the images prior to feeding them to their customized network of stacked convolution layers followed by fully connected layers.

Xiaogang Li et al had proposed a CNN based transfer learning approach on AlexNet, VGG16, VGG19. Experiments were performed on 1200 and 1014 fundus images from the MESSIDOR and DR1

datasets. Pre-trained models on the ImageNet were fine tuned on the DR datasets. Feature extraction was also experimented upon in this paper for transfer learning.

Carson Lam et al worked in classifying 4-ary data on a large dataset of 35,000 training images [9]. Their pre-processing involved systematic cropping using the Otsu's method followed by normalization. Contrast limited adaptive histogram equalization was further used for contrast adjustment. The model was trained on the 22 layered GoogleNet by transfer learning after removing the last dense layer. A Tesla K80 GPU hardware was used to aid training. Peak test accuracies of 74.5%, 68.8% and 57.2% were obtained on 2-ary, 3-ary and 4-ary classes respectively.

Maithra Raghu et al had explored transfer learning for the purpose of medical imaging [23]. Their dataset consisted of fundus photographs to diagnose a number of eye diseases including Diabetic Retinopathy graded into 5 classes. They paved an insight into the effects of transfer learning from an unrelated dataset to medical data. Evidence of feature reuse at the lowest layers and over-parametrization of models leading to the deviations from transfer learning had been observed by them.

3. Method and structure

We decided upon our network after studying baseline literature [7,22] and testing the performance of other models [3–5,9,23]. It was observed that deeper layers cause overfitting as our dataset was comparatively smaller. In our network, we used CNN-based transfer learning on the DenseNet model pre-trained on ImageNet.

3.1. Dataset and hardware

The training and testing fundus images were obtained from Kaggle (<https://www.kaggle.com>). 3050 training images were used belonging to 4 classes. The number of images were 1805, 370, 999 and 295 for 1-ary (A), 2-ary (B), 3-ary (C) and 4-ary (D) classes respectively (Fig. 2).

We trained our model on the publicly available Google Colaboratory which is a free Jupyter notebook environment that runs on the cloud [10]. Keras was used as the deep learning package with Tensorflow at the backend.

3.2. Pre-processing

Different pre-processing methods were experimented with, to determine the one which outperforms the others on the task, when the images are fed through our model. In medical image analysis, it is essential to enhance contrast, meanwhile preserving the brightness, for effective classification. Further, different images have different lighting conditions, which needs to be addressed. Contrast enhancement was used to adjust the bright or dark pixels of an image in order to extract the hidden features. The contrast between the retinal background and the blood vessels observed in the fundus images are very low. Thereby, analysis and study of the tiny retinal vasculature and othersuch abnormalities are difficult. Due to this problem, enhancing the retinal area in the fundus photography images are important in order to provide better visualization of hard exudates, blood vessels and in turn the accuracy of detecting the abnormalities increases.

Fundus photography images are reddish in color consisting of a black background. Necessary details on fundus are not present in the background and can be removed to decrease noise. Equalization of the fundus images with a black background results the darkness to increase within the details of the image [20]. Considering this issue, we decided in eliminating the black background with pre-processing. The pixel values in the black background is 0, with the realistic brighter regions having other non 0 values. The

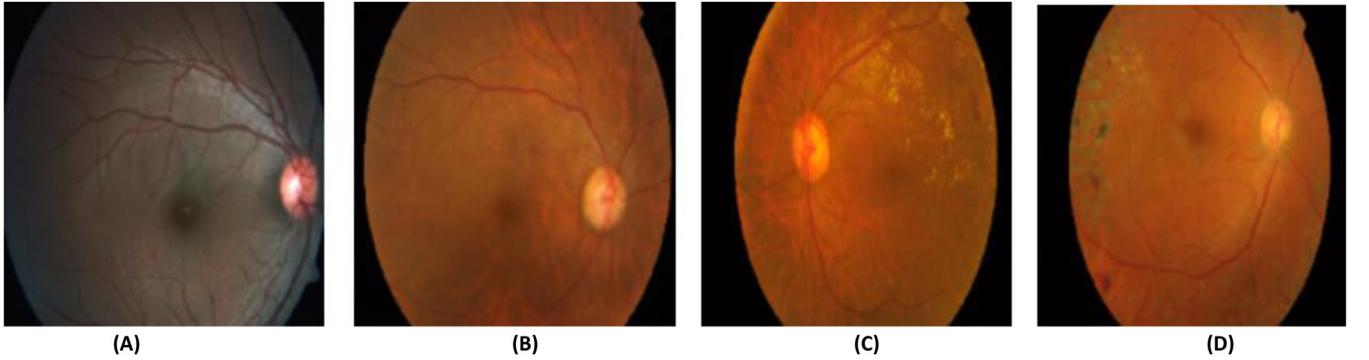


Fig. 2. Data used of 4 classes.

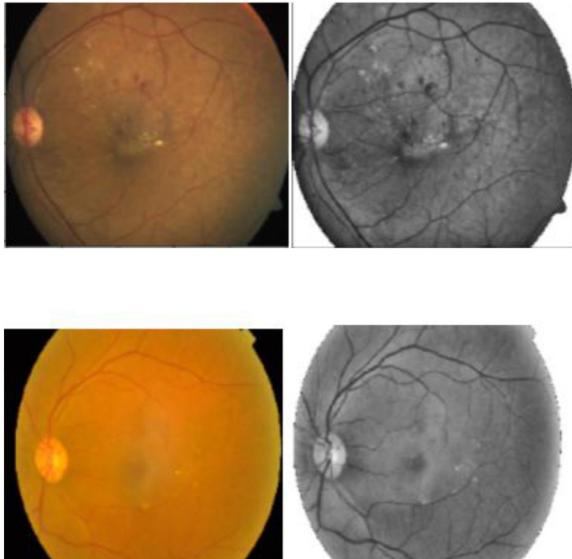


Fig. 3. Depiction of our attempted pre-processing (left-original, right-pre-processed).

black background was converted to 1 and other non black regions were converted to 0 by fixing pixel threshold. After thresholding was performed, the pixel regions containing 0 are substituted with gray scale of the input image, while the grayscale pixels are in turn substituted with the input fundus image. The green plane was extracted thereafter. Contrast Limited Adaptive Histogram Equalization (CLAHE) was then applied to enhance the small regions of relevance in the image. Clipping limit was set to 2.0, and tile grid size was taken as 8. This method, inspired from [21], though produced visually pleasing images and succeeding in bringing out certain subtleties, did not perform exceptionally well on the task (Fig 3).

To further enhance performance, we experimented with a pre-processing method inspired from Ben Graham, used in the Kaggle competition [14], and applied weighted Gaussian blur to the images. It is a 2D convolution operation which reduces noise. The values of sigmaX and sigmaY were taken to 10, as it produced comparatively better results and the kernel size was computed from the sigma values itself. This pre-processing method outperformed our previous experiments, and therefore was used in our final model (Fig 4).

$$\text{In 2 dimensions: } G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

3.3. Model

Stacking of convolution layers were a necessity for the classification of images. The first layers work on classifying the major dis-

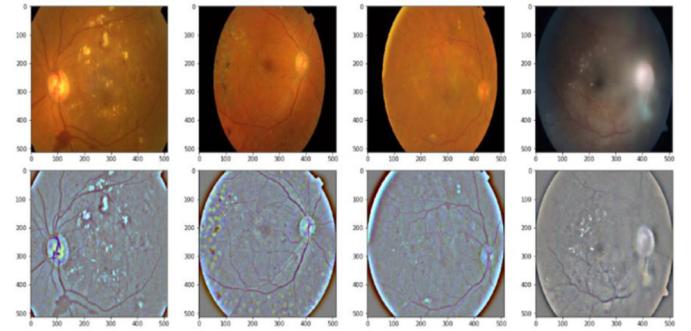


Fig. 4. Depiction of the finalized pre-processing results.

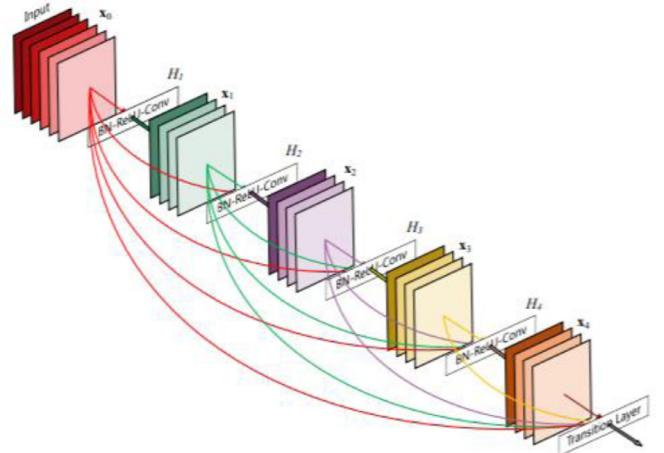
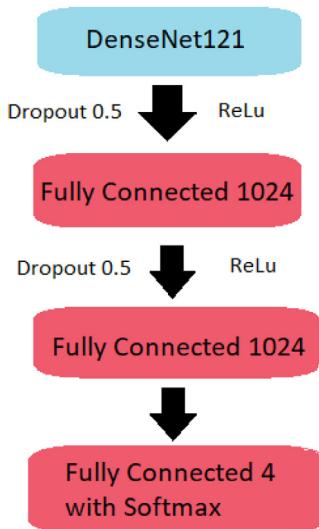
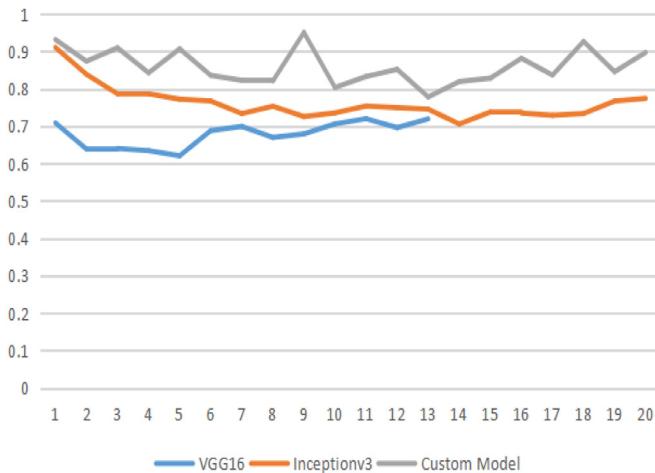


Fig. 5. An illustration of a 5-layer Dense block from Gao Huang et al [11].

tinctions and curves while the last layers are responsible for identifying the distinctive features of classification such as hard exudates and haemorrhage. Custom networks did not prove to be working well in case of a small dataset. We fine tuned and tested our model using several architectures such as Inceptionv1, Inceptionv2, Inceptionv3, Xception, VGG16, ResNet-50, DenseNet and AlexNet [16]. DenseNet121 was chosen as the final network as it was the most effective in classification. This network is the simplest among the other DenseNet architectures.

DenseNet works well in evading the vanishing gradient problem and enabling feature reuse and has achieved state of the art (SOTA) results on the ImageNet, CIFAR and SVHN datasets [11,17]. It is composed of dense blocks where every layer receives concatenated data from all the previous layers, thereby simplifying the residual network pattern. Moreover, it requires fewer parameters than the other convolutional networks. In contrast to ResNets, which also

**Fig. 6.** Model Architecture.**Fig. 7.** Monitoring progress between VGG16, Inceptionv3 and a custom built model. X axis denotes epoch and Y axis denotes validation loss.

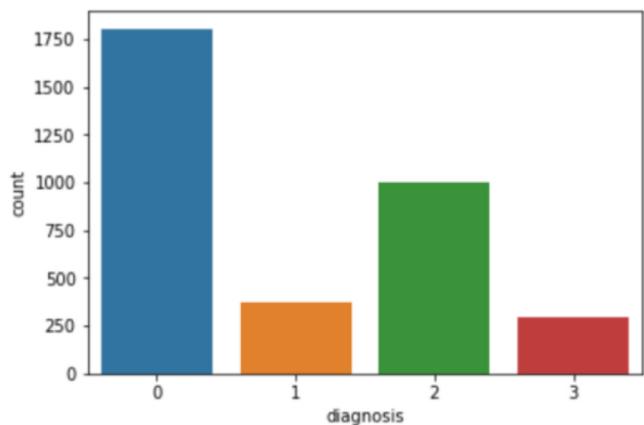
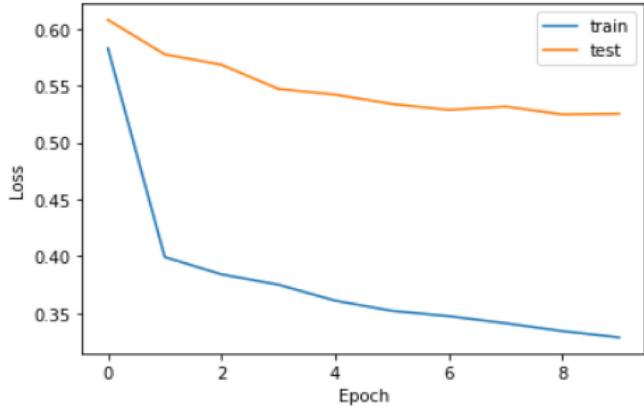
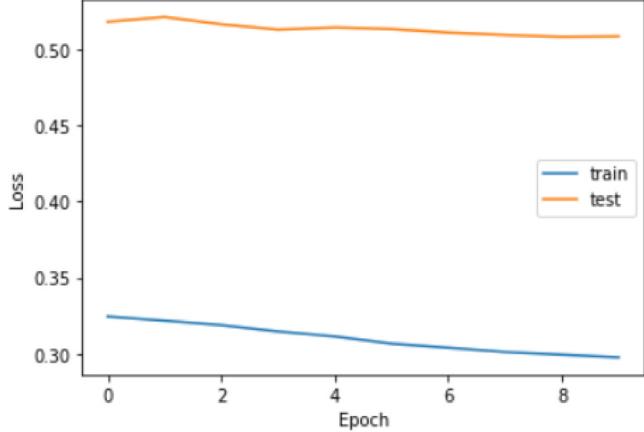
use skip connection, DenseNets concatenate the output and the input feature maps. Due to input concatenation, ease of access to learned maps for the corresponding layers are increased. Fine tuning on the pre-trained DenseNet helped in preserving the salient image features, along with learning the DR classification features.

The fully connected layers were chopped off and the output was flattened to one dimension. 2 fully connected layers of size 1024 were added. ReLu activation ($y=\max(0,x)$) was used in the fully connected layers. To prevent overfitting, a dropout of 0.5 was used. The last layer used Softmax activation function for classification between 4 classes of Diabetic Retinopathy.

$$\text{Softmax Activation : } F(x_i) = \frac{e^{x_i}}{\sum_{j=0}^k e^{x_j}}$$

($i=0,1,2,\dots,k$)

Gradient Descent was used as the optimizer. The loss function was computed using the popularly used categorical cross-entropy. The learning rate was varied with training, with the usage of Nesterov momentum.

**Fig. 8.** Data distribution across the 4 classes depicting skewness.**Fig. 9.** Epoch till 10.**Fig. 10.** Epoch till 20.

3.4. Training

Custom models deep enough were prone to overfitting, which performed well on the training set but suffered on the validation set, while shallow networks cause an underfit. Training time for the Inception networks were considerably high which restricted the number of training epoch. The usage of Xception took less time than the other counterparts. Experimentation among various networks by us has been given below.

As per the observation above, most custom models did not learn enough to find the right trade-off between high variance and bias. Training the model on VGG16 or Inceptionv3 did not prove to

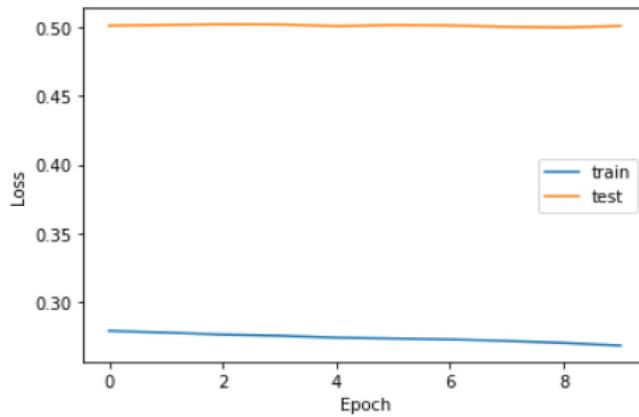


Fig. 11. Epoch till 30.

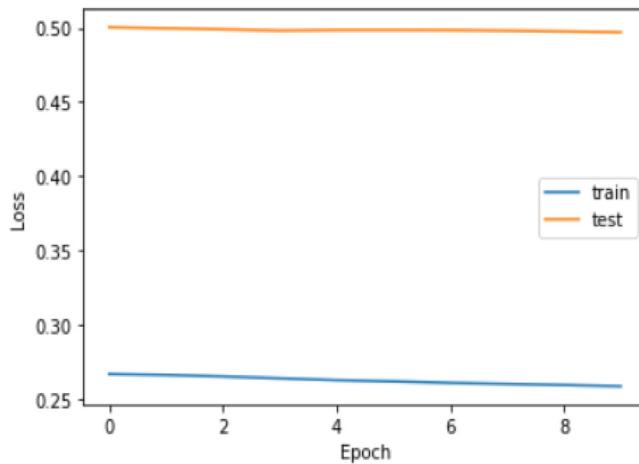


Fig. 12. Epoch till 40.

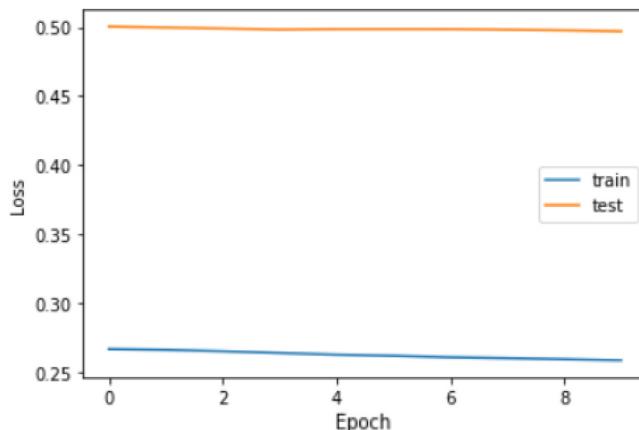


Fig. 13. Epoch till 50.

be of much use either as either of the models were unable to learn useful patterns from the limited amount of data. In Inceptionv3, a slight learning curve was observed, which plateaued for the validation loss while the training loss kept decreasing, indicating a clear situation of high variance. Several custom networks were also experimented by us, inspired from literature on other biomedical images [22]. However, neither were able to learn useful patterns from fewer data representations.

Our final model resorted to using DenseNet121 pre-trained on ImageNet as the base layer. The fully connected layers of the

DenseNet121 were removed and 2 fully connected layers of 1024 were introduced along with a dropout of 0.5. The model was trained for a total of 50 epoch on the pre-processed RGB images, each of dimension size 360×360 . Due to lack of continued training time, the model was saved at regular intervals of 10 epoch and trained. The last 2 layers of the DenseNet121 and the 2 fully connected layers were trained to fine-tune the model in learning the salient features owing to classification between the 4 classes. The implementation of transfer learning and freezing the other layers of the DenseNet helped in training by reducing the training time for the entire network along with preserving the image recognition features learned by the network from ImageNet.

Data skewness was a serious problem, and the small number of training images added to the problem. Class weights were used in ratio of the images present to counteract the problem. Their introduction was observed to boost the corresponding F1 scores of the classes and Cohen's Kappa. Oversampling did not prove to be beneficial in boosting the accuracy of the classes.

The model was trained using gradient descent. Adam and RMSprop were experimented upon but they did not turn out to be instrumental in better classification between the classes or decrease the validation loss. A learning rate of 0.003 was used for the first 30 epoch with a decay of 0.02. Nesterov momentum was used with a momentum of 0.01 in order to speed up training time by the accumulation of past gradients and move over local minima [12]. The usage of Nesterov momentum helped in going back in the case of a gradient offshoot due to the momentum. After 30 epoch, the learning rate was set to 0.001 and the decay was kept constant at 0.02. Training beyond 50 epoch did not cause the model to learn enough and in turn, resulted in overfitting on the training data. EarlyStopping was used for this purpose in order to monitor the validation loss and accuracy.

The plots above represent the losses over the training and test sets for 10 epoch each, the weights being saved and trained over regular intervals.

4. Results

The model was validated on 419 fundus images. The validation process was fast. We obtained a validation accuracy of 84.10 %. Accuracy was not used by us as the final metric due to the skewness of data. We chose Cohen's Kappa as our final metric than F1 score as Kappa provides a relativistic accuracy with respect to the other nearby classes, thus imparting a sense a reliability and originality in identification in case of a medical diagnosis [19].

$$\text{Cohen's Kappa } (\kappa) = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

A literature study on past work highlighted low performance on the middle classes of fundus images. Our model provides F1 scores of 0.64 and 0.74 on 1-ary (Mild DR) and 2-ary (Moderate DR) classes respectively. The Kappa score obtained by us is 0.8836 on the validation data and 0.9809 on the training data.

5. Discussion and conclusion

Our model has approached the classification of 4-class problem in Diabetic Retinopathy on a small dataset using deep learning. Most earlier algorithms dedicated to the classification of DR fundus images on a small dataset evaded the use of deep learning. Our method has produced comparable results with previous literature given data and hardware constraints and the presence of skewed classes. Transfer learning and fine-tuning on the pre-trained DenseNet has proved to be extremely effective on this dataset and the training technique experimented by us paid off well in terms of achieving considerably good classification results.

Our model has no issues in detecting a healthy eye from a fundus photography. The F1 score of our model on a healthy eye is 0.97. The trained network can be used on an user interface and made available at public hospitals for initial screening, with each retinal photograph taking around 0.99 seconds to be graded with minimal hardware requirements, making it robust.

In future, we will be working on enhancing this work by grading the classes based on a semantic segmentation output highlighting the retinal arteries and vessels, which might provide detailed insights to the features that truly enable algorithms to recognize patterns in a fundus photograph responsible for retinopathy. Promising results are expected from the work.

Declaration of Competing Interest

None.

References

- [1] Yann Lecun, Yoshua Bengio, Geoffrey Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [2] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. pages 1–9, 2014.
- [3] J. Nayak, P.S. Bhat, R. Acharya, C. Lim, M. Kagath, Automated identification of diabetic retinopathy stages using digital fundus images, *J. Med. Syst.* 32 (2) (2008) 107–115.
- [4] R. Acharya, C.K. Chua, E. Ng, W. Yu, C. Chee, Application of higher order spectra for the identification of diabetes retinopathy stages, *J. Med. Syst.* 32 (6) (2008) 481–488.
- [5] P. Adarsh, D. Jeyakumari, Multiclass svm-based automated diagnosis of diabetic retinopathy, in: Communications and Signal Processing (ICCP), 2013 International Conference on, IEEE, 2013, pp. 206–210.
- [6] Harry Pratt, Frans Coenen, Deborah M. Broadbent, Simon P. Harding, Yalin Zheng, Convolutional neural networks for diabetic retinopathy, 2016.
- [7] Xiaogang Li, Tiantian Pang, Biao Xiong, Weixiang Liu, Ping Liang, Tianfu Wang, Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification, 2017.
- [8] O. Faust, U.R. Acharya, E.Y. Ng, K.H. Ng, J.S. Suri, Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review, *J. Med. Syst.* 36 (1) (2012) 145–157.
- [9] Carson Lam, Darvin Yi, Margaret Guo, Tony Lindsay, Automated detection of diabetic retinopathy using deep learning, 2018.
- [10] colab.research.google.com/notebooks/welcome.ipynb#.
- [11] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, 2016.
- [12] Ilya Sutskever, James Martens, George Dahl, Geoffrey Hinton, On the importance of initialization and momentum in deep learning, 2013.
- [13] R. Gargya and T. Leng, Automated identification of diabetic retinopathy using deep learning, Elsevier, 2017.
- [14] B. Graham, Kaggle diabetic retinopathy detection competition report, 2015.
- [15] V. Gulshan, L. Peng, M.C. Stompe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *JAMA* 316 (22) (2016) 2402–2410.
- [16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
- [17] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet classification with deep convolutional neural networks, 2012.
- [18] M.D. Abramoff, J.M. Reinhardt, S.R. Russell, J.C. Folk, V.B. Mahajan, M. Niemeijer, G. Quellec, Automated early detection of diabetic retinopathy, *Ophthalmology* 117 (6) (2010) 1147–1154.
- [19] Anthony J. Viera, Joanne M. Garrett, Understanding interobserver agreement: the kappa statistic, 2005.
- [20] K.G. Suma, V. Saravana Kumar, Chapter 5 A Quantitative Analysis of Histogram Equalization-Based Methods on Fundus Images for Diabetic Retinopathy Detection, Springer Science and Business Media LLC, 2019.
- [21] K.G. Suma and V. Saravana Kumar, A quantitative analysis of histogram equalization-based methods on fundus images for diabetic retinopathy detection, 2019.
- [22] Jongwon Chang; Jisang Yu; Taehwa Han; Hyuk-jae Chang; Eunjeong Park, A method for classifying medical images using transfer learning: a pilot study on histopathology of breast cancer, 2017.
- [23] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, Samy Bengio, Transfusion: understanding transfer learning for medical imaging, 2019.



A lightweight CNN for Diabetic Retinopathy classification from fundus images

Gayathri S.*, Varun P. Gopi, P. Palanisamy

National Institute of Technology Tiruchirapalli, Tamilnadu, India



ARTICLE INFO

Keywords:

DR binary and multi class classification
Retinal fundus images
CNN feature extraction
Classifiers
10-fold cross-validation

ABSTRACT

Diabetic Retinopathy (DR) is a complication of diabetes mellitus that damages blood vessel networks in the retina. This is a serious vision-threatening issue in most diabetic subjects. The DR diagnosis by color fundus images involves skilled clinicians to recognize the presence of lesions in the image that can be used to detect the disease properly, making it a time-consuming process. Effective automated detection of DR is a challenging task. The feature extraction plays an excellent role in effective automated disease detection. Convolutional Neural Networks (CNN) have superior image classification efficiency in the present scenario compared to earlier handcrafted feature-based image classification techniques. This work presents a novel CNN model to extract features from retinal fundus images for better classification performance. The CNN output features are used as input for different machine learning classifiers in the suggested system. The model is evaluated through various classifiers (Support Vector Machine, AdaBoost, Naive Bayes, Random Forest, and J48) by using images from generic IDRiD, MESSIDOR, and KAGGLE datasets. The efficacy of the classifier is evaluated by comparing the specificity, precision, recall, False Positive Rate (FPR), Kappa-score, and accuracy values for each classifier. The evaluation results indicate that the proposed feature extraction technique along with the J48 classifier outperforms all the other classifiers for MESSIDOR, IDRiD, and KAGGLE datasets with an average accuracy of 99.89% for binary classification and 99.59% for multiclass classification. Furthermore, for the J48 classifier, the average Kappa-score (K-score) is 0.994 for binary classification and 0.994 for multi-class classification.

1. Introduction

Diabetic Retinopathy (DR) occurs predominantly in people with diabetes Mellitus history. The high blood glucose level triggers the leakage of blood and other fluids from the blood vessels in the retina. It is a cause for loss of vision in diabetic subjects. The main stages involved in DR are Non-Proliferative DR (NPDR) and Proliferative DR (PDR). The major lesions that we are considering for grading [1] are MicroAneurysms (MA), Blood vessels, Haemorrhage, and Exudates. The DR stage with any of these lesions is considered as NPDR and the advanced phase with neovascularization is called PDR. The retina with the presence of these lesions is depicted in Fig. 1. MA is a small swelling form in the wall of tiny blood vessels. In patients with DR, these minute swelling MAs are considered as the earliest visible symptom of DR. They appear in the retina as small red dots [2]. As the disease progress, its size increases. Retinal haemorrhage is another disorder in the retina which is caused by DR. The other reasons for haemorrhage (HM) may be hypertension and retinal vein occlusion. If they are very small, then it resembles MAs. When there are lipid and protein residues in the

leaked blood from the damaged capillaries, it forms yellow flicks in the retina called exudates. Normally the DR is graded into different categories: no DR, mild NPDR, moderate NPDR, severe NPDR, and PDR. Severe NDPR is diagnosed based on the “4-2-1” rule. In this, Intra Retinal Microvascular Abnormalities (IRMA) is one of the important factors. IRMA is the abnormal branching or expansion of retinal blood vessels. The rules in grading DR are mentioned in Table 1. DR in its serious phase is hard to heal. It will result in complete vision loss. Reducing its prevalence worldwide is very crucial. Several approaches have been implemented to detect the symptoms of DR. But the challenging task is the methods of feature extraction. Making the feature extraction for DR classification as precise as possible and also reducing computational costs and time is very important. By considering all these facts, the proposed work is trying to present a simple CNN for feature extraction and combining it with Machine Learning (ML) Classifiers for better classification.

* Corresponding author.

E-mail address: gsgayathriunnithan@gmail.com (Gayathri S.).

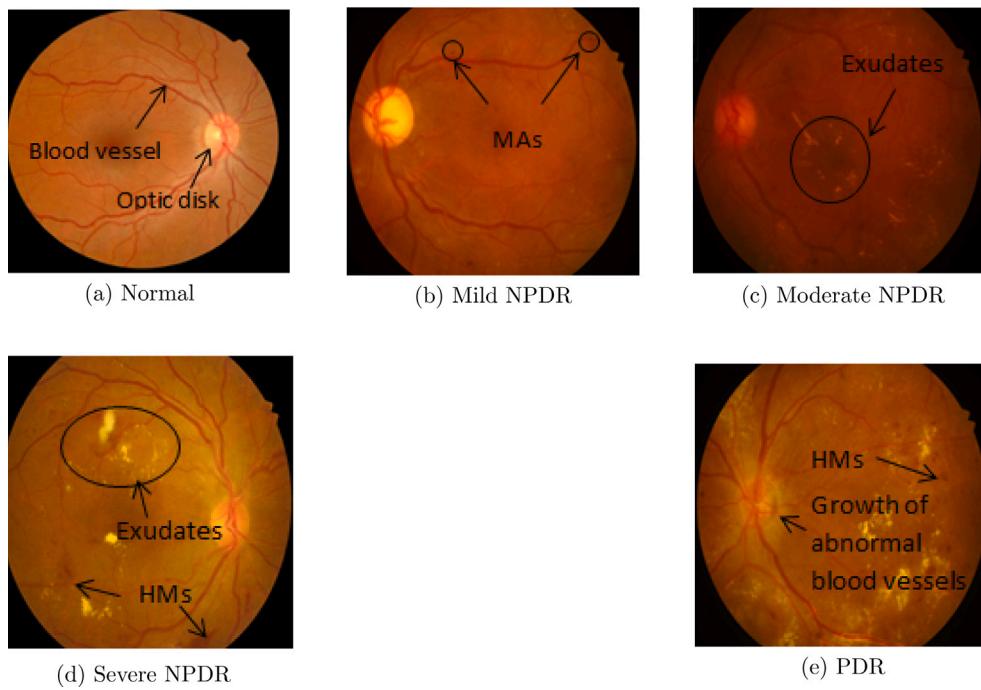


Fig. 1. Different stages of DR in fundus images.

Table 1
Grading of diabetic retinopathy.

Lesion detection	Grading level
Not observed any abnormalities	No DR
Only observed MAs	Mild NPDR
A small number of MAs with or without cotton-wool spots, venous beading, or presence of IRMA which is less than the 4-2-1 guidelines ≥ mild but ≤ severe	Moderate NPDR
Presence of any of the 4-2-1 guidelines • All the four quadrants contains dot blot hemorrhage • venous beading in atleast two quadrants • Prominent IRMA in atleast one quadrant	Severe NPDR
Presence of either Preretinal hemorrhage or neovascularization	PDR

Automated grading has two main issues. The first one is achieving a desirable precision offset and specificity. In this work we are considering five class grading, again it is significantly harder to achieve the same. Second is overfitting in the neural network. These two are particularly related to the training algorithm in the system. Convolutional Neural networks (CNN) are state-of-the-art deep learning techniques that have led to many breakthroughs in the identification and detection of objects in many areas including medical imaging. The task here is to solve the issues regarding the feature extraction stage with a well-defined CNN architecture. The availability of a large database is a crucial stage in deep learning algorithms. To manage this issue, machine learning classifiers can be utilized for classification and CNN for extracting minute features from fundus images for accurate disease classification. Thus handcrafted feature extraction methods can be replaced with efficient and accurate automated methods.

2. Related works

The automatic severity level measurement of DR using Artificial Neural Network (ANN) is explained in [3]. Lesions such as MAs, Blood vessels, hemorrhage, and exudates are derived from the retinal fundus image. The lesions then fed into the multilayer feed-forward

neural network to classify DR into mild, moderate, and severe. The method explained in [4] deals with the two-stage CNN for detecting the abnormal lesions in the fundus image. Here, the lesion type and its location in the image was pointed out and graded the DR. In [5], features like area, perimeter, and count from the four lesions extracted and the classification is performed using ANN and graded DR into mild, moderate or severe. The automatic grading in [6] is obtained from the results of a validated red lesion detection method. Assessment is performed on a public database by leave one out validation method. They tried to prove the feasibility of automatic DR screening. Diabetic or non-diabetic classification of the retinal fundus is carried out in [7]. The image was divided into four sub-images. After extracting the features, they applied Haar wavelet transformation. To select a better feature, they used Principal Component Analysis (PCA). Then backpropagation neural network and one rule classifier were used for the classification process. The work in [8] is based on the screening of fundus images with different illumination and fields of view. They graded the DR using different classifiers such as k-Nearest Neighbor (k-NN), AdaBoost, Support Vector Machines (SVM), and Gaussian mixture model (GMM). According to their performance analysis, GMM and k-NN performed well than others. The main task was to implement the feature ranking method to reduce the number of features for lesion classification. A combined feature extraction method using ADTCWT (Anisotropic Dual-Tree Complex Wavelet Transform) and Haralick is introduced in [9]. Then the selected features are fed into the classifier for DR grading. A three-stage (preprocessing, image analysis, and classification) model for grading was proposed in [10]. They used a hybrid classifier (a combination of GMM and m-medoids classifier) to improve the classification efficiency. A genetic algorithm was used to learn the weights of the classifier. The results showed that the system detected all the NPDR lesions and graded the severity at some accuracy.

In [11], A CNN method is proposed for the diagnosis and accurate severity classification of DR from fundus images. It obtained an accuracy of 75% on 5000 validation images. In [12], they extracted hard exudates area, blood vessels area, texture, entropies, and bifurcation points. The work used a combination of texture and morphological changes for classification. The parameter(σ) of the Probabilistic Neural Network (PNN) classifier is tuned using genetic algorithm and particle

swarm optimization. The five class grading of DR was also one of the highlights of this work. Authors of [13] explained how to handle the blurred retinal images for the detection of DR. To enhance the system performance, they used a regularized filter deblurring algorithm. The areas of the blood vessel, MAs, exudates are computed and fed into the ANN classifier. In [14], modified Alexnet architecture is employed to categorize the input fundus images. A high level of accuracy CNN with the application of suitable pooling was proposed to classify DR fundus images into the severity of the disease. In [15], a siamese neural network architecture named binocular neural network is proposed for DR classification. They used the Inception-v3 CNN model in the binocular structure. Local features of retinal images are extracted using Local Binary Pattern (LBP) in [16]. Then it is evaluated across ANN, Random Forest, and SVM for the detection task. The sparse coding technique with linear SVM for retinal image classification is proposed in [17]. They used the BoVW technique for feature extraction. According to their evaluation, a dictionary size of 100 achieves better sensitivity and specificity. A CNN using fractional max pooling is developed in [18] for DR classification. They used the network for feature extraction and SVM with Teaching Learning Based Optimization (TLBO) for binary as well as multiclass classification. In [19], Inception-v3 architecture is used for DR grading and their studies included the different size of the input to the network that can be able to improve the classification ability. A modified inception-v3 network (named as inception@4) for DR grading is demonstrated in [20] and the system is evaluated using their database which achieves an accuracy of 88.72%. In [21], a multichannel CNN is proposed for DR classification. The network is evaluated across EYEPACS images and obtained an accuracy of 97.08%.

From the review of literature, it can be seen that the multi-class classification using deep neural networks use heavy and large networks (pre-trained networks and pre-trained networks with transfer learning), which may not be suitable for real-time deployment. Thus, it is necessary to develop methods that can boost accuracy while keeping the network size as small as possible.

3. Materials and methods

The proposed work presents a CNN architecture to extract the features from retinal fundus images for binary and multi-class classification of DR. It is necessary to develop automated methods to improve the accuracy of diagnosis and classify the subjects into various stages of DR. The proposed work presents a method that can minimize the computational complexity and provide better classification performance. Through the layer by layer stacking of convolution, pooling, non-linear activation function mapping, CNN extracts symbolic information from the input data and makes the layer by layer abstraction possible. This process is termed as feed-forward operation. By calculating the error between the predicted value and true value, the detected error is fed back from the last layer by the backpropagation algorithm. The extracted features are then fed into the classifier for binary as well as multi-class classification as shown in Fig. 2. The proposed method is implemented in the Keras framework in python using Tensorflow.

3.1. Feature extraction

The feature extraction is one of the predominant stages in an automated classification system, as perfect and minute features are required for precise prediction. CNN has the ability to extract minute features from the image that are sufficient for good classification. To get the features from the CNN model, it is required to train the CNN network up to the last dense layer with respect to the target variable. To improve the latency of the automated system, it is better to use CNN for feature extraction, followed by a simple classifier.

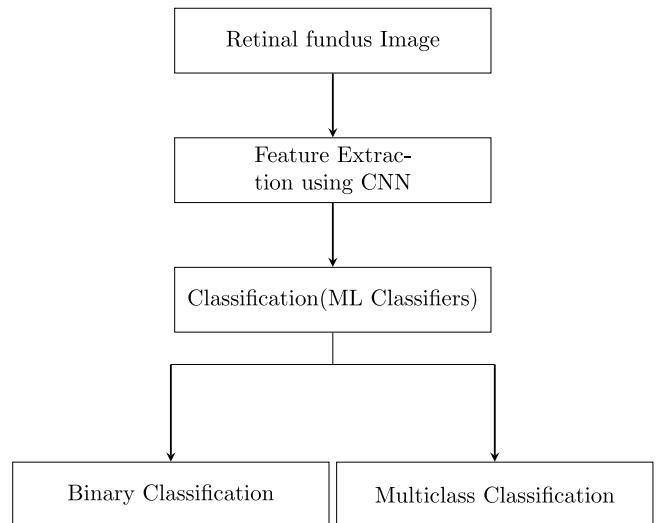


Fig. 2. Proposed work.

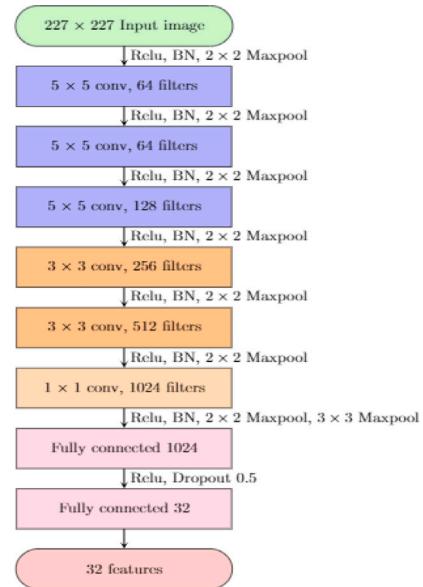


Fig. 3. Proposed CNN Architecture.

3.1.1. Proposed network architecture

The CNN architecture for feature extraction is demonstrated in Fig. 3 and the learnable parameters for each layer are listed in Table 2. In the designed network six convolutional layers and two fully connected layers are used. From the last fully connected layer 32 relevant features are obtained and used for classification.

It is essential to resize the image appropriately before feeding them into CNN. In the proposed work, the input fundus image is resized to a dimension of $227 \times 227 \times 3$ pixels corresponding to the breadth, height, and the three color channels representing the depth of the input fundus image. The function of the CNN is to reduce the input images to a form that is simpler to process without losing critical features to obtain a good prediction. The function of each layer in the CNN [22] is narrated below.

- **Convolutional layer** calculates the output of neurons as a dot product of a small portion of the image with their corresponding weights. Along the length and breadth, this process is repeated. These layers use the parameter sharing system to regulate the number of parameters.

Table 2

Tensor sizes & number of parameters in the proposed architecture.

Layer name	Tensor size	No. of parameters
Input image	227 × 227 × 3	0
Conv-1	227 × 227 × 64	4864
Max pool-1	114 × 114 × 64	0
Conv-2	114 × 114 × 64	1,02,464
Max pool-2	57 × 57 × 64	0
Conv-3	57 × 57 × 128	2,04,928
Max pool-3	29 × 29 × 128	0
Conv-4	29 × 29 × 256	2,95,168
Max pool-4	15 × 15 × 256	0
Conv-5	15 × 15 × 512	11,80,160
Max pool-5	8 × 8 × 512	0
Conv-6	8 × 8 × 1024	5,25,312
Max pool-6	4 × 4 × 1024	0
Maxpool-7	2 × 2 × 1024	0
FC-1	1024 × 1	41,95,328
FC-2	32 × 1	32,800

Total learnable parameters: 65,41,024.

- In **Rectified Linear Unit (ReLU)** layer, simplest non-linear activation function is employed here. This layer replaces all the negative activations with 0 by introducing non-linearity to the system and by applying the function $f(k) = \max(0, k)$. where, k is the neuron input.
- **Pooling layer** reduces the number of inputs to the next layer of feature extraction, thus allowing us to have many more different feature maps. Max pooling is a method of discretization based on samples. The goal is to down-sample an input representation that reduces its dimensionality and enables assumptions about features contained in binned sub-regions to be made. Max pooling is performed by applying a max filter to (generally) the initial representation's non-overlapping subregions.
- **Fully connected layer** neurons, as seen in regular Neural Networks, have complete links to all prior layer activations. Therefore, their activations can be calculated with a matrix multiplication accompanied by a bias offset.

The parameters are trained using a backpropagation algorithm taking cross-entropy as the loss function. In addition to this, Batch normalization [23,24] dropout strategy, L_1 , L_2 regularization are included to avoid overfitting. The total number of learnable parameters in the proposed network architecture is 65,41,024.

3.2. Classifiers

3.2.1. Support vector machine (SVM)

Support vector machines [25] are supervised learning methods with associated learning algorithms. If the vectors are non linearly separable in a space, then the SVM helps to make it linearly separable in a higher-dimensional space. In this work, the SVM is implemented using a Radial Basis Function (RBF) [26] kernel with gamma value selected as the reciprocal of the product of the total number of features and the variance. The algorithm [26] for training the SVM is explained in Algorithm 1.

3.2.2. Random forest

Random Forest [27] is an ensemble model classifier, in which a group of trees is developed together with each has an independent random vector. i.e., the K th tree generates a random vector Φ_K which is independent from previously generated random vectors ($\Phi_1, \Phi_2, \dots, \Phi_{K-1}$) but have same distribution [28]. In this work, the number of trees used in the classifier is 100 [28]. The steps in the pseudo-code generation are as follows:

1. Select the features randomly from total features.

Algorithm 1 Pseudo code for SVM

Require: S and t load with labeled data for training

consider initially $\eta = 0$

START:

1. γ assume random value initially

2. **repeat**

3. **Do for:** $\{s_i, t_i\}, \{s_j, t_j\}$

4. find η_i and η_j and optimize

5. **end for**

6. **until** η and γ become unchanged

Ensure: Retain support vectors where $\eta_i > 0$

2. From the selected features find out the mother node for the tree using best split method
3. Again use the best split method to split the others into branches
4. Repeat steps 1–3 until form a root node with target as the leaf nodes
5. construct the forest by iteration (doing steps 1–4) for n times to create n trees

3.2.3. Multi layer perceptron (MLP)

MLP [29] is a multi-layer feed-forward network that maps inputs to outputs in a nonlinear manner. The MLP base structure contains an input layer, a hidden layer, and an output layer, with each node fully connected to the nodes in the next layer with appropriate weights, which is schematically represented in Fig. 4. MLP uses a backpropagation method for training, there might be a non-linear activation function that is not seen in other neural networks. In MLP, the sigmoid function is generally used, and it is described in Eq. (1). In the proposed work only one hidden layer is used by considering the advantages of single hidden layer MLP which is mentioned in [29].

$$y_i(s_i) = (1 + e^{-s_i})^{-1} \quad (1)$$

where, y_i depicts the i th node output and the weighted sum of the input synapses is denoted by s_i . In back propagation algorithm [30], the motive is to reduce the error propagated in the network by adjusting the weights at each node. The error $e_j(n)$ at the j th output node in the n th data point can be calculated using the actual output value $a_j(n)$ and predicted output value $y_j(n)$ as in Eq. (2).

$$e_j(n) = a_j(n) - y_j(n) \quad (2)$$

To minimize the error in the entire output, the corrections in weights at each node are done by Eq. (3) and the new weight for each node can be acquired from Eq. (4).

$$\sigma(n) = \frac{1}{2} \sum_j [e_j^2(n)] \quad (3)$$

$$\Delta W_{ji}(n) = -\alpha \frac{\partial \sigma(n)}{\partial j(n)} y_i(n) \quad (4)$$

where, α is the learning rate, $y_i(n)$ is the previous node output. The iterative process continues until the error becomes unchangeable.

3.2.4. J48 classifier

This is the open-source java implementation of C4.5 [31] Decision tree. This is a decision tree algorithm mainly designed for data mining. In this, the information gain ratio is evaluated to select each node test feature. This procedure is termed as feature selection. While operating, the attribute with the largest information gain will be selected as the test feature for the present node. Let us assume that F is a set of input feature vectors given to the classifier which contains F_1, F_2, \dots, F_n instances. Suppose there are t distinct values for t distinct classes

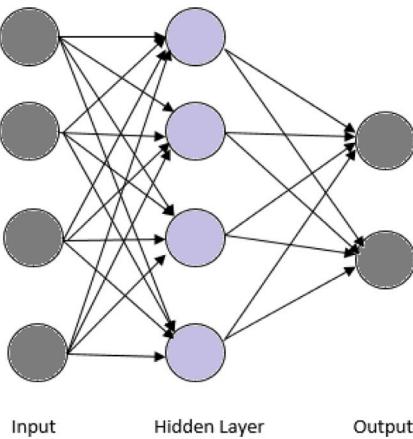


Fig. 4. Schematic of Multi Layer Perceptron.

C_i (where, $i = 1, 2, \dots, n$). Then the gain ratio G_A of sub-attribute A in each attribute can be calculated using Eq. (5).

$$G_A = G(A)/S_A(F) \quad (5)$$

Where, $G(A)$ is the information gain of attribute A which can be obtained by taking the difference between total information $I(D)$ and the attribute information $I_A(D)$ as shown in Eq. (6).

$$G(A) = I(D) - I_A(D) \quad (6)$$

If P_i is the distinct class probability, then $I(D)$ and $I_A(D)$ can be calculated using Eqs. (7) and (8) respectively.

$$I(D) = \sum P_i \log_2(P_i) \quad (7)$$

$$I_A(D) = -\sum \frac{|F_j|}{|F|} I(F_j) \quad (8)$$

The split information value $S_A(F)$ of attribute can be formulated as:

$$S_A(F) = -\sum \frac{|F_j|}{|F|} \log\left(\frac{|F_j|}{|F|}\right) \quad (9)$$

Actually, the fraction $\frac{|F_j|}{|F|}$ acts as the j th partition weight. By utilizing all these equations the C4.5 decision tree can be developed which forms appropriate conditions that can be used for classification. Then during testing, it will classify the input feature vector according to the conditions.

3.3. Performance analysis

The performance of the classifier is evaluated based on K-fold evaluation [32]. In this assessment methodology, the entire dataset available is divided into K-sub parts during the training itself ($K=1, 2, 3, \dots$). Then each subsection is treated as a validation set for each iteration. The general steps in K-fold validation are as follows:

1. Randomly shuffle the dataset
2. Data will split into K-sub groups (If $K=10$, then split the data into 10 groups)
3. The evaluation process is performed for each group
 - Use one group as a test set
 - Use the remaining groups as the training dataset
 - Train the classifier with this dataset and evaluate the model with the test data
 - Retain the evaluation score and repeat the steps by selecting other group.

Table 3
Confusion matrix for binary classification.

		Predicted diagnosis	
		DR	NO DR
Actual diagnosis	DR	J	K
	NO DR	M	N

Table 4
Confusion matrix for severity grading.

		Predicted diagnosis			
		Normal	Mild	Moderate	Severe
Actual diagnosis	Normal	A_{11}	A_{12}	A_{13}	A_{14}
	Mild	A_{21}	A_{22}	A_{23}	A_{24}
	Moderate	A_{31}	A_{32}	A_{33}	A_{34}
	Severe	A_{41}	A_{42}	A_{43}	A_{44}

4. Summarize the model efficiency using the evaluation scores.

The findings are stored in the form of a confusion matrix [33]. The confusion matrix structure that depicts a binary classifier's characteristics is shown in Table 3. In that matrix, J , K , M , and N represents the number of True Positives (TP), False Negatives (FN), False Positives (FP), and True Negatives (TN) respectively. TP and TN give the results of correctly classified data while FP and FN give the incorrectly classified details. Similarly the confusion matrix for multi-class classification is demonstrated in Table 4. The number of TPs, FPs, FNs, TNs in multiclass classification can be easily acquired through formulas in Eq. (10) for each actual class i by taking p predicted classes. Using these values, we can calculate the accuracy, F-score, specificity, precision, and recall of the classifier to examine system efficiency.

$$\text{Number of TPs}, J_i = A_{ii} \quad (10)$$

$$\text{Number of FNs}, K_i = \sum_{j=1}^p A_{ij} - J_i$$

$$\text{Number of FPs}, M_i = \sum_{j=1}^p A_{ji} - J_i$$

$$\text{Number of TNs}, N_i = \sum_{j=1}^p \sum_{k=1}^n A_{ik} - J_i - M_i - K_i$$

Accuracy defines the overall power of the system. For binary class and multi class, it can be obtained from the confusion matrix using the formula in Eqs. (11) and (12) respectively. For binary classification,

$$\text{Accuracy} = \frac{J + N}{J + K + M + N} \quad (11)$$

For multi class classification, each class accuracy:

$$\text{Accuracy}_i = \frac{J_i}{J_i + K_i + M_i + N_i} \quad (12)$$

False Positive Rate(FPR) gives the rate of incorrect positive predictions. The best FPR rate for a good classifier is 0.0.

For binary classification,

$$FPR = \frac{M}{M + N} \quad (13)$$

For multi class classification, each class FPR:

$$FPR_i = \frac{M_i}{M_i + N_i} \quad (14)$$

Precision gives the positive prediction value. This value provides the information on how efficiently our system avoids FPs. It can be measured as,

For binary classification,

$$Precision = \frac{J}{J + M} \quad (15)$$

For multi class classification, each class Precision:

$$Precision_i = \frac{J_i}{J_i + M_i} \quad (16)$$

Table 5

Confusion matrix for the evaluation of each classifier using IDRiD Database for binary classification.

DR		No DR		DR		No DR	
DR	269	10		DR	279	0	
No DR	3	131		No DR	5	129	
(a) SVM				(b) Random forest			
DR		No DR		DR		No DR	
DR	278	1		DR	278	1	
No DR	4	130		No DR	0	134	
(c) MLP				(d) J48			

Table 6

Confusion matrix for the evaluation of each classifier using MESSIDOR Database for binary classification.

DR		No DR		DR		No DR	
DR	650	4		DR	653	1	
No DR	525	21		No DR	1	545	
(a) SVM				(b) Random forest			
DR		No DR		DR		No DR	
DR	647	7		DR	654	0	
No DR	6	540		No DR	1	545	
(c) MLP				(d) J48			

Table 7

Confusion matrix for the evaluation of each classifier using KAGGLE database for binary classification.

DR		No DR		DR		No DR	
DR	9221	95		DR	9315	1	
No DR	90	25720		No DR	0	25810	
(a) SVM				(b) Random forest			
DR		No DR		DR		No DR	
DR	9293	23		DR	9315	1	
No DR	49	25761		No DR	0	25810	
(c) MLP				(d) J48			

Recall, also called as sensitivity gives the information about how efficiently the model reduces FNs. This can be calculated as,

For binary classification,

$$\text{Recall} = \frac{J}{J + K} \quad (17)$$

For multi class classification, each class recall:

$$\text{Recall}_i = \frac{J_i}{J_i + K_i} \quad (18)$$

F1-score determines the model's accuracy. This score gives the harmonic mean of precision and recall.

For binary classification,

$$\text{F1-score} = \frac{2J}{2J + M + K} \quad (19)$$

For multi class classification, each class F1-score:

$$(F1\text{-score})_i = \frac{2J_i}{2J_i + M_i + K_i} \quad (20)$$

High values of these measures except False positive rate indicates the good performance of the classifier.

Specificity quantifies how efficiently the false positives(FPs) are reduced in a model. The sum of FPR and Specificity gives 1.

For binary classification,

$$\text{Specificity} = \frac{N}{M + N} \quad (21)$$

Table 8

Detailed efficiency measures of different classifiers using IDRiD Database for binary classification.

Classifier	FP rate	Specificity	Precision	Recall	F1 Score	Class
SVM	0.022	0.978	0.989	0.964	0.976	DR
	0.036	0.964	0.929	0.978	0.953	Normal
Random forest	0.037	0.963	0.982	1.00	0.991	DR
	0.00	1.00	1.00	0.963	0.981	Normal
MLP	0.030	0.97	0.986	0.996	0.991	DR
	0.004	0.996	0.992	0.970	0.981	Normal
J48	0.00	1.00	1.00	0.996	0.998	DR
	0.004	0.996	0.993	1.00	0.996	Normal

For multi class classification, each class Specificity:

$$\text{Specificity}_i = \frac{N_i}{M_i + N_i} \quad (22)$$

The overall accuracy, FPR, Precision, Recall, F1-score and specificity can be achieved in multiclass classification by taking the mean values acquired for each class using Eqs. (12), (14), (16), (18), (20) and (22) respectively.

To summarize the performance of the system, the weighted average of each class performance measures are required. If P_1 and P_2 denotes the performance measures obtained for class 1 (C_1) and class 2 (C_2) respectively then the weighted average of performance measure W_{PM} can be calculated using the following equation:

$$W_{PM} = \frac{(P_1 * |C_1|) + (P_2 * |C_2|)}{|C_1| + |C_2|} \quad (23)$$

The Kappa statistic (K-score) is a quality metric of the classifier that assesses the interrater reliability. It is a measure relating an Observed Accuracy (A_O) to an Expected Accuracy (A_E). It can be computed as:

$$K\text{-score} = \frac{(A_O - A_E)}{(1 - A_E)} \quad (24)$$

4. Experimental results and discussions

The experimental analysis of the proposed method is conducted using a PC with Nvidia GeForce, RTX2080 11 GB GPU. The entire network is trained using backpropagation by automatic differentiation algorithm using stochastic mini-batch gradient descent to update the model parameters. The best learning rate has been found experimentally to be 0.003 for a mini-batch of 64. A momentum factor of 0.9 is used to make the training less noisy and converge faster to the objective. Initially, there was heavy overfitting of the data. Using a dropout factor of 0.5 and L_2 regularization improved validation accuracy. For others, the default parameter settings are used. The results obtained and discussions on the work are described in this section.

4.1. Datasets

The performance of the proposed system is assessed individually using IDRiD, MESSIDOR, KAGGLE databases that have 413, 1200, and 35126 retinal fundus images respectively. For binary classification, 279 DR and 134 normal images are used from IDRiD, 654 DR and 546 images are used from MESSIDOR, 9316 DR and 25810 normal images from KAGGLE is used. For multi-class classification: 134 normal, 20 mild NPDR, 136 moderate NPDR, 74 severe NPDR and 49 PDR from IDRiD database; 546 normal, 153 mild DR, 247 moderate DR and 254 severe DR images from MESSIDOR database; 25810 normal, 2443 mild NPDR, 5292 moderate NPDR, 873 severe NPDR, and 708 PDR images from KAGGLE database is used.

Table 9

Detailed efficiency measures of different classifiers using MESSIDOR Database for binary classification.

Classifier	FP Rate	Specificity	Precision	Recall	F1 Score	Class
SVM	0.02	0.98	0.983	0.976	0.979	DR
	0.038	0.962	0.969	0.994	0.981	Normal
Random forest	0.002	0.998	0.998	0.998	0.998	DR
	0.002	0.998	0.998	0.998	0.998	Normal
MLP	0.011	0.989	0.991	0.989	0.990	DR
	0.011	0.989	0.987	0.989	0.988	Normal
J48	0.002	0.998	0.998	1.00	0.999	DR
	0.00	1.00	1.00	0.998	0.999	Normal

Table 10

Detailed efficiency measures of different classifiers using KAGGLE Database for binary classification.

Classifier	FP Rate	Specificity	Precision	Recall	F1 Score	Class
SVM	0.003	0.997	0.990	0.990	0.990	DR
	0.01	0.99	0.996	0.997	0.996	Normal
Random forest	0.00	1.00	0.999	0.999	0.999	DR
	0.002	0.998	0.999	1.00	0.999	Normal
MLP	0.002	0.998	0.995	0.998	0.996	DR
	0.002	0.998	0.999	0.998	0.999	Normal
J48	0.00	1.00	1.00	1.00	1.00	DR
	0.00	1.00	1.00	1.00	1.00	Normal

Table 11

Weighted average values for each performance measures for binary classification.

Database	Classifier	FP rate	Specificity	Precision	Recall	F1 Score
IDRiD [34]	SVM	0.027	0.973	0.97	0.969	0.969
	Random forest	0.025	0.975	0.988	0.988	0.988
	MLP	0.021	0.979	0.988	0.988	0.988
	J48	0.001	0.999	0.998	0.998	0.998
MESSIDOR [35]	SVM	0.024	0.976	0.980	0.979	0.979
	Random forest	0.002	0.998	0.998	0.998	0.998
	MLP	0.011	0.989	0.989	0.989	0.989
	J48	0.001	0.999	0.999	0.999	0.999
KAGGLE [36]	SVM	0.008	0.992	0.995	0.995	0.995
	Random forest	0.001	0.999	0.999	0.999	0.999
	MLP	0.002	0.998	0.998	0.998	0.998
	J48	0.00	1.00	1.00	1.00	0.999

Table 12

Kappa statistic & validation accuracy of each classifier using IDRiD Database for binary classification.

Classifier	Correctly classified instances	K-score	Accuracy (%)
SVM	400	0.929	96.85
Random forest	408	0.972	98.78
MLP	408	0.972	98.78
J48	412	0.995	99.76

Table 13

Kappa statistic & validation accuracy of each classifier using MESSIDOR Database for binary classification.

Classifier	Correctly classified instances	K-score	Accuracy (%)
SVM	1175	0.958	97.92
Random forest	1198	0.996	99.8
MLP	1187	0.978	98.92
J48	1199	0.998	99.92

4.2. Analysis of system for binary classification

In binary classification, the system predicts whether the retinal fundus image belongs to DR or normal. The four classifiers used for analysis are SVM, Random Forest, MLP, and J48. The confusion matrices obtained by using IDRiD, MESSIDOR, and KAGGLE datasets are

Table 14

Kappa statistic & validation accuracy of each classifier using KAGGLE Database for binary classification.

Classifier	Correctly classified instances	K-score	Accuracy (%)
SVM	34941	0.985	99.47
Random forest	35100	0.998	99.92
MLP	35054	0.995	99.79
J48	35125	0.99	99.99

Table 15

Confusion matrix for the evaluation of each classifier for multiclass classification using IDRiD Database.

	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	130	0	1	3	0
Mild NPDR	0	0	20	0	0
Moderate NPDR	8	0	128	0	0
Severe NPDR	10	0	0	63	1
PDR	0	0	0	35	14

(a) SVM					
	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	125	0	1	8	0
Mild NPDR	0	5	15	0	0
Moderate NPDR	0	1	135	0	0
Severe NPDR	11	0	0	61	2
PDR	7	0	0	11	31

(b) Random forest					
	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	123	0	4	7	0
Mild NPDR	0	0	20	0	0
Moderate NPDR	5	0	31	0	0
Severe NPDR	3	0	0	68	3
PDR	0	0	0	3	46

(c) MLP					
	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	133	0	0	1	0
Mild NPDR	0	19	1	0	0
Moderate NPDR	1	0	135	0	0
Severe NPDR	0	0	0	73	1
PDR	0	0	0	0	49

(d) J48					
	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	133	0	0	1	0
Mild NPDR	0	19	1	0	0
Moderate NPDR	1	0	135	0	0
Severe NPDR	0	0	0	73	1
PDR	0	0	0	0	49

tabulated in Tables 5–7 respectively. From the confusion matrices, it is able to compute the performance measures that can be used to evaluate system performance. The detailed performance measures of normal and DR images using all the datasets individually is demonstrated in Tables 8–10. From the tabulation, the FPR of the J48 classifier is comparatively lower than other classifiers and almost producing high specificity, precision, recall, and F1-score values. Table 11 summarizes all the performance measures for binary classification by taking weighted average values indicate that the performance of the J48 classifier is better than others. The accuracy and K-score of each classifier using each dataset are listed in Tables 12–14. The interrater reliability indicated by K-score reflects the degree to which the data gathered in the experiment are accurate representations of the calculated variables. The K-score ranges from 0.81 to 1.00 can be considered as almost perfect agreement [37]. The proposed model with all classifiers thus gives K-score above 0.9 for binary classification, which suggests the system's reliability. The J48 classifier generates the highest K-score of the four classifiers. The average accuracy of the model using three datasets with the J48 classifier is 99.89%.

4.3. Analysis of system for multiclass classification

For the analysis of multiclass classification, categorized data from IDRiD, MESSIDOR, and KAGGLE databases are used. The confusion matrices for multiclass classification using all the three datasets are

Table 16

Confusion matrix for the evaluation of each classifier for multiclass classification using MESSIDOR database.

	Normal	Mild	Moderate	Severe
Normal	534	12	0	0
Mild	5	140	8	0
Moderate	1	14	230	2
Severe	0	0	11	243

(a) SVM				
	Normal	Mild	Moderate	Severe
Normal	545	1	0	0
Mild	5	142	6	0
Moderate	3	2	237	5
Severe	0	0	4	250

(b) Random forest				
	Normal	Mild	Moderate	Severe
Normal	543	3	0	0
Mild	3	148	2	0
Moderate	0	2	237	8
Severe	0	0	9	245

(c) MLP				
	Normal	Mild	Moderate	Severe
Normal	545	1	0	0
Mild	0	152	1	0
Moderate	0	0	246	1
Severe	0	0	0	254

(d) J48				
	Normal	Mild	Moderate	Severe
Normal	545	1	0	0
Mild	0	152	1	0
Moderate	0	0	246	1
Severe	0	0	0	254

Table 17

Confusion matrix for the evaluation of each classifier for multiclass classification using KAGGLE database.

	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	25716	0	71	23	0
Mild NPDR	0	2365	78	0	0
Moderate NPDR	75	71	5146	0	0
Severe NPDR	421	0	0	447	5
PDR	52	0	0	634	22

(a) SVM					
	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	25810	0	0	0	0
Mild NPDR	0	2440	3	0	0
Moderate NPDR	1	0	5291	0	0
Severe NPDR	29	0	0	840	4
PDR	10	0	0	12	686

(b) Random forest					
	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	25772	0	38	0	0
Mild NPDR	0	2380	63	0	0
Moderate NPDR	47	51	5194	0	0
Severe NPDR	873	0	0	0	0
PDR	708	0	0	0	0

(c) MLP					
	Normal	Mild NPDR	Moderate NPDR	Severe NPDR	PDR
Normal	25809	0	0	1	0
Mild NPDR	0	2442	1	0	0
Moderate NPDR	1	0	5291	0	0
Severe NPDR	0	0	0	872	1
PDR	0	0	0	0	708

(d) J48					
	Normal	Mild	Moderate	Severe	PDR
Normal	25810	0	0	0	0
Mild	0	2440	3	0	0
Moderate	1	0	5291	0	0
Severe	29	0	0	840	4
PDR	10	0	0	12	686

demonstrated in Tables 15–17. The detailed performance metrics are computed and tabulated in Tables 18–20. The weighted average values of these performance metrics are summarized in Tables 21–23. It shows that SVM and MLP classifier's performance is not appropriate for the proposed feature extraction while using IDRiD and KAGGLE. In the case of SVM using the IDRiD database, the classifier misclassified the mild NPDR images as moderate NPDR. This might happen because of

Table 18

Detailed efficiency measures for multiclass classification using IDRiD Database.

Classifier	FPR	Specificity	Precision	Recall	F1-Score	Class
SVM	0.065	0.935	0.878	0.970	0.922	Normal
	0.00	1.00	—	0.00	—	Mild NPDR
	0.076	0.924	0.859	0.941	0.898	Moderate NPDR
	0.112	0.888	0.624	0.851	0.720	Severe NPDR
Random forest	0.003	0.997	0.933	0.286	0.437	PDR
	0.005	0.997	0.833	0.250	0.385	Mild NPDR
	0.058	0.942	0.894	0.993	0.941	Moderate NPDR
	0.056	0.944	0.763	0.824	0.792	Severe NPDR
MLP	0.005	0.995	0.939	0.633	0.756	PDR
	0.029	0.971	0.939	0.918	0.928	Normal
	0.00	1.00	—	0.00	—	Mild NPDR
	0.087	0.913	0.845	0.963	0.90	Moderate NPDR
J48	0.029	0.971	0.872	0.919	0.895	Severe NPDR
	0.008	0.992	0.939	0.939	0.939	PDR
	0.004	0.996	0.993	0.993	0.993	Normal
	0.00	1.00	1.00	0.950	0.974	Mild NPDR
J48	0.004	0.996	0.993	0.993	0.993	Moderate NPDR
	0.003	0.997	0.986	0.986	0.986	Severe NPDR
	0.003	0.997	0.980	1.00	0.990	PDR

Table 19

Detailed efficiency measures for multiclass classification using MESSIDOR Database.

Classifier	FPR	Specificity	Precision	Recall	F1-Score	Class
SVM	0.009	0.991	0.989	0.978	0.983	Normal
	0.025	0.975	0.843	0.915	0.878	Mild
	0.020	0.98	0.924	0.931	0.927	Moderate
	0.002	0.998	0.992	0.957	0.974	severe
Random forest	0.012	0.988	0.986	0.998	0.992	Normal
	0.003	0.997	0.979	0.928	0.953	Mild
	0.010	0.99	0.960	0.960	0.960	Moderate
	0.005	0.995	0.980	0.984	0.982	severe
MLP	0.005	0.995	0.995	0.995	0.995	Normal
	0.005	0.995	0.967	0.967	0.967	Mild
	0.012	0.988	0.956	0.960	0.958	Moderate
	0.008	0.992	0.968	0.965	0.966	Severe
J48	0.00	1.00	1.00	0.998	0.999	Normal
	0.001	0.999	0.993	0.993	0.993	Mild
	0.001	0.999	0.996	1.00	0.998	Moderate
	0.001	0.999	0.996	1.00	0.998	severe

input images. The same happens with the KAGGLE dataset too. Here, the MLP was not able to differentiate severe NPDR and PDR. But with the MESSIDOR dataset, the system performs very well. It is noticeable that the J48 classifier produces less misclassification for all the datasets. The average FPR is comparatively very less and the other performance measures (specificity, precision, recall, and F1-score) approximately equal 1 for the J48 classifier. The K-score and accuracy of the system are tabulated in Tables 24–26, which narrates the best performance of the J48 classifier with CNN feature extraction. The average K-score and accuracy of the J48 classifier for multiclass classification is 0.994 and 99.59% respectively.

The ResNet-50 architecture is also used as a feature extractor and the extracted features are then fed into the same classifiers for multiclass classification. The K-score and Accuracy for each classifier obtained from ResNet-50 extracted features are illustrated in Table 27.

Some works already existed for DR grading using CNN. The comparison of the existing approaches with the proposed method can be analyzed from Table 28. The method in [4] separates the lesions that are relevant for DR grading using a local network and then a global network is used for DR grading. They obtained both the non-weighted and weighted scores for evaluation metrics. The weighted network gives high values of evaluation metrics than the other. So, we compared the weighted values with the evaluation metrics of the proposed method. From the analysis, it is observed that the proposed

Table 20

Detailed efficiency measures for multiclass classification using KAGGLE Database.

Classifier	FPR	Specificity	Precision	Recall	F1-Score	Class
SVM	0.059	0.941	0.979	0.996	0.988	Normal
	0.002	0.998	0.971	0.968	0.969	Mild NPDR
	0.005	0.995	0.972	0.972	0.972	Moderate NPDR
	0.019	0.981	0.405	0.512	0.452	Severe NPDR
	0.00	1.00	0.815	0.031	0.060	PDR
Random forest	0.004	0.996	0.998	1.00	0.999	Normal
	0.00	1.00	1.00	0.999	0.999	Mild NPDR
	0.00	1.00	0.999	1.00	1.00	Moderate NPDR
	0.00	1.00	0.986	0.962	0.974	Severe NPDR
	0.00	1.00	0.994	0.969	0.981	PDR
MLP	0.175	0.825	0.941	0.999	0.969	Normal
	0.002	0.998	0.979	0.974	0.977	Mild NPDR
	0.003	0.997	0.981	0.981	0.981	Moderate NPDR
	0.00	1.00	–	0.00	–	Severe NPDR
	0.00	1.00	–	0.00	–	PDR
J48	0.00	1.00	1.00	1.00	1.00	Normal
	0.00	1.00	1.00	1.00	1.00	Mild NPDR
	0.00	1.00	1.00	1.00	1.00	Moderate NPDR
	0.00	1.00	0.999	0.999	0.999	Severe NPDR
	0.00	1.00	0.999	1.00	0.999	PDR

Table 21

Weighted average values calculated for each measures in Table 19 for multiclass classification using IDRiD database.

Classifier	FP rate	Specificity	Precision	Recall	F1 Score
SVM	0.066	0.934	–	0.811	–
Random forest	0.051	0.949	0.866	0.864	0.853
MLP	0.044	0.956	–	0.891	–
J48	0.003	0.997	0.990	0.990	0.990

Table 22

Weighted average values calculated for each measures in Table 19 for multiclass classification using MESSIDOR database.

Classifier	FP rate	Specificity	Precision	Recall	F1 Score
SVM	0.012	0.988	0.958	0.956	0.956
Random forest	0.009	0.991	0.978	0.978	0.978
MLP	0.007	0.993	0.978	0.978	0.978
J48	0.001	0.999	0.998	0.998	0.998

Table 23

Weighted average values calculated for each measures in Table 20 for multiclass classification using KAGGLE database.

Classifier	FP rate	Specificity	Precision	Recall	F1 Score
SVM	0.045	0.955	0.960	0.959	0.952
Random forest	0.003	0.997	0.998	0.998	0.998
MLP	0.129	0.871	–	0.949	–
J48	0.00	1.00	1.00	1.00	1.00

Table 24

Kappa statistic & validation accuracy of each classifier for multiclass classification using IDRiD Database.

Classifier	Correctly classified instances	K-score	Accuracy (%)
SVM	335	0.736	81.11
Random forest	368	0.812	89.10
MLP	1032	0.849	86
J48	409	0.986	99.03

work shows higher performance than the two-stage CNN while using the KAGGLE database. They compared the two-stage network with Alexnet architecture also, which shows low performance than their global network. When we look at the binocular network which is introduced in [15] provided a K-score of 0.82, a specificity of 0.707, and a recall value of 0.82. When compared to this, the proposed method shows an improved classification performance with the highest values for specificity, recall (value of 1), and K-score (value of 0.99).

Table 25

Kappa statistic & validation accuracy of each classifier for multiclass classification using MESSIDOR database.

Classifier	Correctly classified instances	K-score	Accuracy (%)
SVM	1147	0.936	95.58
Random forest	1174	0.968	97.83
MLP	1173	0.967	97.75
J48	1197	0.996	99.75

Table 26

Kappa statistic & validation accuracy of each classifier for multiclass classification using KAGGLE database.

Classifier	Correctly classified instances	K-score	Accuracy (%)
SVM	33696	0.904	93.93
Random forest	35067	0.996	99.83
MLP	33346	0.873	94.93
J48	35122	0.999	99.99

Table 27

Kappa statistic & validation accuracy of each classifier for multiclass classification using RESNET-50 features.

Database	Classifier	K-score	Accuracy (%)
IDRiD [34]	SVM	0.595	70.46
	Random forest	0.434	60.29
	MLP	0.887	91.76
	J48	0.901	92.46
MESSIDOR [35]	SVM	0.76	83.75
	Random forest	0.251	56.25
	MLP	0.863	88.23
	J48	0.892	91.22
KAGGLE [36]	SVM	0.79	84.13
	Random forest	0.52	69.03
	MLP	0.883	91.72
	J48	0.92	93.46

Another advantage of the proposed method is that using a simple CNN as a feature extractor, the system achieves higher performance on DR classification while the existing methods are using pre-trained networks and transfer learning methods for the classification which may increase the system complexity. That means the proposed network along with the J48 classifier is a better option for DR grading. In [14], with the modified Alexnet architecture they evaluated the system using the MESSIDOR dataset and attained an average accuracy of 96.25%. The proposed system scores more than this for DR classification (Both binary and multi-class). Since the 32 relevant features from CNN are fed into the classifier reduces the computation time of classification.

From the performance analysis, it is clear that the proposed system works well irrespective of the number of the training samples in each dataset, that is proposed CNN with J48 classifier seems to be the best one for DR classification. Thus by using CNN as a feature extractor, it reduces the computational time and complexity. The proposed system also compensates for the issues faced in deep learning classification by producing high precision and sensitivity scores. Our major contribution is reducing the CNN model parameters significantly to enable real-time deployment while improving the classification accuracy. So the novelty comes in using a CNN based feature extractor along with Machine Learning (ML) classifier (SVM, Random Forest, MLP, J48 Classifier) for the classification rather than using a softmax layer (multinomial logistic regression [22]), we found that features extracted from the CNN if further trained in ML classifier would give an outstanding result which can be a benchmark for DR classification. Also, the use of CNN enables the extraction of features directly from the fundus image without any segmentation and handcrafted feature extraction process. The proposed method also assures to consider the whole image and not leaving out any regions that may be affected due to DR. The issue in the proposed method is the variation in the evaluation metrics of each classifier while using different databases. One of the reasons might

Table 28

Comparison of Proposed work with existing methods.

Dataset	Methods	Specificity	Precision	Recall	F1-Score	K-score	Accuracy (%)
IDRID [34]	RESNET-50 + J48 [38]	–	–	–	–	0.901	92.46
	Proposed Method(multiclass)	0.997	0.990	0.990	0.990	0.996	99.03
MESSIDOR [35]	Modified AlexNet [14]	0.97	92.07	0.923	–	–	96.25
	RESNET-50 + J48 [38]	–	–	–	–	0.892	91.22
KAGGLE [36]	Proposed Method(multiclass)	0.999	0.998	0.998	0.998	0.996	99.75
	Y. Yang et al [4]	–	–	–	–	0.75	56
	X. Zeng et al [15]	70.7	–	0.822	–	0.82	–
	Y. H. Li et al.[18]	–	–	–	–	–	86.17
	RESNET-50 + J48 [38]	–	–	–	–	0.92	93.46
	Proposed Method(multiclass)	1.00	1.00	1.00	1.00	0.99	99.9

be the resolution of the images in each database. So, in the future, a feature extraction method that can overcome this difficulty can provide another breakthrough in automated DR detection and grading.

5. Conclusion

Clinicians are currently diagnosing the DR by examining lesions associated with disease-caused vascular anomalies. This approach is quite effective but it is a cost-effective technique and also the minute lesions may not be visible at the primary stage of DR. In the proposed work, an effective CNN is introduced for feature extraction and the features are fed into simple machine learning classifier for binary and multi-class classification of DR. The FPR, specificity, precision, recall, F1-score, K-score and accuracy obtained for each classifier (SVM, Random Forest, MLP, J48) are evaluated and concluded that the proposed network with J48 classifier gives the best performance than others. The proposed model is capable of providing competitive performance in DR classification with an average accuracy of 99.89% for binary classification and 99.59% for multi-class classification. Thus an effective classifier is developed by combining a simple CNN with ML classifiers for both DR detection and grading. Our major contribution is reducing the CNN model parameters significantly to enable real-time deployment while improving the classification accuracy. In the future, the same network may be used for other retinal disease detection with some modifications.

CRediT authorship contribution statement

Gayathri S.: Conceptualization, Methodology, Software, Validation, Writing - original draft, Data curation, Visualization. **Varun P. Gopi:** Supervision, Writing - review & editing. **P. Palanisamy:** Investigation, 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.

References

- [1] M.D. Abramoff, M.K. Garvin, M. Sonka, Retinal imaging and image analysis, IEEE Rev. Biomed. Eng. 3 (2010) 169–208, <http://dx.doi.org/10.1109/RBME.2010.2084567>.
- [2] N. Cheung, J.J. Wang, R. Klein, D.J. Couper, A.R. Sharrett, T.Y. Wong, Diabetic retinopathy and the risk of coronary heart disease, Diabetes Care 30 (7) (2007) 1742–1746, <http://dx.doi.org/10.2337/dc07-0264>.
- [3] J. James, E. Sharifahmadian, L. Shih, Automatic severity level classification of diabetic retinopathy, Int. J. Comput. Appl. 180 (2018) 30–35, <http://dx.doi.org/10.5120/ijca2018916244>.
- [4] Y. Yang, T. Li, W. Li, H. Wu, W. Fan, W. Zhang, Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2017, pp. 533–540.
- [5] M.P. Paing, S. Choomchuay, M.D.R. Yodprom, Detection of lesions and classification of diabetic retinopathy using fundus images, in: 2016 9th Biomedical Engineering International Conference (BMEiCON), 2016, pp. 1–5.
- [6] L. Seoud, J. Chelbi, F. Cheriet, Automatic grading of diabetic retinopathy on a public database, 2015.
- [7] D.K. Prasad, L. Vibha, K.R. Venugopal, Early detection of diabetic retinopathy from digital retinal fundus images, in: 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2015, pp. 240–245, <http://dx.doi.org/10.1109/RAICS.2015.7488421>.
- [8] S. Roychowdhury, D. Koozekanani, K. Parhi, DREAM: Diabetic retinopathy analysis using machine learning, IEEE J. Biomed. Health Inform. 18 (2014) 1717–1728, <http://dx.doi.org/10.1109/JBHI.2013.2294635>.
- [9] S. Gayathri, A.K. Krishna, V.P. Gopi, P. Palanisamy, Automated binary and multiclass classification of diabetic retinopathy using haralick and multiresolution features, IEEE Access 8 (2020) 57497–57504.
- [10] M.U. Akram, S. Khalid, A. Tariq, S.A. Khan, F. Azam, Detection and classification of retinal lesions for grading of diabetic retinopathy, Comput. Biol. Med. 45 (2014) 161–171, <http://dx.doi.org/10.1016/j.combiomed.2013.11.014>.
- [11] H. Pratt, F. Coenen, D.M. Broadbent, S.P. Harding, Y. Zheng, Convolutional neural networks for diabetic retinopathy, in: 20th Conference on Medical Image Understanding and Analysis (MIUA 2016), Procedia Comput. Sci. 90 (2016) 200–205, <http://dx.doi.org/10.1016/j.procs.2016.07.014>.
- [12] M. Mookiah, U.R. Acharya, R.J. Martis, C.K. Chua, C. Lim, E. Ng, A. Laude, Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: A hybrid feature extraction approach, Knowl.-Based Syst. 39 (2013) 9–22, <http://dx.doi.org/10.1016/j.knosys.2012.09.008>.
- [13] E.M. Shahin, T.E. Taha, W. Al-Nuaimy, S. El Rabiae, O.F. Zahran, F.E.A. El-Samie, Automated detection of diabetic retinopathy in blurred digital fundus images, in: 2012 8th International Computer Engineering Conference (ICENCO), 2012, pp. 20–25, <http://dx.doi.org/10.1109/ICENCO.2012.6487084>.
- [14] T. Shanthi, R. Sabeanian, Modified alexnet architecture for classification of diabetic retinopathy images, Comput. Electr. Eng. 76 (2019) 56–64, <http://dx.doi.org/10.1016/j.compeleceng.2019.03.004>.
- [15] X. Zeng, H. Chen, Y. Luo, W. Ye, Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network, IEEE Access 7 (2019) 30744–30753.
- [16] J. de la Calleja, L. Tecuapetla, M. Auxilio Medina, E. Bárcenas, A.B. Urbina Nájera, LBP and machine learning for diabetic retinopathy detection, in: E. Corchado, J.A. Lozano, H. Quintián, H. Yin (Eds.), Intelligent Data Engineering and Automated Learning – IDEAL 2014, 8669 (2014) 110–117, http://dx.doi.org/10.1007/978-3-319-10840-7_14.
- [17] D. Sidibé, I. Sadek, F. Mériadeuc, Discrimination of retinal images containing bright lesions using sparse coded features and SVM, Comput. Biol. Med. 62 (2015) 175–184, <http://dx.doi.org/10.1016/j.combiomed.2015.04.026>.
- [18] Y.-H. Li, N.-N. Yeh, S.-J. Chen, Y.-C. Chung, Computer-assisted diagnosis for diabetic retinopathy based on fundus images using deep convolutional neural network, Mob. Inf. Syst. 2019 (2019).
- [19] J. Sahlinen, J. Jaskari, J. Kivinen, L. Turunen, E. Jaanio, K. Hietala, K. Kaski, Deep learning fundus image analysis for diabetic retinopathy and macular edema grading, Sci. Rep. 9 (1) (2019) 1–11.
- [20] Z. Gao, J. Li, J. Guo, Y. Chen, Z. Yi, J. Zhong, Diagnosis of diabetic retinopathy using deep neural networks, IEEE Access 7 (2018) 3360–3370.
- [21] M.M. Butt, G. Latif, D.A. Iskandar, J. Alghazo, A.H. Khan, Multi-channel convolutions neural network based diabetic retinopathy detection from fundus images, Procedia Comput. Sci. 163 (2019) 283–291.
- [22] J. Wu, Convolutional neural networks, 2017, Published online at <https://cs.nju.edu.cn/wujx/teaching/15CNN.pdf>.
- [23] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015, arXiv preprint [arXiv:1502.03167](http://arxiv.org/abs/1502.03167).
- [24] N. Bjorck, C.P. Gomes, B. Selman, K.Q. Weinberger, Understanding batch normalization, in: Advances in Neural Information Processing Systems, 2018, pp. 7694–7705.

- [25] J.P. Kandhasamy, S. Balamurali Kadry, L.K. Ramasamy, Diagnosis of diabetic retinopathy using multi level set segmentation algorithm with feature extraction using SVM with selective features, *Multimedia Tools Appl.* (2019) 1573–7721., <http://dx.doi.org/10.1007/s11042-019-7485-8>.
- [26] G. Daqi, Z. Tao, Support vector machine classifiers using RBF kernels with clustering-based centers and widths, in: 2007 International Joint Conference on Neural Networks, 2007, pp. 2971–2976, <http://dx.doi.org/10.1109/IJCNN.2007.4371433>.
- [27] A. Roychowdhury, S. Banerjee, Random forests in the classification of diabetic retinopathy retinal images, in: S. Bhattacharyya, T. Gandhi, K. Sharma, P. Dutta (Eds.), *Advanced Computational and Communication Paradigms*, vol. 475, Springer Singapore, 2018, pp. 168–176, http://dx.doi.org/10.1007/978-981-10-8240-5_19.
- [28] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32, <http://dx.doi.org/10.1023/A:1010933404324>.
- [29] G.-B. Huang, Y.-Q. Chen, H.A. Babri, Classification ability of single hidden layer feed forward neural networks, *IEEE Trans. Neural Netw.* 11 (3) (2000) 799–801, <http://dx.doi.org/10.1109/72.846750>.
- [30] H. Saifuddin, H. Vijayalakshmi, Prediction of diabetic retinopathy using multi layer perceptron, *Int. J. Adv. Res.* 4 (2016) 658–664, <http://dx.doi.org/10.2147/IJAR01/714>.
- [31] S. Sharma, J. Agrawal, S. Sharma, Classification through machine learning technique: C4. 5 algorithm based on various entropies, *Int. J. Comput. Appl.* 82 (2013) 28–32, <http://dx.doi.org/10.5120/14249-2444>.
- [32] S. Yadav, S. Shukla, Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification, in: 2016 IEEE 6th International Conference on Advanced Computing (IACC), 2016, pp. 78–83, <http://dx.doi.org/10.1109/IACC.2016.25>.
- [33] S. Visa, B. Ramsay, A. Ralescu, E. Knaap, Confusion matrix-based feature selection, in: CEUR Workshop Proceedings, vol. 710, 2011, pp. 120–127.
- [34] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, F. Meriaudeau, Indian diabetic retinopathy image dataset (IDRiD): A database for diabetic retinopathy screening research, *Data* 3 (2018) 1–8.
- [35] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, J.-C. Klein, Feedback on a publicly distributed database: The Messidor database, *Image Anal. Stereol.* 33 (3) (2014) 231–234, <http://dx.doi.org/10.5566/ias.1155>.
- [36] <https://www.kaggle.com/c/diabetic-retinopathy-detection/data>.
- [37] M. McHugh, Interrater reliability: The kappa statistic, in: Biochemia medica : časopis Hrvatskoga društva medicinskih biokemičara / HDMB, vol. 22, 2012, pp. 276–282, <http://dx.doi.org/10.11613/BM.2012.031>.
- [38] S. Targ, D. Almeida, K. Lyman, Resnet in resnet: Generalizing residual architectures, 2016, arXiv preprint [arXiv:1603.08029](https://arxiv.org/abs/1603.08029).



Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy

Sraddha Das ^a, Krity Kharbanda ^a, Suchetha M ^{a,*}, Rajiv Raman ^b, Edwin Dhas D ^a

^a Centre for Healthcare Advancement, Innovation & Research, Vellore Institute of Technology, Chennai, India

^b Shri Bhagwan Mahavir Vitreoretinal Services, Sankara Nethralaya Eye Hospital, Chennai, India

ARTICLE INFO

Keywords:

Diabetic retinopathy
Maximum principal curvature
Hessian matrix
Squeeze-excitation
Bottleneck
Convolutional neural network

ABSTRACT

Diabetic retinopathy is ophthalmological distress, diabetic patients suffer due to clots, lesions, or haemorrhage formation in the light-sensitive region of the retina. Blocking of vessels leads, due to the increase of blood sugar leads to the formation of new vessel growth, which gives rise to mesh-like structures. Assessing the branching retinal vasculature is an important aspect for ophthalmologists for efficient diagnosis. The fundus scans of the eye are first subjected to pre-processing, followed by segmentation. To extract the branching blood vessels, the technique of maximal principal curvature has been applied, which utilizes the maximum Eigenvalues of the Hessian matrix. Adaptive histogram equalization and the morphological opening, are performed post to that, to enhance and eliminate falsely segmented regions. The proliferation of optical nerves was observed much greater in diabetic or affected patients than in healthy ones. We have used a convolution neural network (CNN) to train the classifier for performing classification. The CNN, constructed for classification, comprises a combination of squeeze and excitation and bottleneck layers, one for each class, and a convolution and pooling layer architecture for classification between the two classes. For the performance evaluation of the proposed algorithm, we use the dataset DIARETDB1 (standard Diabetic Retinopathy Dataset) and the dataset provided by a medical institution, comprised of fundus scans of both affected and normal retinas. Experimental results show that the proposed algorithm provides improved results, when compared to traditional schemes. The model yielded an accuracy of 98.7 % and a precision of 97.2 % while evaluated on the DIARETDB1 dataset.

1. Introduction

People with Diabetes face a medical condition called Diabetic Retinopathy (DR), which is of two types, (a) Non-Proliferative Diabetic Retinopathy (NPDR), the milder form, (b) Proliferative Diabetic Retinopathy (PDR), the advanced form. The initial signs of the DR are the exudates, which indicate the milder condition, i.e. NPDR. In NPDR, the patient will only have blurry vision, but as the disease develops, the retina then grows new blood vessels that highly affect the vision. These abnormal blood vessels may leak or bleed easily, causing blood clots/blobs to emerge in the retina. Damage to the network of vessels that nourish the retina is mostly the key cause of DR. In the superior stages of the PDR, the blood vessels are entirely blocked, forming lesions in the blood vessels. The most visible lesions which appeared are the microaneurysms and hemorrhages. Micro-aneurysms are the first visual symptom of DR, which appear as small round-shaped red dots in the fundus. Currently, a trained ophthalmologist detects the DR by assessing

the fundus image manually. There is a need for an automated DR screening system that detects the condition quickly and accurately. Various methods have been worked on to build a system with high accuracies, such as a unique unsupervised method. For the detection and segmentation of the fovea in the retinal image, pixel-level exudate detection using deep-learning-based algorithms are used. In deep learning, CNN is a class of deep neural network which contains different layers of neurons. Each neuron of a layer is connected to all the neurons of the next layer. CNN has wide applications in the field of image classification.

In this paper, we have implemented a scheme that can be used to detect the DR. For diabetic patients, the unusual growth of blood vessels in the eye is the most prominent trait. Therefore, estimating the branching retinal vasculature is an important aspect for ophthalmologists, to perform an efficient diagnosis. The common classification algorithm involves the steps such as pre-processing, segmenting the fundus scans of the eye to have a clear view of the branches, followed by

* Corresponding author.

E-mail address: suchetha.m@vit.ac.in (S. M.).

a classification algorithm. Fundus scans are commonly used for the diagnosis of DR. Fig. 1 shows the sample fundus scan images where the blood vessels can be easily diagnosed. Different algorithms are used in segmentation that includes thresholding, edge-based segmentation, region-based segmentation, etc. After referring to different techniques for segmentation, we worked on the technique of Maximal principal curvature, which utilizes the maximum Eigenvalues of the Hessian matrix. Also, for the classification process, different algorithms are commonly used which include support vector machine, clustering-based approaches such as Fuzzy C means (FCM), and K-means. Several classification techniques also exist that include the methods of attention-aware, self-attention, ConvLSTM, and One-dimensional deep attention convolution network.

Our proposed method includes a novel deep learning architecture to classify segmented fundus images like 'Normal' and 'Abnormal', i.e., healthy and diabetic retinopathy affected. Several approaches utilize the CNN to learn features in the retinal scans; we have worked on segmented images. For this, a method of finding maximum principal curvatures is deployed, which extracts the blood vasculature effectively. Process of adaptive histogram equalization and the morphological opening is performed to enhance and eliminate falsely segmented regions. To improve the performance of the classifier, we constructed a two-way classification CNN model. One way for classification contains a combination of the squeeze, excitation, and bottleneck layers. The other one includes convolution and pooling layer architecture. The architecture of both modules is independent and distinct to perform robust feature extraction. A test image is taken as input into both networks, and only if the classification is unanimous, the result is accepted. Due to this, the model gives a significantly better performance than prior methods.

The paper is structured in the following manner. Section II contains a description of related works in DR detection and classification, Section III elaborates on the working of the proposed algorithm. Section IV contains the results and discussion of the proposed work and finally, Section V concludes the paper.

2. Related works

Detection of eye-related diseases from fundus images is an emerging area for researchers these days. Especially, automated detection of DR has become an important topic for researchers working in the field of medical image processing and computer vision.

Several papers focus on the segmentation process because the classification result mainly depends on the segmentation output. Wrong segmentation results can lead to the wrong classification because the features are extracted from non-DR pixels. The following paper focuses on different segmentation approaches. Especially, the authors, Santosh et al. [1] presented a method to segment the retinal fundus images with remarkable accuracy. It proceeds in two stages: Pre-processing of the

input image and post-processing of the pre-processed image by using maximum principal curvature. Gehad et al. [2] proposed a blood vessel segmentation approach. The technique can be used in retinal image analysis to extract the retinal image vessels. Mathematical morphology and K-means clustering are used to segment the vessels. Smoothing operation on the retinal image using mathematical morphology to enhance the blood vessels and suppress the background information. Finally, the K-means clustering algorithm is used to enhance the image. The proposed approach is tested on the DRIVE dataset and is compared with alternative approaches. This algorithm succeeded with an average accuracy of 95.10 %. Orlando et al. [3] present a discriminatively trained method based on extended description and evaluation for blood vessel segmentation in fundus images. The problem statement for this paper is to overcome the difficulty when Standard segmentation priors such as a Potts model or total variation deal with thin and elongated structures. The model is trained in a way that parameters of the method are learned automatically through a structured output support vector machine. Also, the authors R.Manjula et al. [4] developed a paper that aims to employ image processing techniques to enhance and measure the dimensions of the retinal blood vessels. Three techniques have been implemented to carry out the process of segmentation; Gaussian method, mathematical morphology method, and multi-scale analysis method. Gaussian method uses a Gaussian resolution hierarchy with an objective to detect thick and thin vessels. It is a faster technique but it is suitable only for detecting thick vessels. The mathematical morphology technique is more suitable to detect the fine details of thin vessels with further precision. The third technique, the multi-scale analysis method, is preferable for invariant analysis along with the transformation of images as it detects the thick and thin vessels without noise. Memari et al. [5] propose an automatic retinal vessel segmentation that utilizes fuzzy c-means clustering and level sets. Contrast limited adaptive histogram equalization is used to contrast enhance the retinal images, while the noise is condensed through mathematical morphology techniques along with subsequently matching filtering steps that use Gabor and Frangi filters, which improve the blood vessel network preceding the clustering. For extracting an initial blood vessel network, a genetic algorithm enhanced spatial fuzzy c-means scheme is used. The segmentation is further refined by an integrated level set approach. This method of segmentation achieved a mean accuracy of 96.1 %. Also, Budai et al. [6] aim to reduce the running time of the algorithm and segmenting the vessels in the fundus image. A method is proposed to reduce calculation time, that achieves high accuracy and increase sensitivity against the Frangi method. The authors have also work to avoid potential problems like specular reflexes of thick vessels while building this approach. They used two public databases, DRIVE and STARE, each obtaining accuracy of 95.72 % and 93.86 % respectively.

The segmentation methods such as superpixels based segmentation, watershed segmentation, and active contour methods are also very

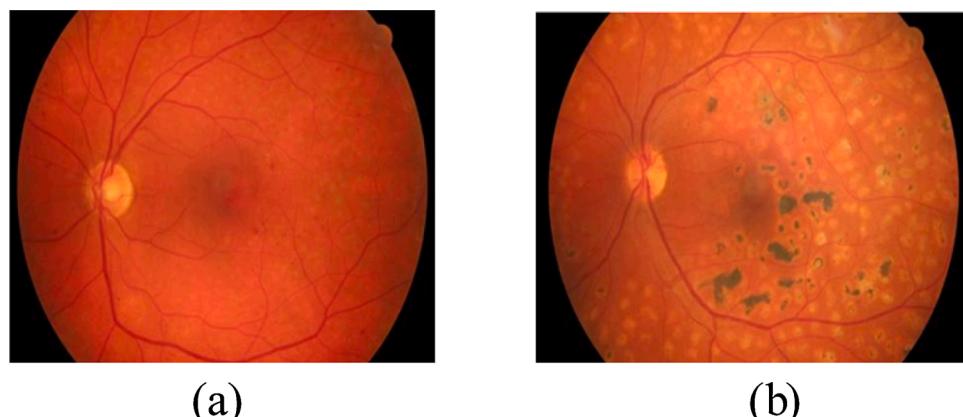


Fig. 1. Scans of fundus images (a) without DR (b) with DR.

popular segmentation schemes. In superpixels, based segmentation [7, 8], the input image is segmented into small homogeneous regions or superpixels. Superpixel segmentation provides an over-segmentation of a picture by gathering pixels into homogeneous bunches dependent on intensity, texture, and different other highlighted features. Every superpixel belonging to the same region label and arrangement of superpixels provides a map, which takes into account the local and nonlinear deformations of interesting regions. A small number of superpixels with more pixels reduce the computational cost and memory usage and make the superpixel-based feature more reliable and can minimize the risk of assigning wrong labels to the superpixels. The watershed transform [9,10] is a popular and reliable unsupervised model, applied to solve diverse image-segmentation problems, which is used by the field of mathematical morphology. This has been applied to solve various difficult problems of image segmentation, successfully. On a topographic surface, watershed transformation helps to determine the watershed lines. However, applying the watershed transform directly to the original image leads to over-segmentation. As the morphological watershed transformations are sensitive to noise and contrast in an image. The over-segmentation can be sufficiently serious to render incorrect segmentation results. Very often Active contour method (ACM) is used in digital images for several applications, including edge detection, segmentation, stereo matching, and shape recognition. It is considered as a most valuable tool for segmenting regions of interest (ROIs) in medical images. ACM [11,12] has two models, edge-based models, and region-based models. Edge-based models use local edge information to fit the boundaries of the approximated shape. Region-based active contour methods find the optimal energy for which the model fits the image best based on statistics calculated from subregions.

Several papers exist for detecting and classifying the DR present in the fundus images. For example, Dulanji et al. [13] exhibit that different morphological operations could be incorporated for accurate detection of exudates while focusing on optic disc detection. As it is the essential factor for the removal of false positives. Otherwise, the optic disc could be falsely recognized as exudates. They achieved sensitivity and specificity of 94.5 % and 88.46 % respectively. The authors Renoh et al. [14], introduces a unique unsupervised method to detect and segment OD and fovea from the retinal images. The proposed method consists of three steps, namely, coarse ONH center detection, fine-tuned ONH center and boundary detection, and fovea detection. They have proposed a method that automatically detects the optic disc (OD) using histogram-based template matching along with the maximum sum of vessel information in the retinal image. The optical disc detection accuracy came out to be 95 % and fovea detection accuracy to be 97.26 %. Also, Ramasubramanian et al. [15] work on hemorrhages and propose a solution for the efficient detection of hemorrhages without segmenting or eliminating any regions. Here a method for the automatic detection of hemorrhages in color retinal images is proposed and validated. The color retinal images used as an input that are captured from the diabetic patients are enhanced, using a bag of features based on intensity, color and texture, by an effective pre-processor. Finally, the features are classified with the help of a partial least square classifier. They obtained the AUC-ROC value of 0.98. The authors, Avula et al. [16] deploy a deep learning model to detect the exudates in the retinal images. Here they have tried to generate an image in which hard exudates are detected by predicting the class of every pixel. The model presented in this paper is developed using Tensorflow deep learning framework. For this method sensitivity and specificity attained was 98.29 % and 41.35 % respectively. Shuang et al. [17] focus on potential exudate candidate points that are first extracted with morphological ultimate opening methods and then the candidate points are passed for classification to the trained CNN deep networks. Before putting the input to the designed algorithm they first pre-process the image to remove the optic disc and blood vessels. The CNN network is trained on the 64*64 patches. They achieved a pixel-wise accuracy to be 91.92 %.

Shailesh et al. [18] proposes an improved blood vessel detection technique by morphological iterative process and develops an automated Optic Disc algorithm. They have worked to bring improved techniques for microaneurysm as well as hemorrhages detection, eventually contributing to the overall improvement in the early detection of DR. The method consists of five stages- pre-processing, detection of blood vessels, segmentation of optic disc, localization of fovea, feature extraction and classification. Pre-processing and blood vessel detection are performed using Mathematical morphology operation. Optic disc segmentation is done using Watershed and the classification of the diseases is done by a Radial basis function neural network. This paper developed sensitivity and specificity of 87 % and 93 % respectively. N. Yalçın et al. [19] proposed an approach of two steps. Step 1 is to perform pre-treatments to remove retinal images from different datasets and standardize them to size. In Step 2, classification was made by Convolutional Neural Network. Instead of creating the feature set manually as in other traditional methods, the deep learning network automatically constructs itself in a very short time by using the CPU and GPU in the training phase where 98.5 % success was attained. Shah et al. [20] came up with an AI-based algorithm, which helps to detect DR. The authors have validated using an internal dataset consisting of 1533 macula-centered fundus images and an external MESSIDOR dataset. Interobserver agreement using kappa value was calculated for both the sets and two out of three agreements for DR grading was considered as ground truth to compare with AI results. The outcome was a sensitivity of 90.4 % and specificity of 91 %. Xu et al. [21] presented a complex architecture containing a compound of stacked layers; 15 layers out of which 13 are convolutional layers and two fully connected layers. It is a two-class classification model, with each class having a precision value of 0.81 and 0.88 respectively. Also, the authors discovered that when pre-processing is done using contrast limited adaptive histogram equalization along with ensuring dataset fidelity by expert verification of class labels enhances the recognition of refined features. Andonova et al. [22] used the MESSIDOR database and proposed a CNN with 4 convolutional layers. Input images are first pre-processed, transformed, and normalized. This enhanced their quality. For classification purposes, they are split into multiple groups (clusters). This method developed max accuracy of 82.5 %. Also, Prentasic et al. [23] developed an algorithm based on CNN to detect exudates from the color fundus images. CNN or deep neural networks helps to segment neuronal membranes in electron microscopy. The convolutional neural network in this paper calculates the probability of a pixel in one of the two classes; exudate or non-exudate class. This method attains a sensitivity of 77 %. Adem K [24] proposed a hybrid approach of circular Hough transform and CNN algorithms for detecting the exudates. Three DiaretDB0, DiaretDB1, and DrimDB public datasets were used to assess. While testing with the images not included in the training set, it was found to have a correct classification ratio of 99.17 % in DiaretDB0, 98.53 % in DiaretDB1, and 99.18 % in DrimDB.

The following papers focus on different approaches that classify the DR using Neural Networks. The authors, Kejie et al. [25] put forward various concepts such as a shortcut connection and an attention mechanism for multi-task deep learning models which can be used to optimize the process of representation sharing. This Attention-aware Multi-task Convolutional Neural Network can automatically learn appropriate sharing via end-to-end training. To suppress redundant contents contained in the representations the attention mechanism is introduced in this paper. M. Li et al. [26] presents a discussion on the design of a cost-effective universal retinal fundus camera and the building of a new algorithm for the identification of prominent vision-threatening diseases. They also introduced an autoencoder neural network and a novel 3D self-attention convolutional neural network. Hemanth et al. [27] in their study included image processing with histogram equalization, and the contrast limited adaptive histogram equalization techniques and performed the classification using a convolutional neural network. 400 retinal fundus images from the MESSIDOR database were used to

validate the method which achieves a precision of 94 %. Shuyuan et al. [28] proposed an algorithm to automatically extract and classify the different features of the signals. In the first step, one-dimensional sparse filters are designed to learn hierarchical features of raw signals. Second, an attention layer is constructed to weigh and assemble feature maps, to derive more context-relevant representation. Wen-Shuai et al. [29] use long short-term memory (LSTM), as a special deep learning structure to improve the classification performance. Zhao et al. [30] have used visualized feature maps of DenseNet and CondenseNet CNN architectures. The authors also propose a relative-squeezing bottleneck design to improve the computing efficiency of Convolutional Neural Networks. Three benchmark datasets: CIFAR-10, CIFAR-100, and ImageNet are used to evaluate the method. Jun et al. [31] proposed a new algorithm called a deep residual squeeze and excitation network, to enhance the representation ability of the network, as a building blocks in DRSEN. Also, to reduce the parameters of the network, work has been done to improve the up-sampling module and the global residual pathway in the network. Lahmri [32] suggests SENet architecture, that generalizes extremely challenging datasets. In this process, deep learning convolutional neural networks are used for automatic feature extraction. And the Student *t*-test is applied to the high dimensional features set, which is extracted by CNN, to select the best ten features. Finally, to perform the classification task, the selected CNN-based features are fed to a nonlinear support vector machine tuned by Bayes optimization.

Instead of using a single-stage classifier as discussed in most of the classifiers, the proposed method includes robust two independent CNN modules for classification, containing the squeeze, excitation, and bottleneck units. The maximum principal curvature algorithm performs a highly efficient segmentation of the fundus images to extract the proliferative blood vasculature as the problem of vessel trench detection is categorized by high magnitudes of the curvature. The adaptive histogram equalization and the morphological opening processes are used to eliminate the wrongly segmented regions. In the early advanced stages of DR, detection of new blood vasculature growth serves as a preventive step before the blood vessels leak and hamper vision later on. Certain constraints entail as well. As stated in most of the articles about diabetic retinopathy, in the advanced conditions of the DR, the patients develop an abnormal growth of blood vessels, which tend to leak and hamper vision. Sometimes when patients are near the advanced stage of the DR or in milder conditions may not have enough proliferation of blood vessels to be detected.

3. Proposed methodology

Our proposed architecture involves two modules used for classification- the memory module and a central CNN. In the memory module, a squeeze operation squeezes a feature. Max-pooling blocks are a major part of the squeeze process that enhances the informational features and suppresses the inefficient features. Batch normalization layers expedite the convergence and thus improve the stability. Non-linear ReLU activation function models each, ‘Normal’ and ‘Abnormal’ data. A bottleneck layer of convolutional layers decreases the model’s complexity. The purpose of the bottleneck layer is to reduce the number of channels by a certain degree using a poor 1×1 convolution so that the latter 3×3 convolution has lesser parameters. The network at the end is widened again with another 1×1 convolution process where our bottleneck includes three convolution layers. The squeeze, excitation, and bottleneck units in the memory module perform robust feature extraction and reduce the overall complexity of the model. Powerful feature extraction is highly essential when dealing with such medical images that contain fine structures. It also enhances the CNN performance with only a minimal increase in the number of total parameters. Finally, ReLU processing is performed on the added results using an FC, followed by the classification layer after softmax activation. The other module, the central CNN, has a basic convolutional layer architecture ending in an FC layer. Two different network architectures are responsible for the

classification of the same image and finally categorizing a particular image to a certain category, only if both the modules give the same result. The images are segmented using the method of maximum principal curvatures, which efficiently extracts the fine blood vasculature. Then adaptive histogram equalization and the morphological opening are performed to enhance and discard incorrect regions. The various layers in the Memory module and bottleneck layer of the suggested deep-learning architecture are given in Fig. 2.

The retinal area in diabetic patients contains eminent vasculature. Our proposed algorithm for the detection and classification of DR in blood vessels caused due to diabetes has been explained below. Fig. 3 depicts the block diagram of the proposed DR detection and classification algorithm.

The proposed work has three stages for DR classification. The first stage is pre-processing, the second stage is segmentation and the third stage is classification. The algorithm of the proposed classification process can be summarized as shown below,

3.1. Algorithm

Step (i): Each image undergoes pre-processing before being used for any step ahead. Pre-processing involves resizing the image to a common size and converting it to a grayscale image format. Usually, the greyscale image is estimated from the R, G, and B values using the relation,

$$G = [0.299R + 0.587G + 0.114B] \quad (1)$$

Since we aim to extract the blood vessels and red clots; we provide higher weightage to red color when compared to other two-color components. Therefore, the greyscale conversion can be modified as,

$$G = [0.587R + 0.299G + 0.114B] \quad (2)$$

Step (ii): The second step is to segment the image; Eigenvalues of a Hessian matrix are calculated to calculate maximum principal curvature. The image that is acquired by scan machines can induce noise in the fundus images. To suppress such noise, a Gaussian filter using a Gaussian function is used to smoothen the image. The morphological opening is used to remove any gaps between pixels in the image.

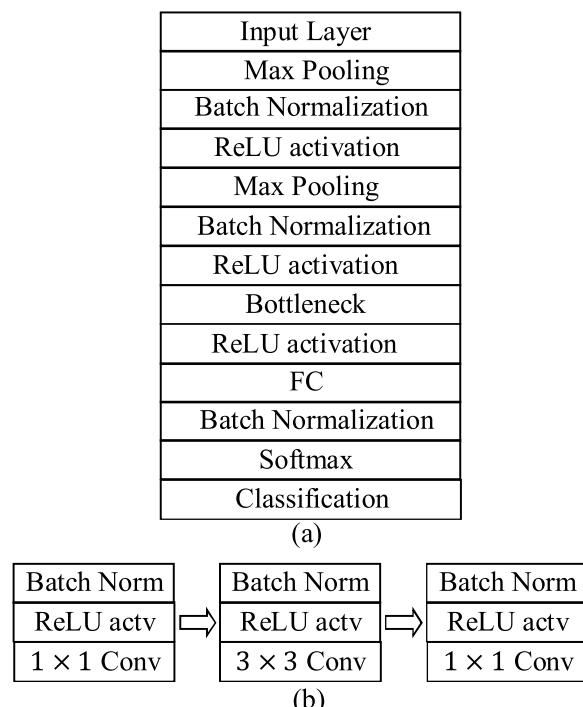


Fig. 2. Architecture of (a) Memory module and (b) Bottleneck layer.

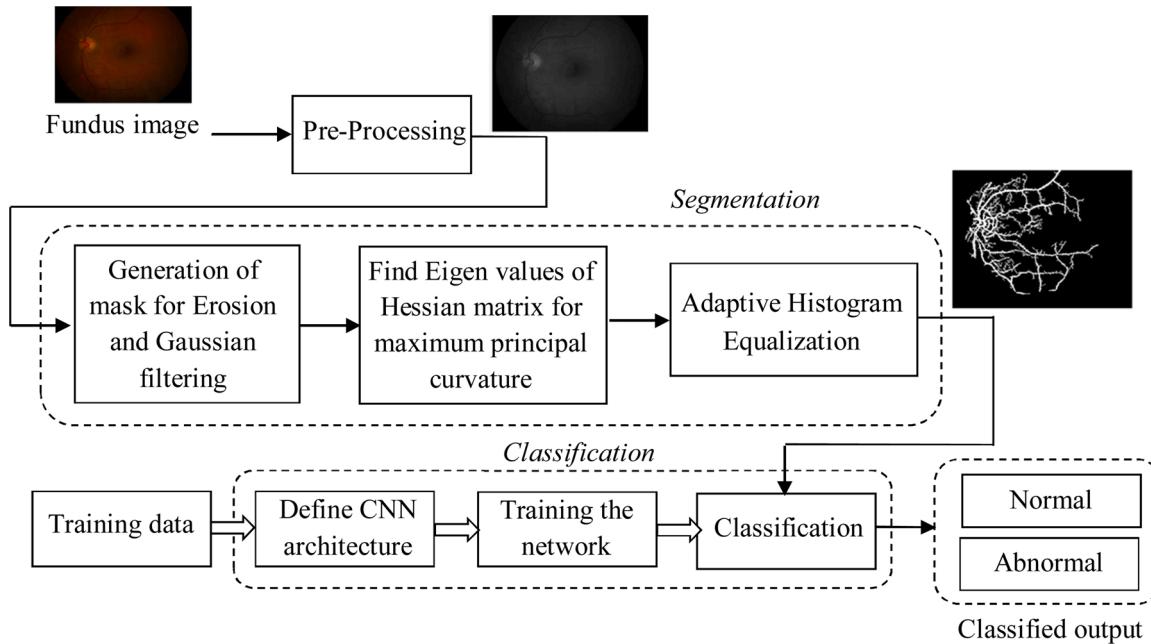


Fig. 3. Overview of the proposed algorithm.

Morphological operator erosion is used to remove the pixels that lie on the boundary which appear like small objects so that only substantive objects remain. An octagonal structuring element of size 24 is used which provides the optimal result. The blood vessels are detected so that further analysis can be done to detect the DR using the maximum principal curvature algorithm. The differences in the gradient of intensities about the neighborhood of a pixel best describe the principal curvature in which the vasculature with a ridge construct in the image emerges. A Hessian matrix can estimate the maximum principal curvature to the direction of the eigenvector which is described for every pixel value and can be defined as the highest Eigenvalue of the calculated second-order derivative. For a Hessian H , given as Eq. (4),

$$H = \begin{pmatrix} I_{xx} & I_{xy} \\ I_{yx} & I_{yy} \end{pmatrix} \quad (3)$$

$Z(x,y)$ is a function of intensity values of the image. The second-order derivative of $Z(x,y)$ with respect to x is Z_{xx} and with respect to y is Z_{yy} . Differentiation of Z_x with respect to y results in mixed partial derivative Z_{xy} . For a unit vector $(a, b)^T$ in the y direction, its second-order form is calculated as shown in Eq. (5).

$$[A \ B] = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \quad (4)$$

Product of Eigenvalues of a Hessian produces its determinant. The eigenvector having the largest Eigenvalue gives the direction of maximum curvature. Eigenvalues or principal curvatures determine the extreme bending of a regular plane at every point or variations in the gradient of intensity with regards to the neighborhood of a pixel in which the upcoming vasculature is ridge shaped. Vessel trench is detected using the algorithm that is used for the measurement of curvature [12]. The Eigenvalues of the Hessian matrix be $\lambda'_{r,t}$ and $\lambda''_{r,t}$ at any pixel location (r,t) . Therefore, the principal curvature corresponding to minimum and maximum can be expressed as,

$$\hat{\lambda}_{r,t}^- = \min\{\lambda'_{r,t}, \lambda''_{r,t}\} \quad (5)$$

$$\hat{\lambda}_{r,t}^+ = \max\{\lambda'_{r,t}, \lambda''_{r,t}\} \quad (6)$$

Let $U \times V$ be the size of the greyscale image, then the minimum and

maximum principal curvature matrices can be represented as,

$$M^- = [\hat{\lambda}_{r,t}^-]_{U \times V} \quad (7)$$

$$M^+ = [\hat{\lambda}_{r,t}^+]_{U \times V} \quad (8)$$

The matrix for adaptive principal curvature can be expressed as, $M = \frac{1}{2}(M^+ + M^-)$. However, the maximum principal curvature provides better performance to extract the blood vessels, so we consider the maximum principal curvature expressed by Eq. (11).

Step (iii): After segmentation, the image passes through two sections, one, excitation and bottleneck block, and, other, Convolutional Neural Network. The output is obtained individually for both paths.

Step (iv): While training, the section of the squeeze and excitation and bottleneck blocks is trained only on normal images and abnormal images. This serves as a memory module for the process. The next step involves, feeding the whole training the CNN with the entire training data, i.e. both abnormal and normal images.

Step (v): For the test image to be classified, it first passes through the excitation and bottleneck blocks, giving *output 1*, while the test image passes through CNN, giving *output 2*.

Step (vi): If both the *output 1* and *output 2* have classified the image as normal, then the image is classified as a normal image. If both the *output 1* and *output 2* has classified the image abnormal, then it is declared and displayed as abnormal. Else, the image is further processed using the CNN algorithm to obtain the classified result. The overview of the proposed system for output generation of diabetes detection is shown in Fig. 4.

To train the CNN architecture effectively, the training data is split into two groups, the ‘Normal’, and the ‘Abnormal’. The CNN takes an input mage of 336×448 and passes it through several layers of convolution and max-pooling. The first convolution layer contains 10 number of 9×9 filters preceding a 2×2 max-pooling layer and the next convolution layer contains 10 number of 6×6 filters followed by a 3×3 max-pooling layer. Following this is an FC layer of output size 2, batch normalization layer, softmax layer, and finally the classification layer which also uses ReLU activation. At a learning rate of 0.00001, training of the network is done for 20 epochs. Since the segmented images show the volume of blood vessels present, the model is trained in a way that if

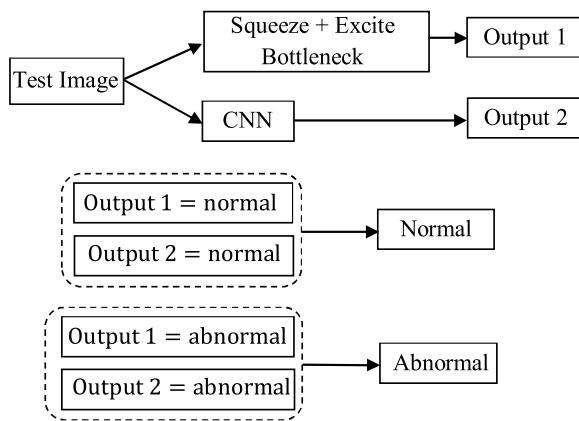


Fig. 4. Overview of output generated by the proposed method.

the segmented image shows an abnormal growth of blood vessels, then it is classified as abnormal. Whereas, if the segmented image shows no unusual growth of the blood vessels, then it is classified as normal. So, the feature that is used to decide the condition of the patient is the blood vessel structure present inside the retina. The central CNN is trained on both classes that consist of convolutional layers with an FC layer at the end to classify any input image to one of the two categories, ‘Normal’ or ‘Abnormal’. Only if the prediction of either memory module coincides with that of the main CNN, then the image of the fundus is classified into that particular category, else it is further processed till classification completes. The next section shows the experimental results and discussion of the proposed work.

4. Experimental outcomes and discussion

The proposed algorithm was assessed using the images taken from the dataset and the practical images obtained from the hospital.

4.1. Data collection and processing

Our data contains high-resolution fundus (HRF) images [5] that are distributed into two groups, i.e. healthy, and diabetic retinopathy, each comprising twenty images. For testing purposes, eye hospital Sankara Nethralaya has provided us with 80 patient images from their patient database. Also, we have used the standard Diabetic Retinopathy Dataset [33] (DIARETDB1), which contains the fundus images of 89 patients. The DIARETDB1 dataset contains the retina fundus that was acquired at the 50° field of view. The dataset contains 84 DR images and 5 normal images each having a size of 1500 × 1152 pixels. The dataset images are annotated by 4 medical experts for the presence of microaneurysms, soft exudates, hard exudates, and hemorrhages [34,35]. The experimental

results were evaluated based on k-fold cross validation where k is chosen as 5. The proposed architecture was trained using $k - 1 = 4$ folds and the model was tested on the remaining 5th fold. The performance metric was recorded for all the k tests (iterations) and the average score is considered as the performance of the proposed model. The performance was measured without data augmentation and with the data augmentation process. The data augmentation includes flipping such as horizontal and vertical flipping and rotation such as 45°, 90°, 135°, 180°, 225°, 270° and 315° rotation. We have done pre-processing such as cropping the background and resized the image to 512 × 512 and converted to grayscale to make the image to be uniform. Figs. 5 and 6 shows the sample images from the dataset DIARETDB1 and hospital respectively.

4.2. Experimental outcomes

The proposed system for the detection of Diabetes based on the proliferation of optical vasculature has been implemented using the tool MATLAB 2018b with Computer Vision System Toolbox. In our work, the maximum principal curvatures are determined using the ‘lambda’ function in MATLAB which performs region growing along the direction of their eigenvectors leading to the segmentation of the vasculature. When an image is fed in, it first goes through the steps of pre-processing, including resizing, converting to greyscale, and getting segmented output by the method of maximum principal curvature, which branches out the blood vessels. The standard deviation of the segmentation result varies between 0.43 to 3.79. To overcome any deficiencies in the data, utilizing the data available fully, augmentation techniques of data were carried out which included flipping and rotation. The image is then ready to be feed into both the memory block and the CNN. The central CNN architecture involves, two max-pooling layers, a couple of convolution layers, two FC layers, ReLU activation layers, with a softmax layer concluding in a classification layer. The memory block which includes the squeeze plus excitation and the bottleneck block, too involves convolution layers, batch Normalization layers, ReLU layers, and max-pooling layers. It takes twenty epochs, 0.0001 learning rate, and stochastic gradient descent with momentum optimizer.

While training, the whole training data is divided in the ratio of 4:1 as prediction and validation data. The validation data is further utilized to calculate the training accuracy and also other performance parameters. When a test image is taken and fed into the network, the image undergoes the segmentation process, and further, it was classified as an ‘Abnormal’ or ‘Normal’ image.

Fig. 7 shows some of the sample segmentation results for the dataset images DIARETDB1. The proposed method segments the fine blood vessels which can be best suited for classification. The segmentation process does not skip fine details present in the fundus image to the branching of blood vessels. Also, Fig. 8 shows the segmentation results

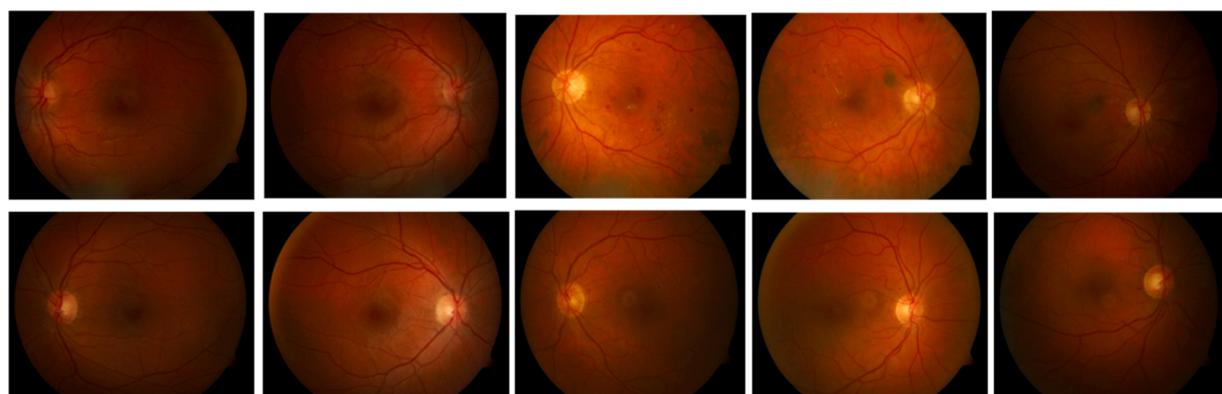


Fig. 5. Sample test images from Standard Diabetic Retinopathy Dataset (DIARETDB1) [30].

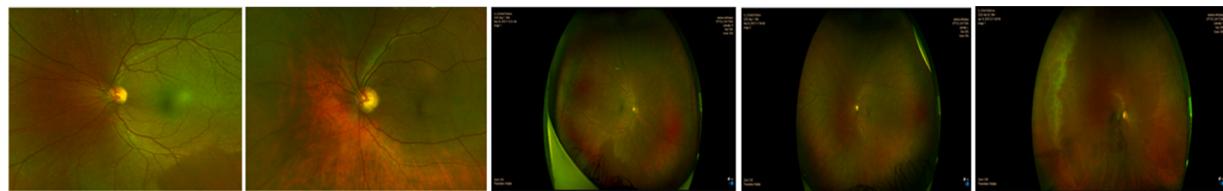


Fig. 6. Sample test images obtained from the Hospital.

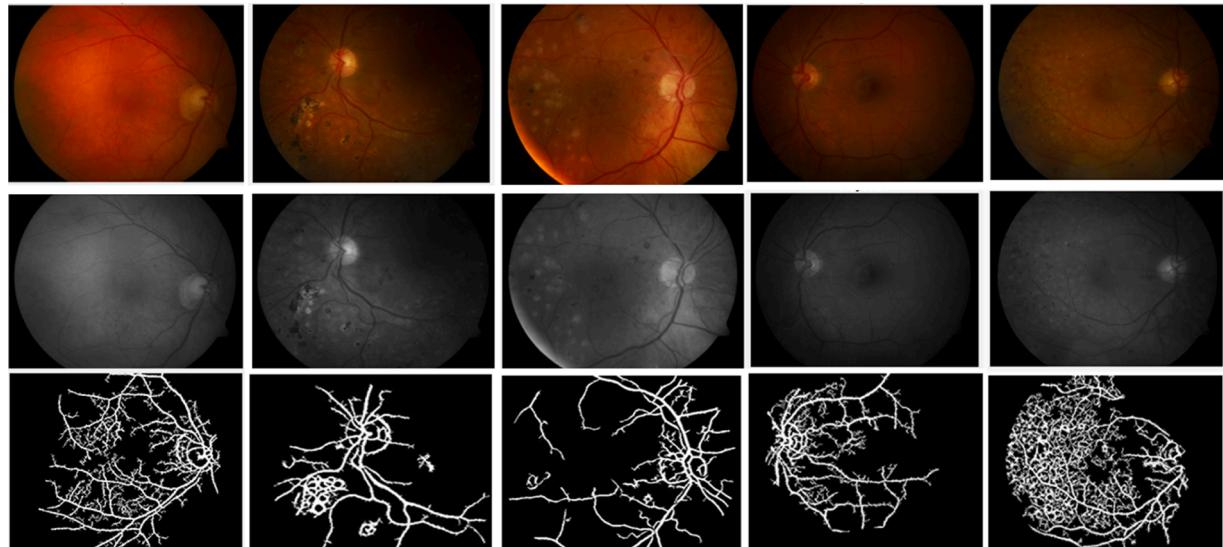


Fig. 7. Some of the segmentation results for the dataset DIARETDB1 images that are classified as abnormal (*First row*: Input image, *Second row*: Greyscale converted, *Third row*: Segmented output).

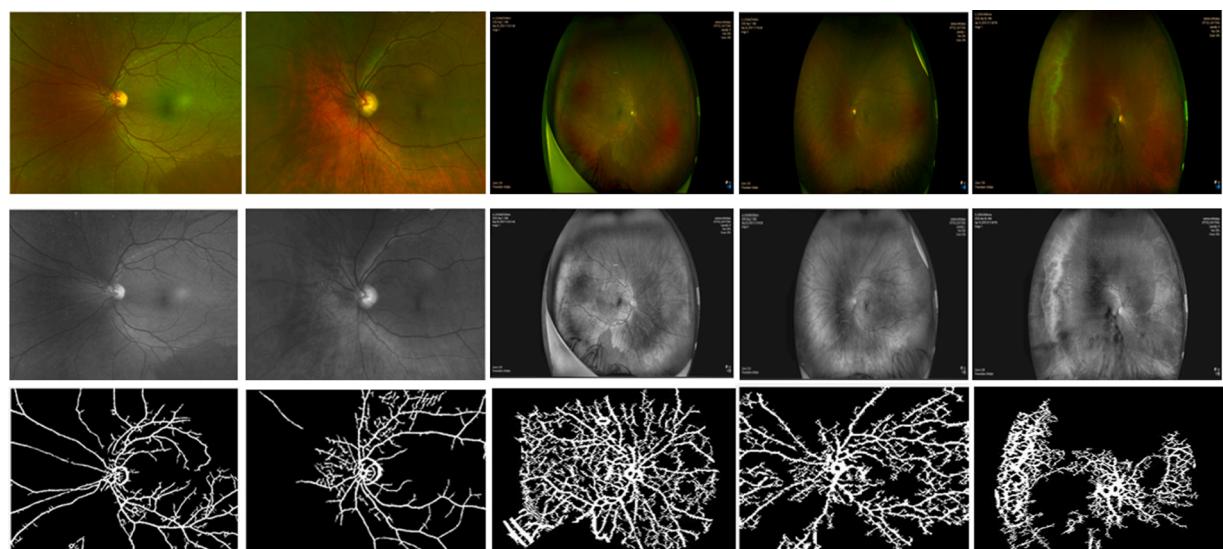


Fig. 8. Some of the segmentation results for the hospital images (*First row*: Input image, *Second row*: Greyscale converted, *Third row*: Segmented output).

for the hospital images. Even though the input image highly varies from the dataset images, the proposed segmentation process also obtains the fine details present in the hospital images without missing any major blood vessels.

The segmentation result of the proposed method highly depends on the greyscale conversion, the traditional conversion using Eq. (4) that misses the fine details as shown in Fig. 9 (c). But the greyscale conversion by providing high weightage to R channel using Eq. (5) provides

fine details at the output as shown in Fig. 9(e).

The performance of the algorithm was validated using the parameters such as precision, recall, sensitivity, and accuracy. The accuracy determines the overall correctness of the classifier, where the precision indicates the authenticity of the predicted positives. While the proportion of actual DR cases/ normal that are properly identified as such is called Sensitivity, that of actual non-DR/ abnormal that are properly identified as such is Specificity. The precision, specificity, recall, and

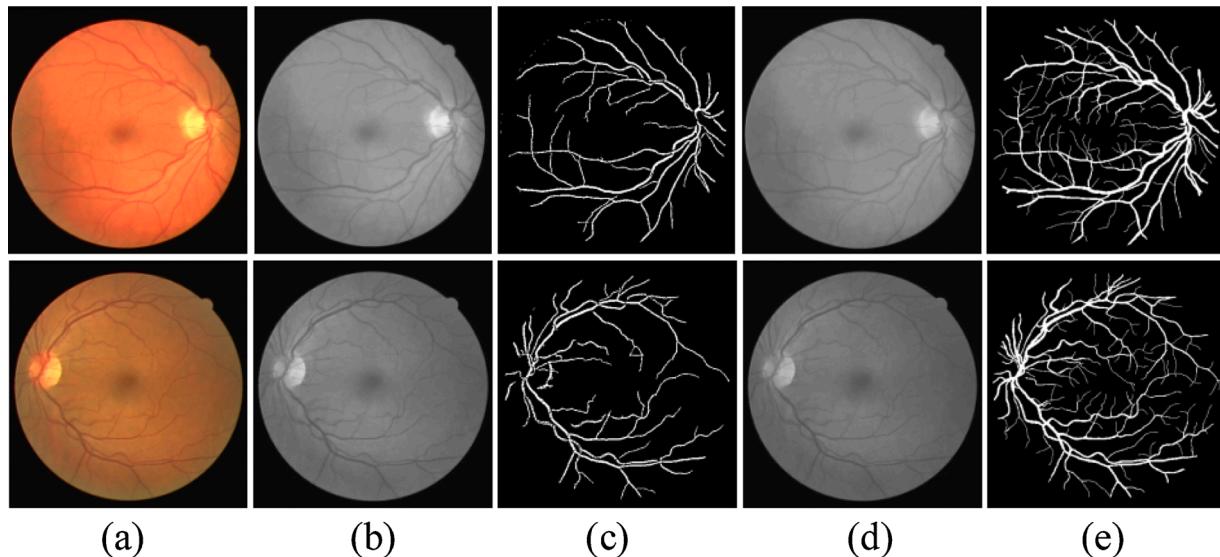


Fig. 9. Some segmented results with different Greyscale conversion (a) Input image, (b) Greyscale converted using Eq. (4),(c) Segmented output with for greyscale conversion using Eq. (4),(d) Greyscale converted using Eq. (5),(e) Segmented output with for greyscale conversion using Eq. (5).

accuracy can be estimated using the relation (12), (13), (14), and (15) respectively as,

$$\text{Precision} = \frac{T_{pos}}{T_{pos} + F_{pos}} \quad (9)$$

$$\text{Specificity} = \frac{T_{neg}}{T_{neg} + F_{pos}} \quad (10)$$

$$\text{Sensitivity} / \text{Recall} = \frac{T_{pos}}{T_{pos} + F_{neg}} \quad (11)$$

$$\text{Accuracy} = \frac{T_{pos} + T_{neg}}{T_{pos} + T_{neg} + F_{pos} + F_{neg}} \quad (12)$$

Here F_{pos} , F_{neg} T_{neg} , and T_{pos} represent false positive, false negative, true negative, and true positive, respectively. The performance of our algorithm with and without data augmentation was compared with the traditional algorithms such as Deep Learning [36], Red Lesions [37], Multi-Sieving [38], Ensemble Approach [39], and circular Hough-CNN [24] using the parameters like precision, specificity, recall and accuracy as shown in Table 1.

4.3. Experimental discussions

The precision, specificity, recall, and accuracy of the proposed method were estimated separately for both the dataset image and the hospital image. The precision, specificity, recall, and accuracy of the proposed algorithm, when tested using DIARETDB1 dataset images, provides a higher value when compared to the hospital images. For the DIARETDB1 dataset, the precision, specificity, recall, and accuracy of the proposed system without data augmentation were estimated to be 96.4 %, 97.2 %, 98.3 %, and 96.92 % respectively. For the same DIARETDB1 dataset the precision, specificity, recall, and accuracy increase to 97.2 %, 98.2 %, 99.6 %, and 98.7 % respectively with data augmentation. For the hospital dataset the precision, specificity, recall, and accuracy is estimated to be 96.1 %, 97.6 %, 97.2 %, and 97.4 % respectively with data augmentation. When comparing the performance of our method with the traditional method using the DIARETDB1 dataset, the proposed method with data augmentation has an improvement of 15.2 % in precision in juxtaposition to the scheme Red lesions [37]. The Specificity of the proposed method improves by 0.5 % when compared with the circular Hough-CNN [24]. The recall and accuracy of the proposed method increase only by 0.004 and 0.17 % respectively when compared to the scheme circular Hough-CNN [24]. Fig. 10 shows the graphical comparison of precision, specificity, recall, and accuracy of the proposed method with the traditional schemes.

To avoid the overfitting in the DIARETDB1 dataset, data augmentation is used that includes flipping such as horizontal and vertical flipping and rotation such as 45° , 90° , 135° , 180° , 225° , 270° and 315° rotation. The usage of this data augmentation method avoids overfitting.

Table 1
Comparison of proposed and traditional methods in terms of performance.

Schemes	Precision (%)	Specificity (%)	Recall	Accuracy (%)
Deep Learning [36]	79	82	0.84	86.1
Red Lesions [37]	82	83.1	0.881	92
Multi-Sieving [38]	81.5	92.1	0.978	96.1
Ensemble Approach [39]	80.3	85.5	0.881	86.9
Circular Hough – CNN [24]	–	97.7	0.992	98.53
Proposed (DIARETDB1) without data augmentation	96.4	97.2	0.983	96.92
Proposed (Hospital) without data augmentation	95.32	96.3	0.965	95.42
Proposed (DIARETDB1) with data augmentation	97.2	98.2	0.996	98.7
Proposed (Hospital) with data augmentation	96.1	97.6	0.972	97.4

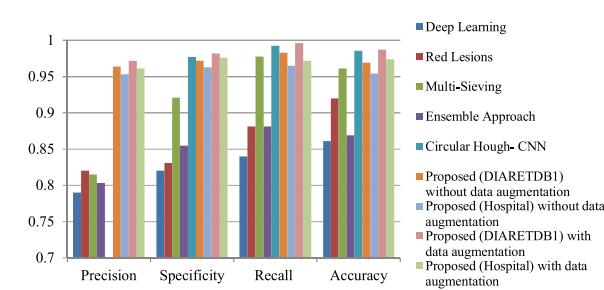


Fig. 10. Performance Comparison of proposed methods with the traditional classification schemes.

In the case of the hospital dataset, both the data augmentation and 20 % dropout of neurons at the hidden layer were used to avoid overfitting. Also, we have used 80 % data for the training set, and 20 % of data for validation to test the overfitting and found that no overfitting occurs with a learning rate of 0.00001 for 20 epochs. The area under the curve is observed from the ROC curve plot obtained from plotting the true positive rate against the false-positive rate. For the proposed methods for the dataset 'DIARETDB1' and hospital images, the ROC curve shows a better performance than the traditional schemes. The hospital images and the images from the DIARETDB1 dataset almost show similar ROC characteristics as shown in Fig. 11. The average classification for the proposed algorithm is estimated as 73 milliseconds.

Table 2 and Fig. 12 show the comparison of Precision, Specificity, and accuracy for each fold with its average value. The precision, specificity, and accuracy obtained on each fold are found to be closer to its average value. Thus, for each fold, there is not much variation in the performance metrics in the k-fold cross-validation.

Accuracy and its variance for DIARETDB1 and hospital dataset are depicted in Table 3 and Fig. 13. When learning 20 epochs, the DIARETDB1 has an accuracy of 98.7 % which is higher than the accuracy of the hospital dataset that has the accuracy of 97.4 %. Especially in the 1st epoch, the DIARETDB1 has an accuracy of 91.3 % and the hospital dataset has an accuracy of 90.2 %. The variance is high at the 11th epoch for the DIARETDB1 dataset where the accuracy is 96.12 %. For the Hospital dataset, the variance is maximum at the 14th epoch, which has an accuracy of 95.23 %. The next section shows the conclusion of the proposed work.

5. Conclusion

This paper proposed a Diabetic retinopathy detection and classification algorithm that uses CNN. Blood clots/blobs, exudates, and abnormal growth of blood vessels can be observed in the fundus of patients suffering from DR. This method initially pre-processes the images and the branching blood vessels are extracted through the segmentation process. Maximum principal curvature, which utilizes the maximum Eigenvalues of the Hessian matrix, has been applied to extract the branching blood vessels post. The adaptive histogram equalization and the morphological opening is performed to enhance and eliminate falsely segmented regions. Further, two sub-networks of neural networks are created excitation and bottleneck, and CNN. If both the sub-networks classify the same label as output, then the image is classified as that label. The experimental results are evaluated using the DIARETDB1 dataset and the images obtained from the hospital. The precision, specificity, recall, and accuracy of the proposed method with data

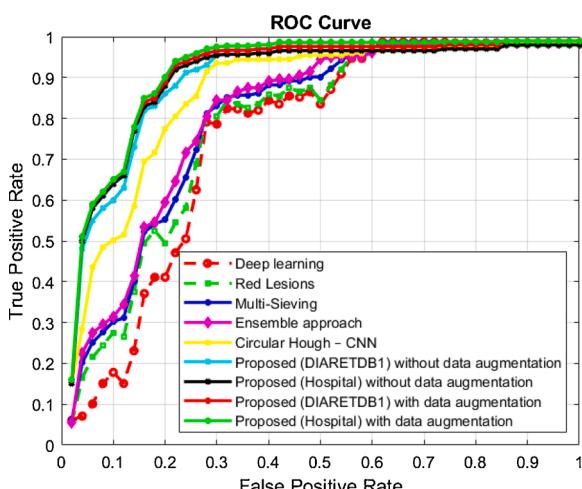


Fig. 11. Comparison of ROC for the proposed and traditional methods.

Table 2

Comparison of Precision, Specificity, and Accuracy on each fold.

Testing fold	Precision	Specificity	Accuracy
1st fold	95.74	98.12	98.54
2nd fold	95.12	96.34	97.98
3rd fold	96.33	97.19	95.86
4th fold	95.98	97.56	96.21
5th fold	97.42	98.77	98.52
Average	96.1	97.6	97.4

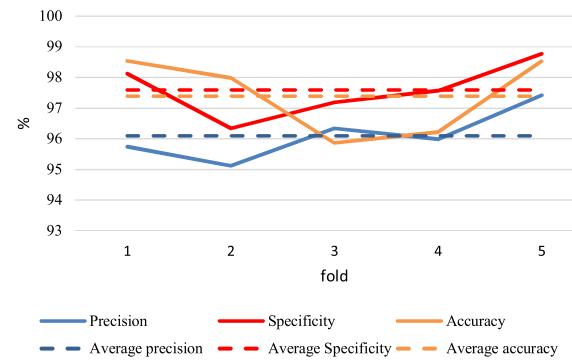


Fig. 12. Comparison of Precision, Specificity, and Accuracy on each fold.

Table 3

Accuracy and dispersion of DIARETDB1 and Hospital dataset.

Epoch	DIARETDB1		Hospital	
	Accuracy	Dispersion	Accuracy	Dispersion
1	91.3	0.12	90.2	0.37
2	91.9	0.25	90.52	0.47
3	92.42	0.52	90.83	0.62
4	92.78	0.93	91.14	0.78
5	93.17	0.54	91.36	0.63
6	93.62	0.92	91.57	0.47
7	93.91	0.14	91.87	0.76
8	94.12	0.31	92.34	0.53
9	94.43	0.76	92.67	0.86
10	94.71	2.31	92.91	0.96
11	96.12	2.62	93.51	1.43
12	97.32	0.82	93.81	1.05
13	97.54	0.92	94.12	2.04
14	97.83	0.41	95.23	2.53
15	97.97	0.42	96.02	0.62
16	98.21	0.41	96.54	0.81
17	98.45	0.63	96.89	0.23
18	98.61	0.75	97.2	0.25
19	98.7	0.43	97.4	0.93
20	98.7	0.43	97.4	0.93

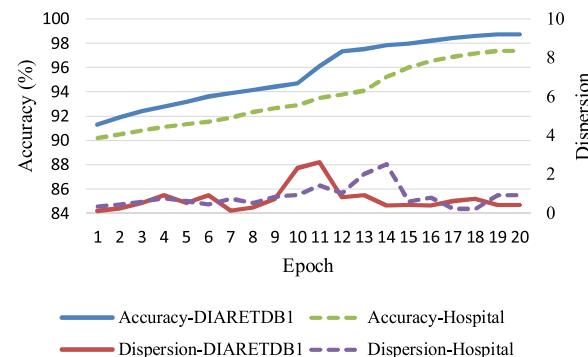


Fig. 13. Comparison of Accuracy and dispersion for DIARETDB1 and Hospital dataset.

augmentation for the DIARETDB1 dataset images was estimated to be 97.2 %, 98.2 %, 99.6 %, and 98.7 % respectively, which is higher than the traditional schemes. In the future, the classification can be further diversified based on the period of suffering. If the image is categorized as ‘Abnormal’ i.e. the patient has been detected with DR in the fundus image, additional processing can be done to classify it, based on how long the patient has been facing the problem. This could help to administer the treatment given to each patient and regulate the dosage of medication.

CRediT authorship contribution statement

Sraddha Das, Krity Kharbanda: Software implementation, Visualization

Suchetha M: Supervision, Conceptualization, Methodology, Data curation, Analysing, Investigation

Rajiv Raman: Investigation, Testing and validation

Edwin Dhas D: Visualization, Writing- Reviewing and Editing

Declaration of Competing Interest

The authors declare no conflict of interest.

References

- [1] N.C. Santosh, Y. Radhika, Optimized maximum principal curvatures based segmentation of blood vessels from retinal images, *Biomed. Res.* (2019), <https://doi.org/10.35841/biomedicalresearch.30-19-068>.
- [2] Gehad Hassan, Nashwa El-Bendary, Aboul Ella Hassani, Aly Fahmy, M. Shoeb, Vaclav Snasel, Retinal blood vessel segmentation approach based on mathematical morphology, *Procedia Comput. Sci.* 65 (2015) 612–622, <https://doi.org/10.1016/j.procs.2015.09.005>.
- [3] J.I. Orlando, E. Prokofyeva, M.B. Blaschko, A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images, *IEEE Trans. Biomed. Eng.* 64 (1) (2017) 16–27, <https://doi.org/10.1109/TBME.2016.2535311>.
- [4] R. Manjula Sri, J. Jyothirmai, D. Swetha, Analysis of retinal blood vessel segmentation in different types of diabetic retinopathy, *Int. J. Eng. Adv. Technol. (IJEAT)* (2019).
- [5] N. Memari, A.R. Ramli, M.I.B. Saripan, et al., Retinal blood vessel segmentation by using matched filtering and fuzzy C-means clustering with integrated level set method for diabetic retinopathy assessment, *J. Med. Biol. Eng.* (2019) 713–731, <https://doi.org/10.1007/s40846-018-0454-2>.
- [6] Attila Budai, Rüdiger Bock, Andreas Maier, Joachim Hornegger, Georg Michelson, Robust vessel segmentation in fundus images, *Int. J. Biomed. Imaging* (2013), 154860, <https://doi.org/10.1155/2013/154860>.
- [7] Z. Tian, L. Liu, Z. Zhang, B. Fei, Superpixel-based segmentation for 3D prostate MR images, *IEEE Trans. Med. Imaging* 35 (3) (2015) 791–801.
- [8] D.C.T. Nguyen, S. Benamou, M. Mignotte, F. Lavoie, Superpixel and multi-atlas based fusion entropic model for the segmentation of X-ray images, *Med. Image Anal.* 48 (2018) 58–74.
- [9] Y.L. Huang, D.R. Chen, Watershed segmentation for breast tumor in 2-D sonography, *Ultrasound Med. Biol.* 30 (5) (2004) 625–632.
- [10] H. Masoumi, A. Behrad, M.A. Pourmina, A. Roosta, Automatic liver segmentation in MRI images using an iterative watershed algorithm and artificial neural network, *Biomed. Signal Process. Control* 7 (5) (2012) 429–437.
- [11] M. Ciechlewski, An edge-based active contour model using an inflation/deflation force with a damping coefficient, *Expert Syst. Appl.* 44 (2016) 22–36.
- [12] M. Ciechlewski, J.H. Spodnik, Semi-automatic corpus callosum segmentation and 3d visualization using active contour methods, *Symmetry* 10 (11) (2018) 589.
- [13] D. Lokuarachchi, K. Gunarathna, L. Muthumal, T. Gamage, Automated detection of exudates in retinal images, in: *IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)*, Penang, Malaysia, 2019, pp. 43–47, <https://doi.org/10.1109/CSPA.2019.8696052>.
- [14] R.J. Chalakkal, W.H. Abdulla, S.S. Thulaseedharan, Automatic detection and segmentation of optic disc and fovea in retinal images, *IET Image Process.* 12 (11) (2018) 2100–2110, <https://doi.org/10.1049/iet-ipr.2018.5666>, 11.
- [15] S. Selvaperumal, Ramasubramanian Bhoopalan, An efficient approach for the automatic detection of hemorrhages in colour retinal images, *IET Image Process.* (2018) 12, <https://doi.org/10.1049/iet-ipr.2017.1036>.
- [16] A. Benzamin, C. Chakraborty, Detection of Hard exudates in retinal fundus images using deep learning, in: 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, 2018, pp. 465–469, <https://doi.org/10.1109/ICIEV.2018.8641016>.
- [17] S. Yu, D. Xiao, Y. Kanagasingam, 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), Seogwipo, 2017, pp. 1744–1747, <https://doi.org/10.1109/EMBC.2017.8037180>.
- [18] Shailesh Kumar, Abhinav Adarsh, Basant Kumar, Amit Singh, An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation, *Opt. Laser Technol.* (2020), <https://doi.org/10.1016/j.optlastec.2019.105815>.
- [19] N. Yalçın, S. Alver, N. Uluhatun, Classification of retinal images with deep learning for early detection of diabetic retinopathy disease, in: 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1–4, <https://doi.org/10.1109/SIU.2018.8404369>.
- [20] Payal Shah, Divyansh K. Mishra, Mahesh P. Shanmugam, Bindiya Doshi, Hariprasad Jayaraj, Rajesh Ramanjulu, Validation of Deep Convolutional Neural Network-based algorithm for detection of diabetic retinopathy – artificial intelligence versus clinician for screening, *Indian J. Ophthalmol.* 68 (2020) 398, https://doi.org/10.4103/ijo.IJO_966_19.
- [21] D.Y. Carson Lam, M. Guo, T. Lindsey, Automated detection of diabetic retinopathy using deep learning, *AMIA Summits Transl. Sci. Proc.* 2018 (2018) 147 [PMC free article].
- [22] M. Andonová, J. Pavlovičová, S. Kajan, M. Oravec, V. Kurilová, Diabetic Retinopathy Screening Based on CNN, *International Symposium ELMAR*, Zadar, 2017, pp. 51–54, <https://doi.org/10.23919/ELMAR.2017.8124433>.
- [23] P. Prentasić, S. Lončarić, Detection of exudates in fundus photographs using convolutional neural networks, in: 9th International Symposium on Image and Signal Processing and Analysis (ISPA), Zagreb, 2015, pp. 188–192, <https://doi.org/10.1109/ISPA.2015.7306056>.
- [24] Kemal Adem, Exudate detection for diabetic retinopathy with circular hough transformation and convolutional neural networks, *Expert Syst. Appl.* (2018) 114, <https://doi.org/10.1016/j.eswa.2018.07.053>.
- [25] K. Lyu, Y. Li, Z. Zhang, Attention-aware multi-task convolutional neural networks, *IEEE Trans. Image Process.* 29 (2020) 1867–1878, <https://doi.org/10.1109/TIP.2019.2944522>.
- [26] Thomas A. Siji, Titus Geeverghese, Design of a portable retinal imaging module with automatic abnormality detection, *Biomed. Signal Process. Control* 60 (2020), <https://doi.org/10.1016/j.bspc.2020.101962>.
- [27] D.J. Hemanth, O. Deperioglu, U. Kose, An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network, *Neural Comput. Appl.* 32 (2020) 707–721, <https://doi.org/10.1007/s00521-018-03974-0>.
- [28] S. Yang, C. Yang, D. Feng, X. Hao, M. Wang, One-dimensional deep attention convolution network (ODACN) for signals classification, *IEEE Access* 8 (2020) 2804–2812, <https://doi.org/10.1109/ACCESS.2019.2958131>.
- [29] W. Hu, H. Li, L. Pan, W. Li, R. Tao, Q. Du, Spatial-Spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification, *IEEE Trans. Geosci. Remote. Sens.* 58 (June 6) (2020) 4237–4250, <https://doi.org/10.1109/TGRS.2019.2961947>.
- [30] Qi Zhao, Jiahui Liu, Boxue Zhang, Shuchang Lyu, Nauman Raoof, Wenquan Feng, Interpretable Relative squeezing bottleneck design for compact convolutional neural networks model, *Image Vis. Comput.* 89 (2019), <https://doi.org/10.1016/j.imavis.2019.06.006>.
- [31] Jun Gu, Xuewen Sun, Yue Zhang, Kun Fu, Lei Wang, Deep residual squeeze and excitation network for remote sensing image super-resolution, *Remote Sens.* 11 (1817) 2019, <https://doi.org/10.3390/rs11151817>.
- [32] S. Lahmiri, Hybrid deep learning convolutional neural networks and optimal nonlinear support vector machine to detect presence of hemorrhage in retina, *Biomed. Signal Process. Control* 60 (2020), <https://doi.org/10.1016/j.bspc.2020.101978>.
- [33] L. Qiao, Y. Zhu, H. Zhou, Diabetic retinopathy detection using prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms, *IEEE Access* (2020), <https://doi.org/10.1109/ACCESS.2020.2993937>.
- [34] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, H. Kang, Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening, *Inform. Sci.* 501 (2019) 511–522.
- [35] W.L. Alyoubi, W.M. Shalash, M.F. Abulkhair, Diabetic retinopathy detection through deep learning techniques: a review, *Inform. Med. Unlocked* (2020), 100377.
- [36] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, J. Pietilä, DIARETDB1 diabetic retinopathy database and evaluation protocol, *Technical report (PDF)* (2021).
- [37] K.M. Adal, P.G. Van Etten, J.P. Martinez, K.W. Rouwen, K.A. Vermeer, L.J. van Vliet, An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images, *IEEE Trans. Biomed. Eng.* 65 (6) (2017) 1382–1390, <https://doi.org/10.1109/TBME.2017.2752701>.
- [38] L. Dai, R. Fang, H. Li, X. Hou, B. Sheng, Q. Wu, W. Jia, Clinical report guided retinal microaneurysm detection with multi-sieving deep learning, *IEEE Trans. Med. Imaging* 37 (5) (2018) 1149–1161, <https://doi.org/10.1109/TMI.2018.2794988>.
- [39] S. Qummar, F.G. Khan, S. Shah, A. Khan, S. Shamshirband, Z.U. Rehman, I.A. Khan, W. Jadoon, A deep learning ensemble approach for diabetic retinopathy detection, *IEEE Access* 7 (2019) 150530–150539, <https://doi.org/10.1109/ACCESS.2019.2947484>.



Contents lists available at ScienceDirect



Optik

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

Original research article

Retinal blood vessel segmentation from diabetic retinopathy images using tandem PCNN model and deep learning based SVM



T. Jemima Jebaseeli^{a,*}, C. Anand Deva Durai^b, J. Dinesh Peter^a

^a Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore 641114, Tamilnadu, India

^b Department of Computer Science and Engineering, King Khalid University, Abha 61421, Saudi Arabia

ARTICLE INFO

Keywords:

Image segmentation
Blood vessel
Diabetic retinopathy
Neural network
Support vector machine
Fundus image
Feature extraction
Deep learning

ABSTRACT

Diabetic Retinopathy (DR) occurs due to Type-II diabetes. It causes damages to the retinal blood vessels and reason for visual impairment. The predicted center is around the probability of variation in the estimation of retinal veins, and the crisp enrolls vessel development inside the retina. To witness the changes segmentation of retinal blood vessels has to be made. A framework to upgrade the quality of the segmentation results over morbid retinal images is proposed. This framework utilizes Contrast Limited Adaptive Histogram Equalization (CLAHE) for eliminating the background from the source image and enhances the foreground blood vessel pixels. Tandem Pulse Coupled Neural Network (TPCNN) model is endorsed for automatic feature vectors generation, and Deep Learning Based Support Vector Machine (DLBSVM) is proposed for classification and extraction of blood vessels. The DLBSVM parameters are fine-tuned via Firefly algorithm. The STARE, DRIVE, HRF, REVIEW, and DRIONS fundus image datasets are deliberated to assess the recommended techniques. The results render that the proposed technologies improve the segmentation with 80.61% Sensitivity, 99.54% Specificity, and 99.49% Accuracy.

1. Introduction

The retina is a reason for vision; it observes the light and transforms as brain signals to produce the vision. Diabetic Retinopathy (DR) made the outer retinal layer thick. The outer retinal layer is damaged and settles dots and spots around the retina. The lipids and proteins inside the retinal layers are developed as a tiny blot of exudates, microaneurysms, and cotton wool spots [1]. At the advanced stage, the Proliferative Diabetic Retinopathy (PDR) leads to the tractional detachment of retina [2,3]. There are various segmentation procedures recommended by various researchers. But, these methods work only on fundus images without pathological effects [4]; till there are difficulties in segmenting the vascular vessel tree map without any discontinuities.

2. Reviews of related studies

Retinal vessel segmentation is an imperative instrument for revealing the deviations that happen in veins and it provides data around the territory of vessels. The precise segmentation of retinal blood vessel is clinically significant [5]. The screening programs have been launched in many countries. The computerized system considerably eases the load of experts. Ana Salazar Gonzalez et al. [6], employed a graph cut technique. The prior learning process is incorporated to obtain the optimal segmentation. But, the methods are sensitive to noise thus, training datasets would need further improvement. Shahab Aslani et al. [7], adopted a hybrid feature

* Corresponding author.

E-mail address: jemima_jeba@karunya.edu (T.J. Jebaseeli).

vector. Here, 13 Gabor features are used. If these features are reduced, this may result in detrition of accuracy. George Azzopardi et al. [8] designed a BCOSFIRE filter. Its parameters influence the performance of the filter. Also, there are misclassifications in the reproduced width of the minor vessel. This tends to increase the false positives computations. S. Wilfred Franklin et al. [9] have proposed the Multilayer Perceptron Neural Network (MPNN) to segment the retinal vessels. Here, the weight of the feed forward network is changed by the backpropagation algorithm. Since it is a kind of pixel processing based approach, it has a less amount of accuracy of 95.03%. R. Geetha Ramani et al. [10] introduced the Principle Component Analysis (PCA) method for feature vector generation. At the final stage, mathematical morphological operation [11] and Connected Component Analysis (CCA) is performed. The classification of vessel cluster decreases the sensitivity measure in some cases. M.R.K. Mookiah et al. [12] have proposed the system with 13 substantial features for the Probabilistic Neural Network (PNN) and SVM [13,14]. The algorithm depends on training inputs. This is not suitable for real-time clinical images. Shuangling Wang et al. [15] demonstrated two superior classifiers: CNN [16] is used as feature extractor, and ensemble RF is the trainable classifier. The execution performance of the learning algorithms to a significant degree relies upon the training data [17]. In 2006, Geoff Hinton initially proposed the idea of Deep Learning [18,19]. For instance, Multi-layer Kernel Machines (MKM) [20,21], utilizes the arc-cosine kernel; however, the kernel determination is having a testing issue, which incredibly influences the nature of data representation. The D-SVM [22] is having a dispute to choose the parametric-type such as kernel function, SVMs, and feature's weights are yet unanswered. Moreover, Multilayer Support Vector Machine (MLSVM) [23,24], is trained using gradient-based learning on a min-max plan for optimization. The structure of MLSVM is like a neural network but differs from the deep learning structure. As an outcome, while experiencing numerous parameters of every layer, this strategy is inclined to fall into a local optimum.

2.1. Challenges

The blood vessel segmentation methods [25–27], prevalently utilizes vector geometry, statistical data, image filters, and machine learning techniques to generate the low-level feature vectors to detect the vessel. These methods depend on the utilization of high-quality features or heuristic presumptions to solve the issue. These methods do not use the generalized learning pattern to generate the feature vectors. Hence, it is vulnerable to subjectivity and its strategic nature causes shortcomings. The primary challenge is to design an end-to-end framework which learns from the data with no space learning based heuristic data to recognize both coarse and fine vascular structures. There are several approaches suggested for enhancing PCNN parameters to attain virtuous outcomes. There are no methods to spontaneously determine all the parameters of PCNN to achieve the clear-cut image segmentation [28,29]. While considering the earlier approaches, several algorithms are useful to extort the features from unfussy images and receptive to noise. Also, many algorithms which are incompetent in recognizing the blood vessels from pathological depigmented images.

2.2. Motivations

The unconstrained retinal map is used for the macular weakening treatment. Regardless of a substantial number of inquiries on the extraction of retinal veins, there is a requirement for exact retinal blood vessel extraction framework.

- i There are a few impediments that need change, including deceitful vessel extraction which is caused due to low segmentation between the retinal vessels and the background. In like way, the impairment of the blood vessel tree whose topological construction is complex.
- ii There is a couple of disenchantment that must be redesigned among the retinal veins and the non-vessel image.
- iii The organize network misfortune amidst retinal vessels makes a complex topological structure.
- iv In some method, it is inadequate to demonstrate the whole segmentation result if there ought to emerge an event of veins that obscure a long way from the mid to its broadened heading.
- v Losing one branch point may prompt a lacking vein system and raises the error. Grounded on this motivation, the proposed TPCNN model is competently deciding the paramount alternative of segments similar to the intended object. Subsequently, it is proficient enough to observe the little veins without any discontinuities.

2.3. Novelty of the proposed techniques

The Tandem PCNN (TPCNN) model extract the blood vessels from the obsessive fundus images and discloses all tiny vessels at their cross boundaries which are suppressed by depigmentation. Deep Learning Based SVM (DLBSVM) improves the classification accuracy in categorizing the vessel pixels from non-vessel pixels. Compared with the conventional methods the proposed methods have the following novelties:

- i The scheming of pulse threshold to fire the neuron is computed dynamically.
- ii The linking weight redefines the strength of the active neuron.
- iii The proposed TPCNN model extracts the blood vessels within four iterations.
- iv The cross-channel linking of TPCNN model enables the visibility of vessels at the crossover points.
- v Classification and extraction of retinal blood vessels via DLBSVM help the network to attain better data expression to increase the final classification result.
- vi Firefly algorithm optimizes the parameters of TPCNN.

Table 1
Fundus Image Databases.

Database	No. of Images	Image size (pixels)
DRIVE	40	584 × 584
STARE	20	700 × 605
REVIEW	16	1360 × 1024
HRF	15	3504 × 2336
DRIONS	110	600 × 400

3. Materials

The fundus images chosen for validation of the proposed approaches are taken from the public datasets such as STARE, DRIVE, HRF, REVIEW, and DRIONS databases [30]. Table 1 describes the number of fundus images taken for the proposed model's experimentation.

The following are the test set and training set images available in the public databases used by various researchers for their experimental setup.

- i DRIVE database has 20 test images and 20 training images.
- ii STARE database contains 20 images with two sets of manual segmentation provided by two experts. Here, there is no division into training and testing images.
- iii REVIEW database encompasses 16 test images.
- iv HRF database comprises 15 test images.
- v DRIONS database encloses 50 test images and 60 training images.

We have used all the above 201 fundus images to assess the performance of the proposed techniques.

4. Methods

The architecture of the proposed segmentation method is shown in Fig. 2. The blood vessels are extracted by the subsequent ways. At first, the image enhancement and noise removal are done through CLAHE. Then, the feature vectors are produced by Tandem PCNN model. The retinal blood vessels are classified using DLBSVM and segmented using Tandem PCNN's feature element vectors. Firefly algorithm generates values to optimize the parameters of DLBSVM.

4.1. Preprocessing of fundus images

The false photographic artifacts and illumination inconsistencies in the fundus image are preprocessed. The intensity of the green channel has more vessel pixels compared to its background [1,13]. Therefore the green channel is considered and CLAHE preprocesses it. The CLAHE picks the clipping value of the histogram and computes the local histogram amplification of the image. The tile estimate is fixed spontaneously dependent on the vascular geometry of the retinal image. The preprocessed fundus image is shown in Fig. 1.

4.2. Blood vessel segmentation - Tandem PCNN (TPCNN) model

The simple PCNN model is inadequate to complete the segmentation procedure. In multispectral PCNN (m-PCNN), m-channels are fused to create the outcomes. The resultant image provides more details, but it doubles the execution time. Hence, there is a need of

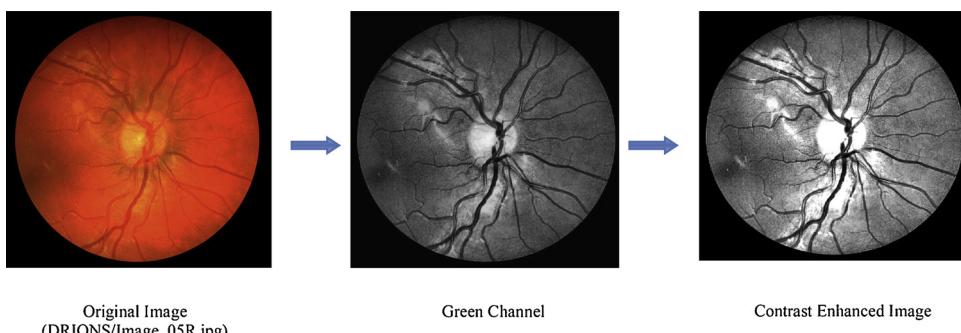


Fig. 1. Preprocessing results of the fundus image (DRIONS).

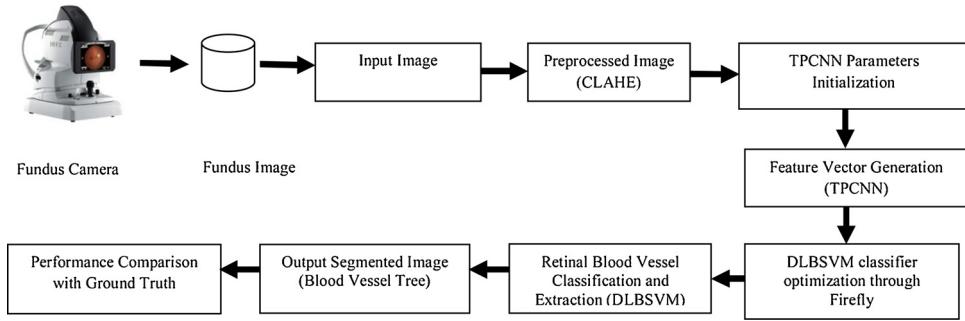


Fig. 2. The recommended frame work for the retinal blood vessel segmentation.

Tandem PCNN model, which fuses the data from two source input images.

Tandem PCNN (TPCNN) model is shown in Fig. 3 which triggers the inter and intra channel linking of the input neurons. The main channel creates an auto wave which stimulates to generate auto waves continuously. Each auto wave generation renowned the edges of the retinal vessels and made it detectable. The cross-channel linking of this model authorizes the visibility of vessels at the crossover points. This model need not bother about the training of the target information. In Tandem PCNN model, there are two input channels such as f^a and f^b are considered here, instead of feeding input and linking the input field in the basic PCNN model.

$$f^a(n) = s^a + m(y(n-1)) \quad (1)$$

$$f^b(n) = s^b + w(y(n-1)) \quad (2)$$

where s^a and s^b are input images respectively. m and w are the feeding functions which decide the encompassing neurons.

$$m(\cdot) = y(n-1) \otimes k \quad (3)$$

Where n is the iteration count. $y(n-1)$ is the yield neurons created in the midst of the past cycle. k is the array of convolution core values.

The inertia weight w is reduced to the minimum value in the diverse iteration. The adjustment of inertia weight plays an imperative role in improving TPCNN model. In the case of fixed inertia weight, the method traps into local optima. The linear decrease in weight leads to missing optimal point value in the final segmentation. To solve this issue, the inertia weight is redefined as follows,

$$w = w_{\max} - \frac{t(w_{\max} - w_{\min})}{t_{\max}} \quad (4)$$

$$w = \begin{cases} w_{\min} - \frac{(w_{\max} - w_{\min}) \times (f^a(n), f^b(n))_t - (f^a(n), f^b(n))_{\min}}{(f^a(n), f^b(n))_{\text{avg}}}, & (f^a(n), f^b(n))_t \leq (f^a(n), f^b(n))_{\text{avg}} \\ w_{\max}, & (f^a(n), f^b(n))_t > (f^a(n), f^b(n))_{\text{avg}} \end{cases} \quad (5)$$

where t is the recent iteration, w_{\max} is the maximum weight, and w_{\min} is the minimum weight.

When the fitness estimation of the entire particles had a tendency to join together expands the inertia inactivity weight. At the point when the fitness estimation of all particles distributed, the inertia weight esteem is decreased. Since the inertia weight esteem is logically altered close by the fitness value. The dynamic inertia weight has the accompanying favorable circumstances for those

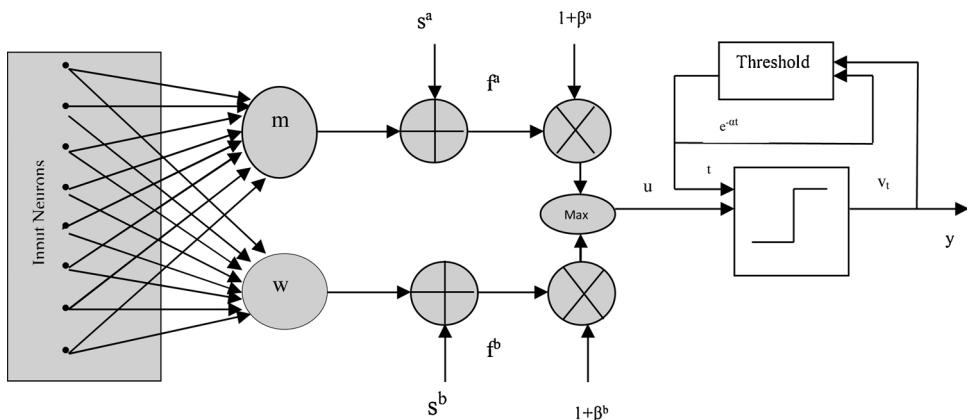


Fig. 3. Tandem Pulse Coupled Neural Network (TPCNN) model.

incredible particles whose fitness esteem are more significant than the normal fitness value; they will normally get the littler weight regard. That suggests they will diminish their speed and be ensured. For those particles whose fitness esteems are more diminutive than the normal fitness esteem, they will normally obtain the more prominent inactivity weight regard. That infers will extend their speed and quickly move closer to the phenomenal particles. The receptive field gets the information and upgrades the enveloping neurons. u is the connecting field, where all data are combined to find the state of the neuron to be terminated is processed as,

$$u(n) = \max((1 + \beta^a f^a(n))(1 + \beta^b f^b(n))) + \delta \quad (6)$$

The weighing coefficients of f^a and f^b are β^a and β^b . It is expected to decide the quality of the linking neuron. δ is the smooth factor to screen the inner activity of the neuron. To consequently determine the estimation of β , the background and the object intensity $[i_1, i_2]$ and $[i_3, i_4]$ are analyzed. The smallest linking input of the neuron intensity (β_{\max}) of i_1 and i_3 are captured by the firing neuron $i_1 < i_3 < i_2$.

$$\beta_{\min} = \max((i_4/i_3 - 1)/f^a(n), (i_2/i_1 - 1)/f^b(n)) \quad (7)$$

The most extreme linking input of the neuron intensity (β_{\max}) i_2 is caught by the firing neuron is,

$$\beta_{\max} = (i_4/i_2 - 1)/\max(f^a(n), f^b(n)) \quad (8)$$

To guarantee all the blood vessels captured by the neighboring neurons are computed using the iterations of Tandem PCNN model is,

$$\beta(n) = i_4(n)/i_2(n - 1)/\max(f^a(n), f^b(n)) \quad (9)$$

Where i_4 is the upper bound and i_2 is the lower bound intensity of fired neuron linking input at the TPCNN iteration process. In dissimilarity to the static value of β , the conceivable object neurons with spatial proximity is given most priority to fire and creates the reasonable outcome. The pulse generator $y(n)$ generates the output pulse.

$$y(n) = \begin{cases} 1 & u(n) > t(n) \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The dynamic threshold $t(n)$ is expressed by the following equation.

$$t(n) = \begin{cases} e^{-\alpha t}(n - 1) & y(n) = 0 \\ v_t & \text{otherwise} \end{cases} \quad (11)$$

v_t and αt are the normalized offset and time constants. The receptive field gathers the encompassing neurons as a contribution to deciding the feeding and linking field values. The neurons of two channels get the external stimuli from the surrounding neurons. At that point, these channels are weighed and the information is fused according to the weighing coefficients. Then the dynamic neuron will be fired based on the attenuation coefficients. The generated outputs are given to the pulse generator to yield the output pulse.

4.2.1. Algorithm of TPCNN

The algorithm for Tandem PCNN model is as follows.

1. Input: s^a, s^b

2. Let $u = 0, y = 0, t = 1$

Compute w and m from empirical metrics.

3. If $s^a = s^b$ then set $o \leftarrow s^a || s^b$

goto step 7.

4. Set $e^{-\alpha t} \leftarrow [0, 1]$

5. Identify the Surrounding Neurons

$sn \leftarrow y[n - 1] \otimes k$

$u(n) \leftarrow \max((1 + \beta^a f^a(n))(1 + \beta^b f^b(n))) + \delta$

If $u(n) > t(n)$ then $y(n) \leftarrow u(n) - sn$

Else $y(n) \leftarrow 0$

End

If $s^a = s^b || \beta^a = \beta^b$ then $o \leftarrow s^a || s^b$

Else $o \leftarrow y$

End

If $y = 0$ then $t \leftarrow e^{-\alpha t}$

Else $t \leftarrow v_t$

End

6. Stop when all neurons are fired goto Step 7. Else goto Step 5.

7. Output: ois the output of Tandem model PCNN.

4.2.2. Parameter setting

In Tandem PCNN model, the parameters are initialized as follows:

- 1 Convolution core K = [0.5, 1, 0.5; 0.5, 1, 0.5; 0.5, 1, 0.5]
- 2 Smooth factor $\sigma = 1.2$
- 3 Time constant $\alpha t = 0.2$
- 4 Standardized offset parameter $v_t = 220$.

4.3. Retinal blood vessel classification - deep learning based SVM (DLBSVM) model

The training data commencing from the Tandem PCNN model of the form $\{(x_1, y_1), \dots, (x_n, y_n)\}$; such that, x_i is the feature vector and y_i is the classification label of the image. DLBSVM has the function $G: X \rightarrow Y$, where X is the input image space and Y is the yield feature space. R is the function with the possible element in the hypothesis space. $x_i \in R^N$, $y_i \in Y = \{1, -1\}$ and $i = 1 \dots n$.

4.3.1. Algorithm of DLBSVM

The algorithm of DLBSVM is stated as follows,

Step 1: Input the training data as follows,

$$T = \{I_1(x, y), \dots, I_n(x_n, y_n)\} \in (R^N \times y)^n$$

Step 2: Initialize the value of similarity measure between the two points: (σ) trades of the misclassification of training against the decision surface, and (c) value of Radial Basis Function parameters in DLBSVM as follows.

$$\sigma = [2^{-8}, \dots, 2^8] \text{ & } c = [2^{-8}, \dots, 2^8].$$

Step 3: Iteration loop

for k = 1...K

1. *Carryout Tandem PCNN*

2. *Construct new Training data*

$$T = \{I(x_1^{k+1}, y_1), \dots, I(x_n^{k+1}, y_n)\} \in (R^N \times y)^n$$

end

Step 4: The final classifier

The classifier depends on the inner product of the test point x and the support vector x_i^{k+1} . The final solution is to calculate the inner products $x_i^{k+1} \cdot x_i$ between all pairs of training points. The DLBSVM classification algorithm has gained in popularity due to its high-performance accuracy.

4.4. Firefly algorithm

The Firefly algorithm optimizes the DLBSVM final classification process. The search technique is established on the social conduct of firefly's communication. The brightness determines the firefly's attractiveness. Hence, the attractiveness between x and x_i^{k+1} is proportional to the brightness of a firefly. This is given as an objective function $F(x)$. Hence, $I(x_i) \propto F(x_i)$, where x_i is the member of the firefly swarm. $I(x_i)$ is the brightness of the corresponding fireflies. The attractiveness between fireflies x_i and x_j^{k+1} are defined by their distance vector $dist_{ij}$.

$$dist_{ij} = \frac{\|x_i - x_j^{k+1}\|}{\sqrt{\sum_{l=1}^n (x_{il} - x_{jl}^{k+1})^2}} \quad (12)$$

where l is the index of candidate fireflies solutions.

The following equation defines the total sum of initial brightness α_i .

$$\alpha = \alpha_i e^{-\varphi dist_{ij}} \quad (13)$$

φ is the light absorption coefficient.

The movement of x_i to x_i^{k+1} during attractiveness is defined as follows,

$$x_{il} = (1 - \alpha)x_{il} + \alpha x_{jl}^{k+1} + u_{il} \quad (14)$$

$$u_{il} = rand_1 - \frac{1}{2} \quad (15)$$

If there is no brighter firefly than the specific fireflies x_{max_i} , then it move arbitrarily according to the following given equation.

$$x_{max_il} = x_{max_il} + u_{max_il} \quad (16)$$

$$u_{max_il} = rand_2 - \frac{1}{2} \quad (17)$$

$rand_1$ and $rand_2$ are obtained from the $U(0, 1)$ uniform distribution.

The Firefly - DLBSVM may converge with the most optimal solution within a limited time when it associates with the feature selection because of its complexity.

Table 2
Performance Measures.

$$\begin{aligned} \text{Sensitivity} &= \frac{TP}{TP + FN} \\ \text{Specificity} &= \frac{TN}{TN + FP} \\ \text{Accuracy} &= \frac{TP + TN}{(TP + FN) + (TN + FP)} \\ TPR &= \text{Sensitivity} \\ FPR &= 1 - \text{Specificity} \end{aligned}$$

5. Performance analysis

The segmented output image contains the blood vessels. Likewise, the ground truth images contain the blood vessels which are classified manually by the specialists. The vessel pixel values of the ground truth images are compared with the segmented output image to discover the worth of the algorithms in predicting the vessel pixels. The proposed method ensures not to flop in classifying the tiny blood vessels which are created by the diseased particles. The proposed Tandem PCNN model segments the blood vessels within four iterations.

5.1. Metrics – I

The performances of the proposed algorithms are evaluated by the parameters such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP is a pixel marked as vessel pixel in both ground truth image and in segmented output image; TN is a pixel marked as non-vessel pixel in both ground truth image and in the segmented output image; FP value is a pixel marked as vessel pixel in the segmented output image but as a non-vessel in the ground truth image; and FN is a pixel marked as non-vessel in the segmented output image but as a vessel in the ground truth image.

As shown in [Table 2](#),

- i Sensitivity denotes the algorithm's capability to detect the blood vessels correctly.
- ii Specificity denotes the algorithm's ability to detect the non-blood vessels correctly.
- iii Accuracy is the ratio of the total number of correctly classified pixels to the number of pixels in the image field of view.
- iv True Positive Rate (TPR) denotes the fraction of pixels correctly detected as vessel pixels.
- v False Positive Rate (FPR) indicates the fraction of pixels incorrectly identified as vessel pixels.

The proposed Tandem PCNN model has enormous performance excellence over the DRIVE database with Sensitivity (80.27%), Specificity (99.80%), and Accuracy (98.98%); STARE database with Sensitivity (80.60%), Specificity (99.70%), and Accuracy (99.70%); REVIEW database with Sensitivity (80.88%), Specificity (98.76%), and Accuracy (99.87%), HRF database with Sensitivity (80.77%), Specificity (99.66%), and Accuracy (98.96%); DRIONS database with Sensitivity (80.54%), Specificity (99.78%), and Accuracy (99.94%) on normal and pathological images as shown in [Table 3](#).

5.2. Metrics – II

The computation time exploration is the significant impediments of the supervised vessel segmentation methods, and the proposed technique is not a special case. The proposed strategy is estimated in a PC with Intel(R) Core(TM) i3-5005U CPU@2.00 GHz processor. As appeared in [Table 4](#), the computational time is less for the proposed method while comparing with other competitive methods. The duration for handling a distinct image is 0.6 s for DRIVE database, and 0.8 s for STARE database. The exhibited effectiveness, robustness, smoothness, and fast execution make the proposed computerized blood vessel segmentation method as a flexible instrument for initial DR detection.

Table 3
Performance outcomes of the proposed method.

Database	Sensitivity%	Specificity%	Accuracy%
DRIVE	80.27	99.80	98.98
STARE	80.60	99.70	99.70
REVIEW	80.88	98.76	99.87
HRF	80.77	99.66	98.96
DRIONS	80.54	99.78	99.94
Average Value	80.61	99.54	99.49

Table 4
Computation time per image.

Methods	DRIVE	STARE
Qiaoliang Li et al. [1]	1.2 min	–
Shahab Aslani et al. [7]	60s	60s
Nagendra Pratap Singh et al. [22]	2.26min	2.4min
A. Colomer et al. [29]	297.07 s	303.64 s
Zhexin Jiang et al. [30]	1.67 s	1.67 s
Chengzhang Zhu et al. [31]	12.16 s	–
Dinesh Pandey et al. [32]	7s	8s
George Azzopardi et al. [2]	10s	10s
Luiz Carlos Rodrigues et al. [33]	35s	–
Toufique Ahmed Soomro et al. [34]	4.5s	4.5s
Sudeshna Sil Kar et al. [35]	2.25s	2.25s
Proposed Method	0.6 s	0.8 s

6. Experimental results and discussion

The proposed system is instigated and tested using MATLAB R2010a and verified over STARE, DRIVE, HRF, REVIEW, and DRIONS databases. The segmented vessel map of the fundus images and the relating ground truth image are shown in Figs. 4 and 5. The prevailed vascular tree pattern of input fundus images are checked with the ground truth fundus images and shows more accuracy while comparing with other compatible methods.

Table 5 provides the performance of diverse approaches proposed in the literature in relations of the typical incoherency measurements. Subjectively, the proposed approach segments the maximum vessels without holding any background noise. Quantitatively, this methodology accomplishes highest specificity and accuracy but the lowest sensitivity. The overall performance of this proposed method achieves the average sensitivity of 80.61%, specificity of 99.54%, and accuracy of 99.49%. The depigmentations are expelled from the vessel points proactively by the Tandem PCNN model which made the system to follow every small vessel unseen by this depigmentation. The configuration of BCOSFIRE filter has to be attuned for dissimilar shapes, such as vessels, bifurcations and crossover points at diverse scales [8]. There are inaccuracies in the reconstructed width of the tiny vessel. This offers a reduction in the true positive counts. There is a major struggle in relinking vessel segments that are situated far from the intersections. The classifier has been trained by choosing 300,000 samples randomly from the DRIVE dataset [1]. Additionally, the proposed model automatically generates its feature vectors by DLBSVM through Tandem PCNN model. Tandem PCNN is an unsupervised model, where its parameters must be set ahead of time to create the feature vectors consequently. The PCNN parameters of 2D Otsu [28], takes $\alpha_0 = 0.2$, $V_L = 1$, $V_0 = 10\,000$, and $\beta = 0.1$. Compared to 2D Otsu technique, the proposed Tandem PCNN model parameters takes less value; also the iteration count for segmenting the image is less. In the proposed model, the Tandem PCNN parameter contains less parameter, and the template will be generated only one time, but it is suitable for all images. Based on the experimentation, these parameters are optimally tuned. Tandem PCNN model is improved than CNN, which is confirmed in the proposed method by combining the virtues of feature learning and traditional classifier. In SVM, there are opportunities to recognize the false vessel pixels; while, DLBSVM significantly identifying the minute vessel pixels from its boundary. CNN neglects to find some tiny vessels around the optic disc. The accuracy of Random Forest is depended upon every classifier and their correlations. Each classifier has less value of correspondences. This is suitable for classification and deterioration on images with incredible noises [15]. Due to the disparity in the image intensity between vessel and background, the optimum blood vessels are sometimes neglected because of the nonappearance of the edges [13].

There are various web-based retinal image segmentation tools such as ROPTOOL, VAMPIRE, SIVA, and LIVE VESSEL [31] accessible for Diabetic Retinopathy diagnosis. The android systems [31,33], developed for small datasets could analyze the information within the restricted area nearby the optic disc. Additionally, the processing time is more for a single image. Clinically this solution is unfeasible. Consequently, the proposed Tandem PCNN model with DLBSVM conquers the impediments happened in web-based and Android applications.

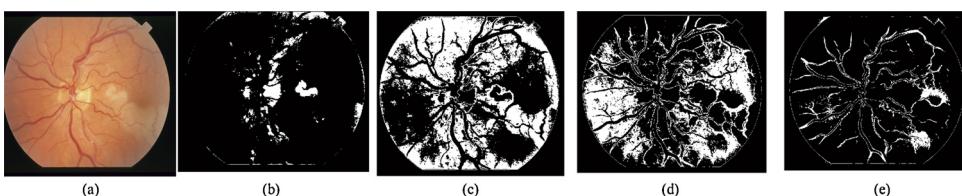


Fig. 4. Segmentation results of STARE Dataset (a) Input image, (b)–(e) Segmentation at four different iterations: $i = 0$, $i = 1$, $i = 2$, $i = 3$, (e) Final Segmentation image.

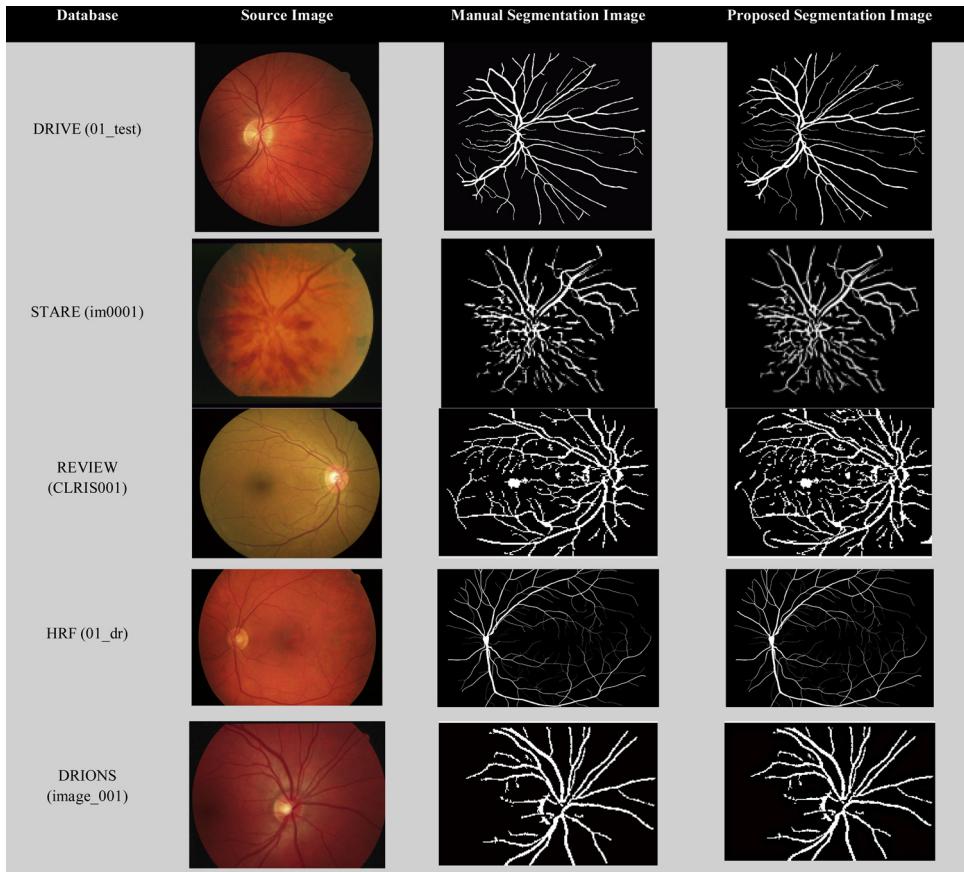


Fig. 5. The comparison of the proposed technique with manual segmentation.

Table 5

Performance comparison with competitive methods for retinal blood vessel segmentation.

Type	Author	Year	Method	Sensitivity %	Specificity %	Accuracy %
Supervised	Qiaoliang Li et al. [16]	2016	Cross-Modality Learning Approach	75.69	98.16	95.27
	Argyrios Christodoulidis et al. [2]	2016	Multi-Scale Tensor Voting Approach	85.06	95.82	94.79
	Shahab Aslani et al. [34]	2016	Multi-scale Gabor wavelet & B-COSFIRE Filters	75.45	98.01	95.13
	Villalobos-Castaldi et al. [35]	2010	Gray Level Co-occurrence Matrix (GLCM)	96.48	94.80	97.59
	Roberto Vega et al. [36]	2015	Lattice Neural Networks with Dendritic Processing	74.44	96.00	94.12
Unsupervised	Elaheh Imani et al. [11]	2015	Morphological Component Analysis	75.24	97.53	95.23
	George Azzopardi et al. [8]	2014	B-COSFIRE Filters	76.55	97.04	96.14
	Miguel A. Palomera-Perez et al. [37]	2010	Parallel Multiscale Feature Extraction and Region Growing	-	-	92.00
	R. GeethaRamani et al. [10]	2016	Gabor Filter & Principal Component Analysis	70.79	97.78	95.36
	Behdad Dashbozorg et al. [38]	2014	Graph-Based A/V Classification	90.00	84.00	87.40
	Lei Zhang et al. [39]	2015	Multi-Scale Textons	78.12	96.68	95.04
	Nagendra Pratap Singh et al. [40]	2016	Gumbel Probability Distribution Based Matched Filter	75.94	-	95.22
	Rashmi Panda et al. [14]	2016	Binary Hausdorff Symmetry Seeded Region Growing	73.37	97.52	95.39
	Proposed TPCNN, DLBSVM and Firefly Model		Average value	80.61	99.54	99.49

7. Conclusions

The proposed techniques are appropriate for the prescreening scheme of Diabetic Retinopathy. The Contrast Limited Adaptive Histogram Equalization improves the contrast of the image and removes the noise from the images captured at different illuminations. Tandem Pulse Coupled Neural Network model fuses two source images and generates the feature vectors to segment the retinal

vasculature. The produced feature vectors are utilized by Deep Learning Based Support Vector Machine technique to classify the vessels from non-vessels. DLBSVM uses numerous layers and decreases the errors in misclassifying the vessel pixels from non-vessel pixels. TPCNN model calibrates its parameters, so that error is limited. Firefly algorithm optimizes the parameters in classifying the vessel pixels. The segmented retinal vessels are compared with the ground truth images and the performance of the algorithm is analyzed. The proposed method facilitates the ophthalmologists in the diagnosis of Diabetic Retinopathy by considering the segmented retinal blood vessels structure. This work can be extended by quantifying the tortuosity of the vessel to distinguish the disease.

References

- [1] M. Islamuddin Ahmed, M. Ashraful Amin, High speed detection of optical disc in retinal fundus image, *SIViP* 9 (2015) 77–85.
- [2] A. Christodoulidis, T. Hurtut, H.B. Tahar, F. Cheriet, A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images, *Comput. Med. Imaging Graph.* 52 (September) (2016) 28–43.
- [3] S.H. Mohammad Alipour, H. Rabbani, M. Akhlaghi, A new combined method based on curvelet transform and morphological operators for automatic detection of foveal avascular zone, *SIViP* 8 (2014) 205–222.
- [4] S.R. Nirmala, S. Dandapat, P.K. Bora, Wavelet weighted distortion measure for retinal images, *SIViP* 7 (2013) 1005–1014.
- [5] Y. Zhao, X. Wang, X. Wang, F.Y. Shih, Retinal vessels segmentation based on level set and region growing, *Pattern Recognit.* 47 (2014) 2437–2446.
- [6] A. Salazar-Gonzalez, D. Kaba, Y. Li, Xiaohui, Segmentation of the blood vessels and optic disk in retinal images, *IEEE J. Biomed. Health Inform.* 18 (November (6)) (2014).
- [7] S. Aslani, H. Sarnel, A new supervised retinal vessel segmentation method based on robust hybrid features, *Biomed. Signal Process. Control* 30 (September (2016)) 1–12.
- [8] G. Azzopardi, N. Strisciuglio, M. Vento, N. Petkov, Trainable COSFIRE filters for vessel delineation with application to retinal images, *Med. Image Anal.* 19 (2015) 46–57.
- [9] S. Wilfred Franklin, S. Edward Rajan, Computerized screening of diabetic retinopathy employing blood vessel segmentation in retinal images, *Biocybern. Biomed. Eng.* 34 (2014) 117–124.
- [10] R. GeethaRamani, L. Balasubramanian, Retinal blood vessel segmentation employing image processing and data mining techniques for computerized retinal image analysis, *Biocybern. Biomed. Eng.* 36 (1) (2016) 102–118.
- [11] E. Imani, M. Javidi, H.-R. Pourreza, Improvement of retinal blood vessel detection using morphological component analysis, *Comput. Methods Programs Biomed.* 118 (2015) 263–279.
- [12] M.R.K. Mookiah, U. Rajendra Acharya, R.J. Martis, C.K.C.C.M. Lim, E.Y.K. Ng, A. Laude, Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: a hybrid feature extraction approach, *Knowledge Based Syst.* 39 (2013) 9–22.
- [13] R. Vega, G. Sanchez-Ante, L.E. Falcon-Morales, H. Sossa, E. Guevara, Retinal vessel extraction using Lattice Neural Networks with dendritic processing, *Comput. Biol. Med.* 58 (2015) 20–30.
- [14] R. Panda, N.B. Puhan, G. Panda, New Binary Hausdorff Symmetry measure based seeded region growing for retinal vessel segmentation, *Biocybern. Biomed. Eng.* 3 (6) (2016) 119–129.
- [15] S. Wang, ilong Yin, G. Cao, B. Wei, Y. Zheng, G. Yang, Hierarchical retinal blood vessel segmentation based on feature and ensemble learning, *Neurocomputing* 149 (2015) 708–717.
- [16] Q. Li, I.E.E.E. Member, B. Feng, L.P. Xie, P. Liang, H. Zhang, T. Wang, A cross-modality learning approach for vessel segmentation in retinal images, *IEEE Trans. Med. Imaging* 35 (January (1)) (2016).
- [17] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1798–182.
- [18] G.E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.* 18 (7) (2006) 1527–1554.
- [19] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science* 313 (5786) (2006) 504–507.
- [20] Y. Cho, L.-K. Saul, Kernel methods for deep learning, *NIPS* (2009) 342–350 9.
- [21] Y. Cho, L.-K. Saul, Large-margin classification in infinite neural networks, *Neural Comput.* 22 (10) (2010) 2678–269.
- [22] A. Abdullah, R.C. Veltkamp, M.A. Wiering, An ensemble of deep support vector machines for image categorization, *SOCPAR'09* (2009) 301–306.
- [23] M. Wiering, M. Schutten, A. Millea, A. Meijster, L. Schomaker, Deep support vector machines for regression problems, *Proceedings of the International Workshop on Advances in Regularization, Optimization, Kernel Methods, and Support Vector Machines: Theory and Applications* (2013).
- [24] M. Wiering, M. Van der Ree, M. Embrechts, M. Stollenga, A. Meijster, A. Nolte, L. Schomaker, The neural support vector machine, *Proceedings of the 25th Benelux Artificial Intelligence Conference (BNAIC)* (2013).
- [25] M.D. Abramoff, M.K. Garvin, M. Sonka, Retinal imaging and image analysis, *IEEE Rev. Biomed. Eng.* 3 (2010) 169–208.
- [26] N. Patton, T.M. Aslam, T. MacGillivray, I.J. Deary, B. Dhillon, R.H. Eikelboom, K. Yugesan, I.J. Constable, Retinal image analysis: concepts, applications and potential, *Progress Retinal Eye Res.* 25 (1) (2006) 99–127.
- [27] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, Ridge-based vessel segmentation in color images of the retina, *IEEE Trans. Med. Imaging* 23 (April (4)) (2004) 501–509.
- [28] C. Yao, H. Chen, Automated retinal blood vessels segmentation based on simplified PCNN and fast 2D-Otsu algorithm, *J. Cent. South Univ. Technol.* 16 (2009) 0640–0646.
- [29] T. Jemima Jebaseeli, D. Sujitha Juliet, C. Anand Devaduri, Segmentation of retinal blood vessels using pulse coupled neural network to delineate diabetic retinopathy, *Digital Connectivity – Social Impact Volume 679 of the Series Communications in Computer and Information Science*, Springer, 23 November, 2016, pp. 268–285.
- [30] DRIVE: Digital Retinal Image for Vessel Extraction, <http://www.isi.uu.nl/Research/Databases/DRIVE>, STARE: <http://cecas.clemson.edu/~ahoover/stare/>, REVIEW: <http://reviewdb.lincoln.ac.uk/Image%20Datasets/Review.aspx>, HRF: <https://www5.cs.fau.de/research/data/fundus-images/>, DRIONS: <http://www.ia.uned.es/~ejcarmona/DRIONS-DB.html>.
- [31] M.M. Fraz, S.A. Barman, P. Remagnino, A. Hoppe, A. Basit, B.U. yyanonvara, A.R. Rudnicka, C.G. Owen, An approach to localize the retinal blood vessels using bit planes and centerline detection, *Comput. Methods Programs Biomed.* 108 (2012) 600–616.
- [32] A. Bourouis, M. Feham, M.A. Hossain, L. Zhang, An intelligent mobile based decision support system for retinal disease diagnosis, *Decis. Support Syst.* 59 (2014) 341–350.
- [33] P. Prasanna, S. Jain, N. Bhagatt, A. Madabhushi, Decision support system for detection of diabetic retinopathy using smartphones, *7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops* (2013).
- [34] S. Aslani, H. Sarnel, A new supervised retinal vessel segmentation method based on robust hybrid features, *Biomed. Signal Process. Control* 30 (September (2016)) 1–12.
- [35] F.M. Villalobos-Castaldi, E.M. Felipe-Riverón, L.P. Sánchez-Fernández, *J. Vis.* 13 (2010) 263.
- [36] R. Vega, G. Sanchez-Ante, L.E. Falcon-Morales, H. Sossa, E. Guevara, Retinal vessel extraction using Lattice Neural Networks with dendritic processing, *Comput. Biol. Med.* 58 (2015) 20–30.
- [37] M.A. Palomera-Perez, M. Elena Martinez-Perez, H. Benitez-Perez, Jorge Luis Ortega-Arjona, parallel multiscale feature extraction and region growing: application in retinal blood vessel detection, *IEEE Trans. Inf. Technol. Biomed.* 14 (March (2)) (2010).
- [38] B. Dashtbozorg, A.M. Mendonça, A. Campilho, An automatic graph-based approach for Artery/Vein classification in retinal images, *IEEE Trans. Image Process.* 23 (March (3)) (2014).
- [39] L. Zhang, M. Fisher, W. Wang, Retinal vessel segmentation using multi-scale textons derived from key points, *Comput. Med. Imaging Graph.* 45 (2015) 47–56.
- [40] N.P. Singh, R. Srivastav, Retinal blood vessels segmentation by using Gumbel probability distribution function based matched filter, *Comput. Methods Programs Biomed.* 129 (2016) 40–50.



Modified Alexnet architecture for classification of diabetic retinopathy images[☆]



T. Shanthi*, R.S. Sabeenian

Department of Electronics and Communication Engineering, Sona College of Technology, Salem, Tamilnadu, India

ARTICLE INFO

Article history:

Received 28 December 2018

Revised 10 March 2019

Accepted 11 March 2019

Keywords:

Diabetic retinopathy
Convolutional neural network
Messidor database
Classification
Accuracy

ABSTRACT

Diabetic retinopathy (DR) is an illness occurring in the eye due to increase in blood glucose level. Among people in the age group of 70, 50% of deaths are attributed to diabetes. Early identification and appropriate treatment can reduce the loss of sight in many DR patients. Once the symptoms of DR are recognized, the severity of the disease should be evaluated for administering the right medication. This paper focuses on the classification of DR fundus images according to the severity of the disease using convolutional neural network with the application of suitable Pooling, Softmax and Rectified Linear Activation Unit (ReLU) layers to obtain a high level of accuracy. The performance of the proposed algorithm has been validated using Messidor database. In the case of healthy images, images of stage1, stage 2 and stage 3 of diabetic retinopathy, classification accuracies of 96.6% and 96.2%, 95.6% and 96.6% have been achieved.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Diabetes

Diabetes is caused by increase in the glucose level in blood. If this abnormality persists for a long period of time, it will bring about irreversible damage to the blood vessels. A diabetic person is prone for kidney failure, vision loss, bleeding gums, lower limb confiscation, wounds in the feet, and nerve damages. There is a high risk of even heart attack and stroke among diabetic patients. On the basis of the parts that are affected by rise in the blood glucose level, the diseases are named as Diabetic Nephropathy (nephrons in the kidney are damaged), Diabetic Neuropathy (neurons in the brain are affected), and diabetic retinopathy (retina of the eye is affected). The World Health Organization (WHO) predicts that diabetes will be the 7th leading fatal disease [1]. In 1980, there were 108 million diabetic patients only, but in 2014, the diabetic patients population increased by four folds to 422 million. According to statistics, an increase in diabetic patients above 18 years age from 4.7% to 8.5% has been observed. People below poverty line are the major victims of diabetes. In India, there are about 61.3 million people falling in the age group of 20–79 identified as diabetic patients. It is estimated that this figure is likely to increase to 101.2 million by 2030 [2].

* This paper is for CAEE special section SI-cih. Note that it was submitted originally for SI-uacn. Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. S. Smys.

* Corresponding author.

E-mail addresses: shanthi@sonatech.ac.in (T. Shanthi), sabeenian@sonatech.ac.in (R.S. Sabeenian).

Table 1
Different stages of DR.

Stages of DR	Ophthalmoscope recordings	Corresponding label in the proposed algorithm
Normal	Without any abnormalities	Healthy retina
Mild NPDR	Presence of micro aneurysms only	DR stage 1
Moderate NPDR	Microaneurysms are present but in a smaller amount as compared to severe NPDR.	DR stage 2
Severe NPDR	<ul style="list-style-type: none"> • Venous beading in two or more regions • Prominent intraretinal microvascular abnormality (IRMA) in one or more regions • Vitreous/pre- retinal hemorrhage 	DR stage 3
PDR	Neovascularization	

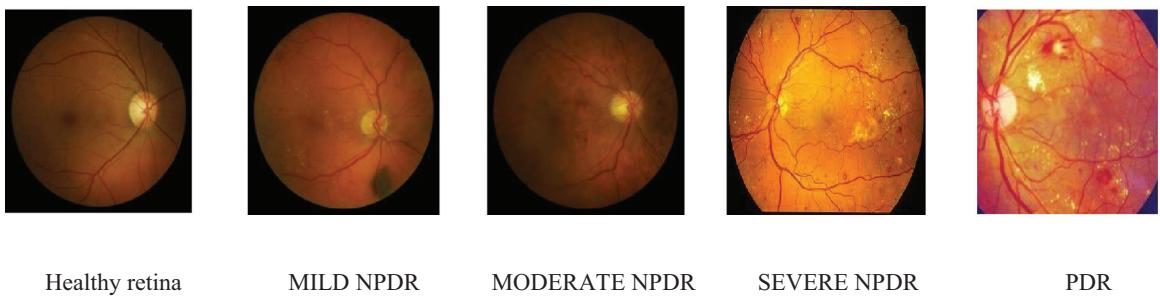


Fig. 1. Stages of DR starting from healthy fundus image [17].

1.2. Diabetic retinopathy

Persistent increase in the blood glucose level damages the retinal blood vessels. This increase in blood glucose level mainly punctures the blood vessels causing leakage of blood into the eye and thereby weakening the vision system. The human body inherently is capable of curing itself. When the brain finds a leakage of blood, it activates the nearby cells to control the situation. This activity leads to an abnormal growth of new blood vessels [3]. The newly formed vessels are weak. They affect the vision of the person in due course. Hence it is imperative that a diabetic patient should always take up a regular eye test. The patient's retina should be checked and monitored regularly by an ophthalmologist. Slit lamp bio microscopy, optical coherence tomography (OCT), fundus fluorescein angiography (FFA), and fundus photography are some of the eye testing techniques used to diagnose the disease in its early stage itself.

1.3. Phases of diabetic retinopathy

The presence of abnormalities and their magnitude define the severity of the disease. The identification of manifestations like micro aneurysms, hemorrhage, neovascularization, venous beading is a major process in diagnosis. Micro aneurysms refer to blood clots of size 100–120 µm that are usually circular in shape. Leakage of blood from a broken blood vessel is called hemorrhage. Abnormal growth of tiny blood vessels is termed as neovascularization. Venous beading refers to the central expansions of the veins lying adjacent to the occluded arterioles. Diabetic retinopathy patients are categorized as Non-Proliferative Diabetic Retinopathy (NPDR) patients and Proliferative Diabetic Retinopathy (PDR) Patients. Further, depending on the severity of the disease, NPDR patients are distinguished as mild-stage, moderate-stage, and severe-stage NPDR patients. Stages in the severity of diabetic retinopathy are described in [Table 1](#). Images of different stages of diabetic retinopathy [17] appear in [Fig. 1](#).

The remaining part of this paper is organized as follows: [Section 2](#) provides a literature review of the research area. The Materials and the proposed method are explained in [Section 3](#). The results and discussion explained in [Section 4](#). Conclusions are finally given in [Section 5](#).

2. Literature review of existing methods for automatic DR detection

Diabetic retinopathy is one of the most prominent ailments among the diabetic patients and patients can be prevented from vision loss if the disease is diagnosed at an earlier stage. Once the disease is diagnosed the patient has to be evaluated for every six months to know the progress of the disease [18]. An efficient algorithm to detect and classify the fundus images will be helpful for the ophthalmologist to a greater extent in eradicating the vision loss due to DR. Researchers have developed a number of algorithms for application in the study of images to facilitate precise diagnosis of diabetic retinopathy. The configuration of human eye includes optic disc and optic nerves. Detection and classification of DR can be carried out by segmenting the images of the portions of the parts from the fundus image or by examining the fundus image for the occurrence of hemorrhages, lesions, micro aneurysms, exudates, etc.

The authors in [5] have presented an algorithm that computes several features such as area occupied by blood vessels, foveal zone irregularities, and micro aneurysms. The proposed method utilizes curvelet coefficients of fundus image and angiograms. Three-stage (Normal, Mild NPDR, and PDR) classification is adopted. The study has been carried out among 70 patients. The proposed method has achieved 100% sensitivity. The researchers in [4,19] have classified DR images on the basis of the presence of micro aneurysms. Features such as circularity and area of micro aneurysms are considered in feature extraction. Datasets namely DRIVE, ROC, and DIARETDB1 are used in this research. The algorithm introduced by the authors has produced 94.44% and 87.5% sensitivity and specificity respectively. The proposed method employs principal component analysis (PCA) to segregate the image of the optic disc from the fundus image. Using Enhanced MDD classifier, the authors have achieved an increase by about 25–40% in the detection probability of exudates near the region surrounding the optic disc in the image. A set of 39 pictures has been taken up for study; the images have been classified using proposed algorithm into 4 normal fundus images and 35 images showing fundus with exudates.

In [6,7], the researchers have described the application of SVM, Bayesian method, and PNN for identifying the stages of DR as NPDR or PDR from the study of fundus images. The database DIARETDB0 pertaining to 130 images has been used. Initially the portions of blood vessels, exudates, and hemorrhages from the images of DR have been segregated. The accuracy obtained by using the proposed method involving the application of PNN, SVM, and Bayes classification techniques is 87.69%, 95.5%, and 90.76% respectively. In [8], the author discusses the validation of results obtained from using trained SVM classifier. Three public datasets, namely, MESSIDOR, DIARETDB1, and DRIVE have been used. An accuracy of 93% has been achieved in classifying exudates and micro aneurysms by segmenting the images of blood vessels.

Reference [9] explains the application of local binary pattern texture feature to detect exudates and reports that an accuracy of 96.73% has been obtained. In Reference [10], a dual classification method is introduced. This method involves boot strapped decision tree to classify the fundus images. Two binary vessel maps are created by reducing the feature vectors dimensionally. The suggested method has produced an accuracy of 95%.

In Reference [11], the author has proposed an algorithm that employs Gabor filtering technique and SVM classifier to categorize the DR images. Prior to the application of classifier, Circular Hough Transform (CHT) and CLAHE techniques are applied to the input images. An accuracy of 91.4% has been realized for the images drawn from STARE database. In Reference [12], the application of Multi-Layer Perception Neural Network (MLPNN) to detect diabetic retinopathy is explained. Nine statistical features are extracted through 64-point Discrete Cosine Transform (DCT) operation. The statistical features thus obtained are fed to the neural network.

In Reference [13], morphological operation using the intensity of the image as threshold to segment is described. The process of segmenting the portions of exudates in the HIS space is explained. Reference [14] discusses the application of CNN architecture along with data augmentation techniques in the classification of DR images. The severity of DR is classified into five stages. Kaggle database is used. An accuracy of 75% is produced.

The researchers in [15] have proposed a method of grouping a set of error dependent networks for image classification. The system is tested with the dataset that contains remote sensing images taken from an agricultural land in a village near Feltwell (UK). The classification was done by pixel basis. Based on the elements of feature vectors the pixels were characterized as one of the five different agricultural areas like sugar beets, carrots, potatoes, bare soil and stubble. Three different types of ensembles (E1, E2&E3) were created. The classification accuracy obtained was 89.83%, 87.87% and 90.46% for E1, E2 and E3ensembles respectively. The authors in [15] have attempted to recognize macular disorders from SD-OCT images using recombined convolutional neural network. The preprocessing stage uses BM3D filter for noise removal. In BM3D approach the image is initially divide the image into 2D blocks and then recombined to form 3D blocks based on the similarity. The low level, mid-level and high level features are extracted sequentially from 18 layered residual networks. The network was tested for different kernel sizes like 3×3 , 5×5 , and 7×7 . Accuracy, precision and recall were used as performance indicators. The kernel of size 3×3 in the recombined residual network achieves a better accuracy of 90%.

3. Materials and the proposed method

3.1. Dataset description

Messidor [16] dataset consists of around 1190 color fundus images with annotations in Excel file. It has been used for testing and training the proposed Alexnet architecture. A camera having three separate charge-coupled devices and attached with Topcon TRC-NW6 Non-Mydriatic Retinal Camera has been employed to take photographs of FoV images at 45° . The images to be fed as input have been taken in 8 bp color level at 1390×1040 , 2160×1884 , and 2216×1166 pixels. Each dataset has been distinguished into four different subsets, namely, Healthy retina, DR stage 1, DR stage 2, and DR stage 3, according to the annotations given for the respective images. Grades on the basis of presence of micro aneurysms and hemorrhages have been assigned to the images. Images showing no signs of micro aneurysms and hemorrhages represent healthy retina. Images containing only a few micro aneurysms correspond to DR stage 1. Images presenting more (between 5 and 15) micro aneurysms and few (less than 5) hemorrhages are classified to stand for DR stage 2. Images showing more (greater than 15) micro aneurysms, more (greater than 5) hemorrhages, and very few neovascularization locations are considered to denote DR stage 3.

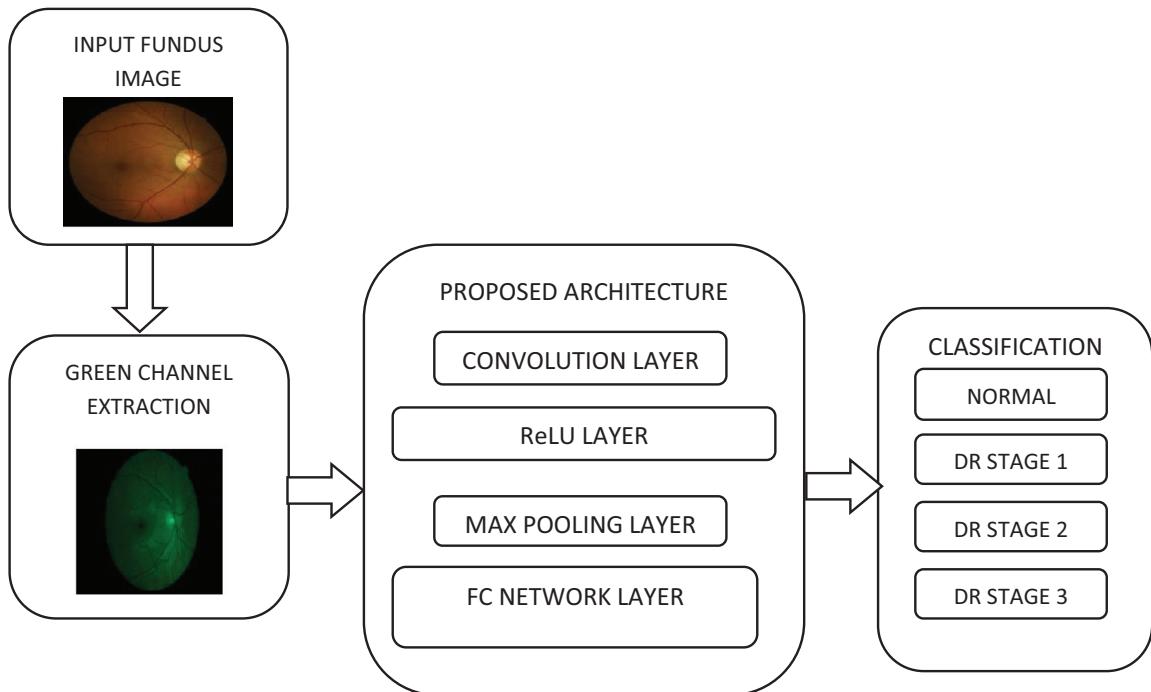


Fig. 2. Flow diagram of the proposed method.

3.2. Method proposed

The present research aims to classify the fundus images with high accuracy into various stages of diabetic retinopathy. There is a massive growth in patients affected by diabetic retinopathy. It is necessary to categorize the patients into different stages of diabetic retinopathy in a swift manner. Through the present research with the application of a modified Alexnet architecture we have striven to increase the classification accuracy in the study of DR images.

3.3. Preprocessing

The Messidor dataset has been arranged into four different subsets, viz., Healthy retina, DR stage 1, DR stage 2, and DR stage 3. The annotations are given for all the images in the Messidor dataset in Excel file. These particulars are much useful in the grouping of subsets. The green channel provides finer details of the optic nerves and other features of the retina. Initially the images are split into RGB channels. Using the green channel, the input fundus image is refined so as to get a higher accuracy in classification. Improved images in different stages are fed to the modified Alexnet architecture to categorize the fundus images into the corresponding stages of diabetic retinopathy. The stages followed in the present research are depicted as a flow chart in Fig. 2

3.4. Convolutional neural network

A convolutional neural network is a deep-learning neural network. It has been emulated from the observation of biological process. It mimics the functions of different layers in the human brain. CNN has been proved to be very efficient in all image processing applications like face recognition, pattern recognition, etc. The proposed architecture carries out the image processing through different layers. The input image is fed to the initial convolution layer and successive output is processed in different layers in the proposed architecture. Convolutional layer takes out a patch of input fundus image, and different filters are applied to the input images. The output from the convolutional layer is fed as input to maximum pooling layer where unwanted pixels are removed.

CNN has different architectures like Le-Net, Alex-Net, Google-Net, Conv-Net, Res-Net, etc. In the present research we have employed Alexnet architecture because it has better computational ability to address complexities than other architectures. CNN works with three-dimensional (length, width, and depth) values. For a normal color image, the dimensions of the image are denoted by a , b , and c , where ' a ' stands for the length of the color image, ' b ' represents the breadth of the color image, and ' c ' denotes the number of color channels present in the input. The primary layer receives the input image, and after processing the final layer of the architecture provides the prediction. Generally, Alexnet architecture has eight layers,

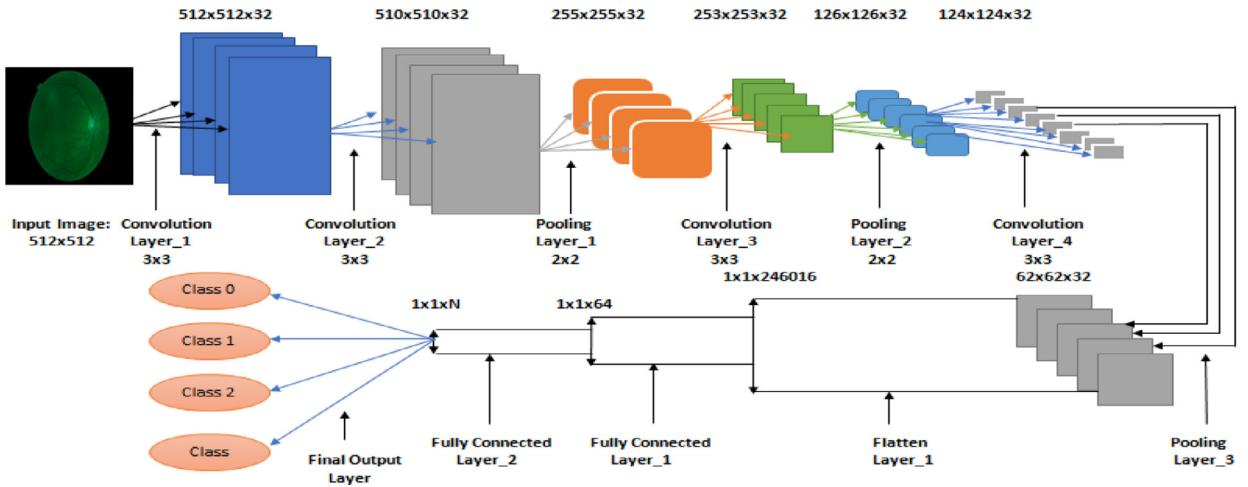


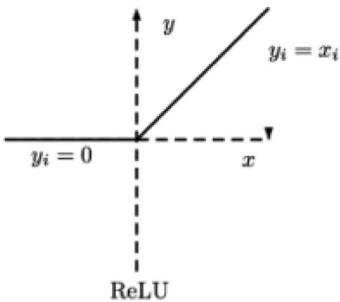
Fig. 3. Modified Alexnet architecture for classification of diabetic retinopathy images.

in which the first five layers are convolutional and maximum pooling layers, followed by three layers fully connected to the neural network.

3.5. Proposed Alexnet architecture

Among various CNN architectures, Alexnet is one of the most efficient architectures that are widely employed to address problems in image classification. The Fig. 3 depicts the Modified Alexnet architecture for classification of diabetic retinopathy images. The operations involved in the implementation of the proposed Alexnet architecture are described below:

- The first step is to resize the input fundus image to the size of 259×259 pixels corresponding to the breadth and height and the three color channels representing the depth of the input fundus image.
- The output of neurons is computed as a scalar product of a small portion of the image with their corresponding weights. This process is repeated along the length and breadth. This operation is performed in convolutional layer.
- In Rectified Linear Unit (ReLU) layer, an element-wise activation function is employed. This layer replaces all the negative activations with 0 by introducing nonlinearity to the system and by applying the function $-f(k) = \max(0, k)$.



Activation function

- In pooling layer, the samples are reduced along the spatial coordinates. This process is known as decimation.
- Fully Connected (FC) layer computes the Class scores for each image and gives the prediction. The probability score for each of the prediction class is computed and the class that is scoring maximum probability score is chosen as the predicted class as shown in Fig. 4.

3.6. Performance metrics

The performance of the architecture is assessed in terms of accuracy, specificity, sensitivity, and precision factor in respect of all DR stages. Alexnet architecture has been employed to train 710 input fundus images. Of these 710 images, 303 input fundus images have been used to evaluate the performance of the algorithm. Such a sample size is necessary to realize perfect classification with minimum computational error. True positive (TP), true negative (TN), false positive (FP), and

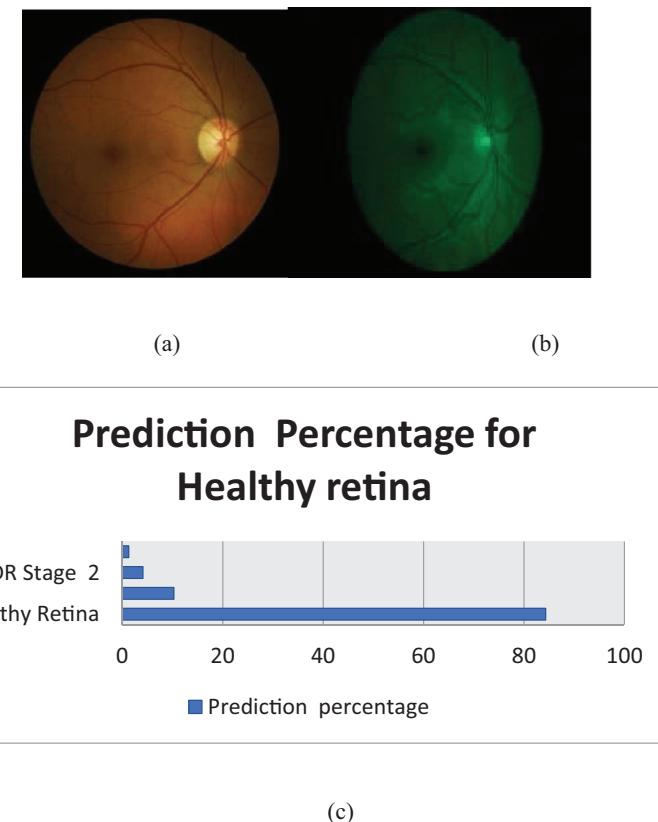


Fig. 4. Prediction for testing image (Healthy retina) (a) input RGB fundus image (b) green plane of the image (c) the test image is predicted as healthy retina as per the prediction percent.

false negative (FN) values for each class of input images vary according to the values obtained through confusion matrix. If a healthy retina is predicted as healthy retina, then it represents a true positive value for that class; all other correct predictions for other classes are called as true negative with respect to this class (healthy retina). When an input healthy retina image is predicted as DR image of any stage (DR stage 1, 2 or 3) then it is termed as false positive. On the other hand, if a DR input image of any stage (DR stage 1, 2, or 3) is predicted as healthy retina, then it is termed as false negative. Once all these values are obtained for each class of images, then the performance metrics like sensitivity, specificity, accuracy, and precision factor determined using the following equations:

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}}$$

$$\text{Precision factor} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Sensitivity measures the proportion of correctly classified images that are correctly identified to the total number of positive classification. Specificity measures how well the algorithm predicts the other classes. Accuracy measures the total prediction rate of the algorithm.

4. Results and discussion

Training the architecture has been carried out with 710 fundus images from the given dataset. The performance of the proposed architecture has been evaluated by testing it on 303 fundus images. It is found that among the 303 images of

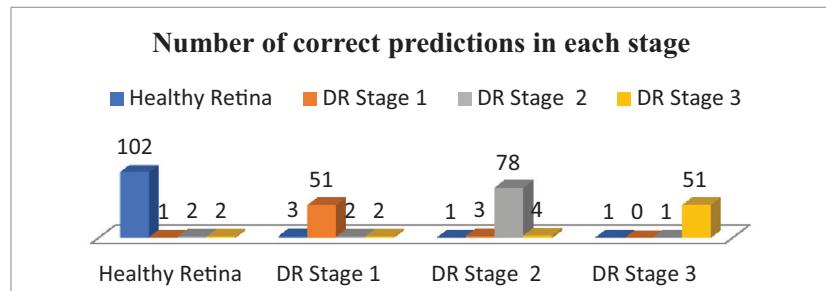


Fig. 5. Classification of different DR grade images.

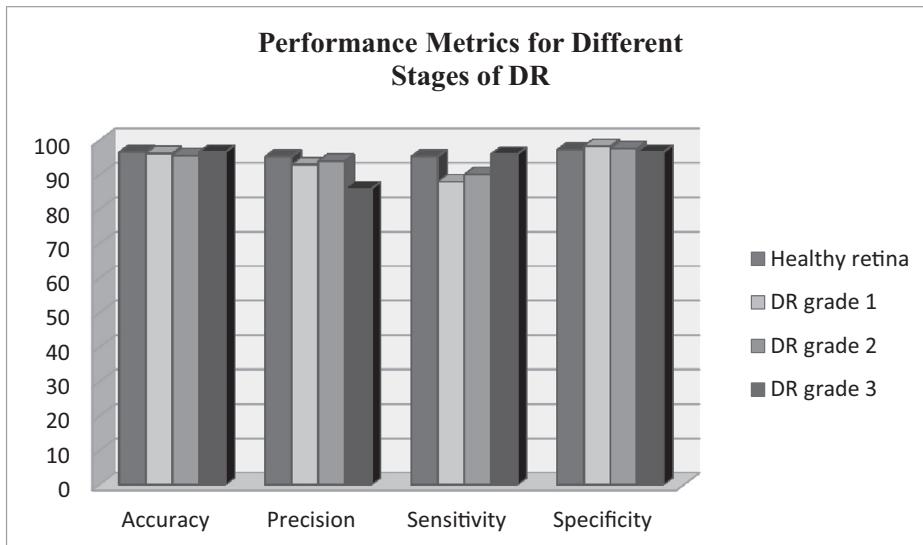


Fig. 6. Performance metrics for different stages of DR.

Table 2
Confusion matrix.

Input class	Different level of diabetic retinopathy				Total no. of images
	Healthy retina	DR stage 1	DR stage 2	DR stage 3	
Healthy retina	102	1	2	2	107
DR stage 1	3	51	2	2	58
DR stage 2	1	3	78	4	86
DR stage 3	1	0	1	51	53
Total no. of images	107	55	83	59	303

Table 3
Manipulations from confusion matrix.

Input grades →	Healthy retina	DR grade 1	DR grade 2	DR grade 3
TP	102	51	78	51
TN	180	231	204	231
FP	5	4	5	8
FN	5	7	8	2

fundus tested with the application of the proposed Alexnet architecture, 107 images are images of healthy retina, 58 images belong to DR stage 1, 86 images pertain to DR stage 2, and 53 fundus images fall in the category of DR stage 3.

Table 2 and Fig. 5 show that 102 images out of 107 are correctly predicted as Healthy retina, 51 images out of 58 are correctly predicted as DR stage 1 image, 78 images out of 86 are correctly predicted as DR stage 2 image, 51 images out of 53 are correctly predicted as DR stage 3 image. Table 3 shows observations from the confusion matrix.

Table 4
Performance metrics for test images with different DR grades.

Input grades	Accuracy	Precision	Sensitivity	Specificity
Healthy retina	96.6	95.3	95.3	97.3
DR grade 1	96.2	93.0	88.0	98.3
DR grade 2	95.6	94.0	90.1	97.6
DR grade 3	96.6	86.0	96.0	96.6

Table 4 presents the performance metrics obtained through using the proposed algorithm. **Fig. 6** shows the performance metrics for different stages of DR.

5. Conclusion

Severe vision loss in diabetic patients can be avoided by detecting and treating diabetic retinopathy at an early stage. The method proposed in this paper aims at providing an optimal solution for the classification of diabetic retinopathy patients according to the severity of the disorder. Deep learning is one of the state of the art techniques to address classification problems and it provides better accuracy. Efficient convolutional neural network architecture to detect and classify the fundus images will be helpful for the ophthalmologist to a greater extent in eradicating the vision loss due to diabetic retinopathy. The probability score for each of the prediction class is computed by the final fully connected layer of the deep learning architecture and the class that is scoring maximum probability score is chosen as the predicted class. The testing and training the evaluation of proposed Alex net architecture is done using Messidor dataset. The dataset from Messidor dataset s segregated into two, 70% of the fundus images are trained by Alexnet architecture and 30% of the fundus images from Messidor dataset are used to evaluate the performance of the algorithm. A modified Alexnet architecture that is used to categorize the input fundus images is employed in the present research and the results obtained are discussed in this paper. The performance of the modified Alexnet architecture is evaluated in terms of performance metrics like accuracy, specificity, sensitivity and precision. Accuracy in classifying the images collected from the Messidor dataset into Healthy retina, diabetic retinopathy stage 1, diabetic retinopathy stage 2 and diabetic retinopathy stage 3 using the proposed modified Alexnet architecture is 96.6%, 96.2%, 95.6% and 96.6% respectively. The results can be improved further by increasing the size of a dataset. In future the algorithm can be tested with Kaggle dataset with more of images.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.compeleceng.2019.03.004](https://doi.org/10.1016/j.compeleceng.2019.03.004).

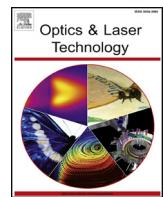
References

- [1] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60(May (6)):84–90.
- [2] Whiting DR, Guariguata L, Weil C, Shaw J. IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract* 2011;94(Dec (3)):311–21.
- [3] Hajeb Mohammad Alipour S, Rabbani H, Akhlaghi MR. Diabetic retinopathy grading by digital curvelet transform. *Comput Math Methods Med* 2012;2012:1–11.
- [4] SB S, Singh V. Automatic detection of diabetic retinopathy in non-dilated RGB retinal fundus images. *Int J Comput Appl* 2012;47(Jun (19)):26–32.
- [5] Singh N, Tripathi RC. Automated early detection of diabetic retinopathy using image analysis techniques. *Int J Comput Appl* 2010;8(Oct (2)):18–23.
- [6] Bhatia K, Arora S, Tomar R. Diagnosis of diabetic retinopathy using machine learning classification algorithm. 2016 2nd international conference on next generation computing technologies (NGCT), October; 2016.
- [7] Sahebrao R, N S, T S, Dhopeshwarkar M. Automated diagnosis non-proliferative diabetic retinopathy in fundus images using support vector machine. *Int J Comput Appl* 2015;125(Sep (15)):7–10.
- [8] Cunha-Vaz BSPJG. Measurement and mapping of retinal leakage and retinal thickness – surrogate outcomes for the initial stages of diabetic retinopathy. *Curr Med Chem-Immuno Endocr Metab Agents* 2002;2(Jun (2)):91–108.
- [9] Anandakumar H, Umamaheswari K. Supervised machine learning techniques in cognitive radio networks during cooperative spectrum handovers. *Cluster Comput* 2017;20:1–11. doi:[10.1007/s10586-017-0798-3](https://doi.org/10.1007/s10586-017-0798-3).
- [10] Omar M, Khelifi F, Tahir MA. Detection and classification of retinal fundus images exudates using region based multiscale LBP texture approach. 2016 international conference on control, decision and information technologies (CoDIT), April; 2016.
- [11] Welikala RA, Fraz MM, Williamson TH, Barman SA. The automated detection of proliferative diabetic retinopathy using dual ensemble classification. *Int J Diagn Imaging* 2015;2(Jun (2)).
- [12] Haldorai A, Ramu A, Chow C-O. Editorial: big data innovation for sustainable cognitive computing. *Mob Netw Appl* 2019;24(Jan):221–3.
- [13] Purandare M, Noronha K. Hybrid system for automatic classification of Diabetic Retinopathy using fundus images. 2016 online International conference on green engineering and technologies (IC-GET); Nov. 2016.
- [14] Bhatkar AP, Kharat GU. Detection of diabetic retinopathy in retinal images using MLP classifier. 2015 IEEE international symposium on nanoelectronic and information systems, December; 2015.
- [15] Partovi M, Rasta SH, Javadzadeh A. Automatic detection of retinal exudates in fundus images of diabetic retinopathy patients. *J Anal Res Clin Med* 2016;4(May (2)):104–9.
- [16] Anandakumar H, Umamaheswari K. A bio-inspired swarm intelligence technique for social aware cognitive radio handovers. *Comput Electr Eng* 2018;71(Oct):925–37.
- [17] Doshi D, Shenoy A, Sidhpura D, Gharpure P. Diabetic retinopathy detection using deep convolutional neural networks. 2016 international conference on computing, analytics and security trends (CAST), December; 2016.

- [18] Giacinto G, Roli F. "Design of effective neural network ensembles for image classification purposes". *Image Vision Comput* 2001;19(9–10):699–707.
- [19] Meng T, Wu C, Jia T, Jiang Y and Jia Z, 'Recombined convolutional neural networks for recognition of macular disorders in SD-OCT images', In 2018 37th Chinese control conference (CCC), pp 9362–9367, IEEE.

T. Shanthi is currently working as an Assistant Professor (Sr.G) in Sona College of Technology and a team member in the research group in SONASIPRO. She has more than 16 years of teaching experience. She is pursuing her Ph.D. Degree under Anna University, Chennai in the area of Digital Image processing. Her areas of interest include image analysis and deep learning.

R.S. Sabeenian is a Professor and Head/ECE in Sona College of Technology. He has 17+ years of teaching experience. He received his Ph.D. Degree in the year 2009 in the area of Image processing. He heads the research group (SONASIPRO). He has published more than 85 research papers in various Journals. He has authored several books and reviewer for various journals.



Full length article

An automated early diabetic retinopathy detection through improved blood vessel and optic disc segmentation

Shailesh Kumar^a, Abhinav Adarsh^a, Basant Kumar^a, Amit Kumar Singh^{b,*}^a ECED, MNNIT Allahabad, Prayagraj, India^b CSED, NIT Patna, Patna, India

HIGHLIGHTS

- Presents an automated early diabetic retinopathy detection scheme for colour fundus image.
- Uses mathematical morphology operation for pre-processing and blood vessel detection.
- The accuracy of the proposed algorithm is evaluated based on sensitivity and specificity.
- The method offer better performance than similar stat-of-the-art techniques.

ARTICLE INFO

Keywords:

Fundus image
Blood vessels
Optic disc
Mathematical morphology
Watershed transform
Radial basis function neural network

ABSTRACT

This paper presents an automated early diabetic retinopathy detection scheme from color fundus images through improved segmentation strategies for optic disc and blood vessels. The red lesions, microaneurysms and hemorrhages are the earliest signs of diabetic retinopathy. This paper essentially proposes improved techniques for microaneurysm as well as hemorrhages detection, which eventually contribute in the overall improvement in the early detection of diabetic retinopathy. The proposed method consists of five stages- pre-processing, detection of blood vessels, segmentation of optic disc, localization of fovea, feature extraction and classification. Mathematical morphology operation is used for pre-processing and blood vessel detection. Watershed transform is used for optic disc segmentation. The main contribution of this model is to propose an improved blood vessel and optic disc segmentation methods. Radial basis function neural network is used for classification of the diseases. The parameters of radial basis function neural network are trained by the features of microaneurysm and hemorrhages. The accuracy of the proposed algorithm is evaluated based on sensitivity and specificity, which are 87% and 93% respectively.

1. Introduction

The aim of mass screening programs for diabetic retinopathy is to detect and diagnose the disorder earlier than it causes loss of vision. Mass screening programs use computerized fundus photography of retina with or without mydriasis (pupil dilation) to detect eye diseases. Automated analysis of retinal pictures has the likelihood to enhance the effectiveness of screening programs in comparison to manual picture evaluation. Diabetic Retinopathy (DR) is an eye retinal disease caused by long standing diabetic condition also referred as diabetic mellitus (DM) [1,2,3]. DR is a vital cause of blindness in the functioning age population between 20 and 60 years. There is 78% chance of having DR if the patient is above 30 years of age and suffers from diabetes for more

than 15 years [1,4].

Because of diabetic retinopathy, multiple lesions are produced on retinal surface. Microaneurysms, hemorrhages, exudates, and cotton wools are the major lesions, which indicate the presence of DR. *Microaneurysms* (MAs) are the lesions that appear at the premature phase of the disease and the number of MAs grow as disease develops [5,6]. *Hemorrhages* are likely to be the lesions that appear after MAs. Hemorrhages appear in case of blood leakage from the vascular tree present in the retina [7,8]. Hence, the detection of MAs and hemorrhages plays important role in automated retinopathy detection at early stages [9,10,11]. At later stages, yellow- white colour plasma starts leaking from blood capillaries, known as *hard exudates* [12]. *Cotton wools* appear next to exudates, which are silver color contusions caused

* Corresponding author.

E-mail addresses: shailesh@mnnit.ac.in (S. Kumar), adarsh.abhinav@gmail.com (A. Adarsh), singhbasant@mnnit.ac.in (B. Kumar), amit.singh@nitp.ac.in (A.K. Singh).

by fats outpouring from veins. MAs, hemorrhages and exudates are the features of non-proliferative diabetic retinopathy (NDPR). NDPR is an early stage of diabetic retinopathy. The advance stage of DR is known as proliferative diabetic retinopathy (PDR) [13,14]. In this stage, the main blood vessels of retina are completely blocked. In order to supply oxygen to retina, new blood vessels grow, named as abnormal blood vessels [15,16]. The vision of patient is severely affected in PDR. The fundamental problem with DR is asymptomatic, it does not affect the vision until it reaches at advance stage. DR may affect one or both eyes simultaneously.

A retinal color fundus image contains a variety of structures such as vascular tree, optic disc, fovea and red lesions, like MAs and hemorrhages. MAs and hemorrhages are considered as dark lesions, whereas exudates and cotton wools are considered as bright lesions. In order to detect lesions from the fundus image, firstly we have to remove the morphological structures present inside the retina. Mathematical morphological operation is used for detection of vascular tree, segmentation of optic disc and localization of fovea. Morphology is a branch of science that studies the shape and structure of creatures and plants. Morphology operation is a non-linear image processing technique [17].

Morphological processing commonly consists of two parts: objective image and structuring element. The size of structuring element is very small as compared to objective image. The structuring element is made of 0's and 1's and the shape of structuring element could be flat such as linear, disc, square, diamond shape and non-flat such as ball shape. The structuring element can be placed inside the concerning image if it fits, and outside if it does not fit. The structural information is derived by identifying the location at which the structuring element fits inside the picture. It depends upon size and shape of structuring element [18,19]. Details of morphology operation is discussed in later section. In this paper, mathematical morphology is widely used for pre-processing and extraction of vascular tree, optic disc and fovea. We have used mathematical morphology operation which is iterated one more time, in order to get improved blood vessels segmentation. Optic disc segmentation is done using Watershed Transform.

After the removal of all morphological structures the retinal fundus image under investigation is left with white spots of nearly circular shape. These structures are described as MAs and hemorrhages. In the next stage, seven features (area, perimeter, circularity, number of objects, major axis, minor axis, aspect ratio) are extracted from MAs and hemorrhages: these features are used for training a classification model for automated detection of DR. Radial Basis Function Neural Network (RBFNN) is utilized as classification model. RBFNN is a special type of neural network which consists of only three layers, including input and output layer. The developed neural network model classifies the fundus image into healthy image and DR image. The proposed DR detection scheme is exploiting the combined benefits of morphology operation (for blood vessel segmentation), Watershed Transformation (Optic disc segmentation) and RBF NN (for DR classification). Overall contribution of the proposed work can be summarized as follows:

1. To propose an improved blood vessels detection technique by morphological iterative process
2. To develop an automated Optic Disc algorithm using marker controlled Watershed Transform
3. Implementation of RBFNN based classification mass screening of diabetic retinopathy.

Rest of the paper is organized as follows. Previous works related to proposed scheme are presented in Section 2. Proposed schemes for microaneurysm and hemorrhages detection from fundus images are described in Section 3. Results and conclusions are explained in Sections 4 and 5 respectively.

2. Previous related work

Automated detection of eye related diseases from fundus image is a hot topic for researchers these days. Automated detection of glaucoma and DR have emerged as an important topic for researchers working in the area of medical image processing and computer vision. A number of DR detection systems and algorithms have been suggested in last two decades. Several preprocessing methods of retinal fundus images have been proposed. Seyed Hossein Rasta et al. [20] and Sharad K Yadav et al. [21] have presented the comparative analysis of illumination correction and contrast enhancement methods. V. Thirilogasundari et al. [22] and Ayyaz Hussain et al. [23] have suggested fuzzy based methods for removal of impulse noise.

Literature review of some of the automated DR detection algorithms from fundus images are discussed here. Daniel Welfer et al. [24] suggested a method for detachment of optic disc in color fundus images using mathematical morphological operation. In this work, optic disc detection has been divided into two parts; First part finds the points on optic disc and second part finds the optic disc boundary. In order to find internal optic disc point, they have applied morphological operations such as 'skeletonization' and 'recursive pruning' on blood vessels. They have used the pixels of main blood vessels, which are in the vicinity of optic disc center, as 'internal marker' and circles of different diameters, centered at optic disc center point, as 'external marker'. K. Parvati et al. [25] proposed image segmentation method based on gray-scale morphology using Watershed Transform for medical images. First of all, morphological gradient is applied on gray-scale image, which shows the edge toughness of every pixel. They have created internal markers and external markers using regional maxima. These markers are superimposed with original image, followed by applying Watershed Transform for segmentation. Over segmentation is common problem of Watershed Transform. The problem arises due to noise and quantization error. Sarni Suhaila Rahim et al. [2] proposed automatic detection of microaneurysm in color fundus images for diabetic retinopathy. They explored several preprocessing, feature extraction and classification techniques. In system I, they have used adaptive histogram equalization, Discrete Wavelet Transform, filtering and morphology process for preprocessing. Area of pixels, mean and standard deviation are the extracted features of DR. In System II, shade correction, vessel segmentation and morphology operation have been used for preprocessing techniques. For classification, they have used decision tree, K-nearest neighbor, radial basis function, polynomial support vector machine. System III consists of only two stages: preprocessing and detection of microaneurysm. Preprocessing has been done by grey scale conversion and adaptive histogram equalization. Circular Hough Transform has been proposed to locate the microaneurysm due to the circular shape of microaneurysm. System IV also consists of two stages. Combination of preprocessing techniques, grey scale conversion and Contrast Limited Adaptive Histogram Equalization (CLAHE) are implemented in first system, whereas grey scale and fuzzy histogram equalization are used in second system. Circular Hough Transform is used for detection of microaneurysms. Jang Pyo Bae et al. [9] proposed a study on hemorrhage detection using hybrid method in fundus images. Detection of hemorrhages from fundus images consists of different stages such as preprocessing and candidate extraction. The objective of preprocessing is to make the brightness of picture uniform and improve the difference among hemorrhages and ground. They applied hue saturation value for brightness correction and CLAHE to fundus images. Template matching with normalized cross correlation has been used for candidate extraction. Shailesh Kumar and Basant Kumar [26] proposed a DR detection system by extracting area and number of MAs from color fundus images. The detection system consists of multiple phases namely; preprocessing, exudate detection, blood vessel extraction, optic disc segmentation, fovea localization, feature extraction and classification of fundus image. They have extracted two features namely, area and number of MAs, after extraction of MAs from retinal images.

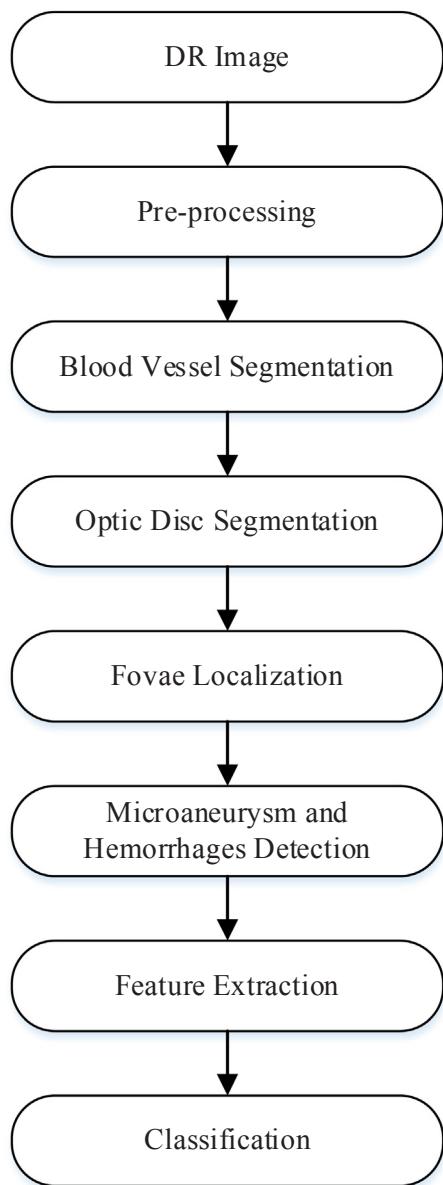


Fig. 1. Block diagram of automatic detection of DR from fundus images.

There are some limitations of the existing DIP based diabetic retinopathy detection methods. Some red lesions are very near to the blood vessels, making it difficult to be distinguished. Microaneurysms usually have very small size ($\leq 125 \mu\text{m}$ of diameter) and small intensity of variation with the background, which makes the detection of microaneurysms a complicated task. The biggest challenge for the detection of red lesion is the segmentation of microaneurysms in the region of low contrast and in presence of white lesions in the retina (exudates). If one has cataract, glaucoma, macular degeneration or any other eye disease then diabetic retinopathy detection is impossible [1,2].

3. Proposed scheme for early diabetic retinopathy detection

3.1. Pre-processing of fundus image

Proposed diabetic retinopathy detection model consists of mainly three parts such as preprocessing, feature extraction and classification. The block diagram for automatic detection of DR is shown in Fig. 1. Preprocessing is the essential part of detection technique using image processing. There are variety of pre-processing methods suggested in

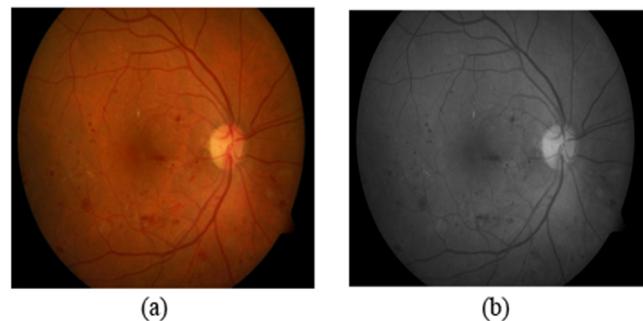


Fig. 2. (a) Resized RGB image (b) Gray image.

the literature [20]. Retinal color fundus images often suffer from non-illumination, low contrast due to anatomy of the eye fundus (the eye fundus with a 3-D concave shape), opaque media inside the eye, wide angle optics of the cameras, variation in size of pupil, geometry of sensor array, and movement of eye during capturing image. Hence, preprocessing plays important role in retinal fundus image analysis. The main objective of preprocessing technique is to increase the detection probability of disease by visual assessment and computer aided segmentation of retinal images. First of all, color fundus images, which are taken from different sources, are resized to 640×480 pixels. In order to reduce processing time, images are converted into gray scale. The resized image and gray scale images are shown in Fig. 2. Retinal color fundus image consists of three channels namely red, green and blue channel. Green channel of the RGB image is used for preprocessing because of the great contrast between blood vessels and background, and the best contrast between optic disc and tissue inside the retina. Red channel is moderately bright and veins of choroid is very noticeable. The retinal vessels are visible however, it shows less contrast than green channel. The gray scale image will be used in localization of optic disc. Blue channel is not used in detection because it contains more noise and less information about the morphological structure of the retina.

Some Basic Morphology Operations

Erosion: The erosion of gray scale image X by structuring element Y is symbolized by $X \ominus Y$ is described as

$$X \ominus Y(p, q) = \min_{s, t \in Y} \{X(p + s, q + t)\} \quad (1)$$

where p, q, s and t are pixels values in X and Y respectively.

Dilation: It is a dual operation to erosion. It is derived from erosion by set complementation. The dilation of gray scale image X by Y is symbolized by $X \oplus Y$, is described as

$$X \oplus Y(p, q) = \max_{s, t \in Y} \{X(p - s, q - t)\} \quad (2)$$

Opening: The opening of binary image X by structuring element Y is denoted by $X \circ Y$ and defined as erosion followed dilation

$$X \circ Y = (X \ominus Y) \oplus Y \quad (3)$$

Closing: The closing operation of binary image X by structuring element Y is denoted by $X \bullet Y$ and defined as dilation followed by erosion.

$$X \bullet Y = (X \oplus Y) \ominus Y \quad (4)$$

Top-hat transformation: The top-hat transformation of a gray scale image I is defined as subtracting its opening from itself.

$$T_{hat}(I) = I - (I \circ b) \quad (5)$$

Bottom-hat transformation: The bottom-hat transformation of gray scale image I is I defined as subtracting the image from its closing operation.

$$B_{hat}(I) = (I \bullet b) - I \quad (6)$$

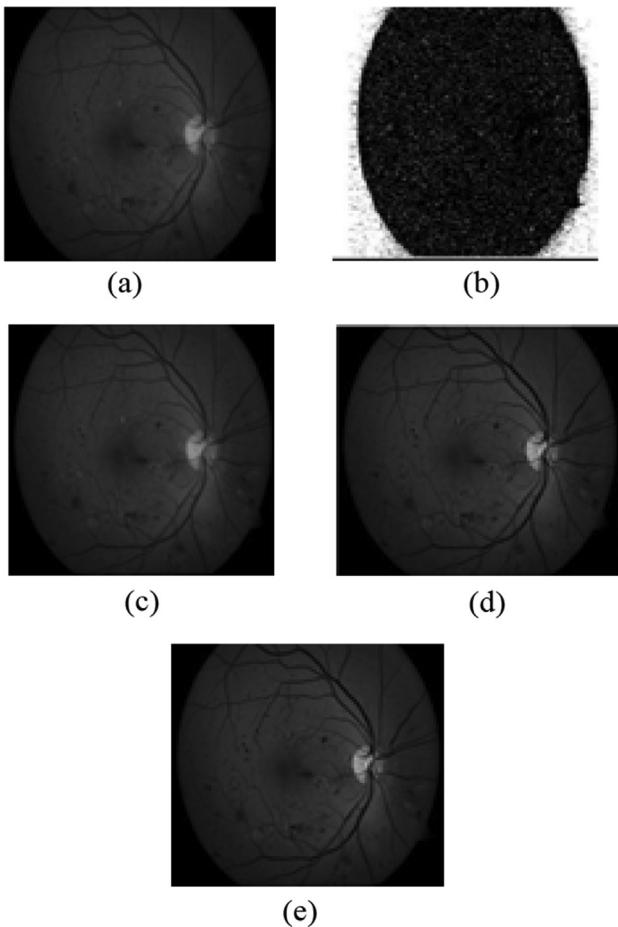


Fig. 3. (a) Green channel image (b) Regional minima image (c) Reconstructed image (d) Enhanced image (e) Smoothed image.

where b is structuring element.

3.2. Detection of blood vessels using iterative morphological approach

Green channel of RGB image is preferred for detection in blood vessel from the fundus image as shown in Fig. 3(a).

The main idea is the identification of foreground and background. We assume foreground as high intensity structures like optic disc and exudates, and background as low intensity structures like MAs, hemorrhages, and blood vessels. Background estimation has been accomplished using Regional Minima (RMIN) operator which emphasize the regional minima pixel, followed by reconstruction by dilation. By identifying the regional minima, the bright area, such as exudates are removed from the green channel image. Bright lesions are segmented from resulting image I_1 as shown by Eq. (7).

$$I_1 = R_{I_g}(RMIN(I_g)) \quad (7)$$

where $RMIN(I_g) = R^*_{I_g}(I_g + 1)$ and $R_{I_g}(RMIN(I_g))$ are reconstruction by dilation using I_g as mask image and $RMIN(I_g)$ as marker image. $R^*_{I_g}(I_g + 1)$ is gray scale reconstruction by erosion using I_g as mask and $(I_g + 1)$ as marker image. The regional minima and reconstructed images are shown in Fig. 3(b) and (c) respectively [24].

Now morphological contrast enhancement method is applied on I_1 image is described by Eq. (8)

$$I_2 = I_1 + \alpha_{th}(I_1) - \beta_{th}(I_1) \quad (8)$$

where $\alpha_{th}(I_1)$ is top hat by opening which contains only high intensity regions of I_1 and $\beta_{th}(I_1)$ is top hat by closing which contains low intensity regions. Then summing up the actual image with top hat

opening α_{th} and subtracting top hat by closing β_{th} , the enhancement of bright constituents are procured. Here ball shaped structuring element with 12-pixel radius and reference height of gray level is used. Enhanced image is shown in Fig. 3(d). Then Butter-worth Low Pass filter of order one is employed to diminish the impact of noise and MAs in detection process. The Butterworth filter of order one is defined by Eq. (9). Now, this filter is applied to I_2 as described in Eq. (10). The resulting smoothed images is shown in Fig. 3(e).

$$H(u, v) = \frac{1}{1 + [D(u, v)/D_0]^2} \quad (9)$$

$$I_3 = I_2 * H(u, v) \quad (10)$$

Next, top-hat transform by closing is applied to enhance the low intensity structure. The low intensity region are blood vessels, MAs and hemorrhages. We enhance the local minima and obtain a binary mask with retinal vein. Top-hat by closing is performed on image I_3 with structuring element B according to Eq. (11). Here, B is diamond shaped structuring element bigger than the maximum thickness of blood vessels. Here diamond shape structuring element with side of 14 pixels has been considered.

$$I_4 = \theta_{TH}^{(B)}(I_3) \quad (11)$$

Now supremum of opening of I_4 is calculated by set of linear structuring element. In this work 24 linear structuring elements are used with 100 pixels length. Linear structuring element is mapped to suit the foremost vessels in right and left eyes [27].

$$I_5 = \bigcup_{R_\phi=1}^{12} \alpha_{R_\phi}^{B_{R_\phi}}(I_4) \bigcup \bigcup_{L_\phi=1}^{12} \alpha_{L_\phi}^{B_{L_\phi}}(I_4) \quad (12)$$

where B_{R_ϕ} is the structuring element utilized to recognise the superior and inferior of right eye and employed angle of gyration.

$$R_\phi = [15 \ 30 \ . \ . \ 90 \ , \ -15 \ -30 \ . \ . \ -90] \quad (13)$$

B_{L_ϕ} is the structuring element employed to the superior and inferior part of principal vein of left eye.

$$L_\phi = [105 \ 120 \ . \ . \ 180 \ , \ 195 \ 210 \ . \ . \ 270] \quad (14)$$

The resulting image of supremum of opening of right eye is shown in Fig. 4(b).

Now reconstruction by dilation is performed on supremum of opening in order to estimate the blood vessels. Reconstructed image of blood vessels is shown in Fig. 4(c).

$$I_6 = R_{I_4}(I_5) \quad (15)$$

where I_4 is mask image and I_5 is marker image. I_6 is gray scale image, it is binarized in which blood vessel pixels consist of 1 and remaining all pixels consist of 0. We binarise I_6 using thresholding technique. The pixel value greater than specified threshold is assigned 1 otherwise 0. The threshold is taken as 8; which has been determined by trial and error experiments. The blood vessels in binary image has been shown in Fig. 4(d).

Now these blood vessels have to be removed from fundus image. For this purpose, firstly CLAHE is applied on green channel image 5 (a) and resulting image is shown in Fig. 5 (b). Green channel CLAHE image was chosen because it has the highest contrast of the blood vessels [21].

$$I_7 = CLAHE(I_g) \quad (16)$$

Further, ASF is performed on CLAHE image (I_7). This is based on successive applications of morphological openings and closings. A disc-shaped structuring element B with radius of 15 pixels was used for this purpose. ASF resulting image is shown in Fig. 5(c). Now CLAHE image (I_7) is subtracted from ASF image (I_8), which is shown in Fig. 5(d).

$$I_8 = \varphi^s(\gamma^s(I_7)) \quad (17)$$

$$\{I_9\} = \{I_8\} - \{I_7\} \quad (18)$$

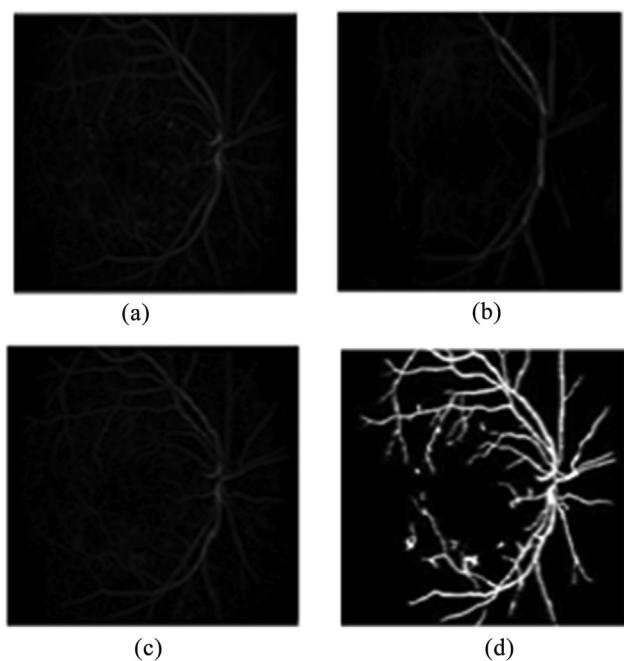


Fig. 4. (a) Top-hat by closing image (b) Supremum of opening image (c) Reconstructed image (d) Blood vessels in binary image.

H-MINIMA (H_{min}) transform is applied on I_9 image in order to eliminate the regions with low contrast in image I_9 . They could be optic disc or exudates. H_{min} transform eliminates the connected components with contrast less than specified threshold. The resulting image of H_{min} transform is shown in Fig. 5(e).

$$I_{10} = H \min(I_9) \quad (19)$$

Finally, regional minima (RMIN) operator is applied on resulting image I_{10} . RMIN operator based on morphological reconstruction by erosion, which converts the grayscale image into binary image without using any threshold. The pixel value of I_{11} is inverted as shown in Fig. 5(f).

$$I_{11} = RMIN(I_{10}) \quad (20)$$

Now blood vessel removal is achieved by subtracting the binary blood vessel image I_6 from RMIN operator image I_{11} . Some of the blood vessels residues are left in the eye image I_{12} which is shown in Fig. 6(a). The above morphological operation steps are applied on the remaining blood vessels residues. Residues after one more iteration are shown in Fig. 6(b). After comparing Fig. 6(a) and (b), it is clearly observed that improved blood vessels removal is achieved after one more iteration of blood vessel segmentation process.

$$I_{12} = imsubtract(I_{11}, I_6) \quad (21)$$

The steps for the detection of blood vessels are summarized below:

- Step 1. Extract green channel from color fundus image.
- Step II. Apply regional minima on green channel image.
- Step III. Perform reconstruction by dilation regional minima image.
- Step IV. Contrast enhancement is accomplished by morphological operator and followed by smoothing.
- Step V. Perform top-hat transform by closing.
- Step VI. Apply 24 times morphology opening on I_4 image with 15° rotation from vertical.
- Step VII. Perform morphology reconstruction by dilation.
- Step VIII. Apply regional minimum operator and binaries it.

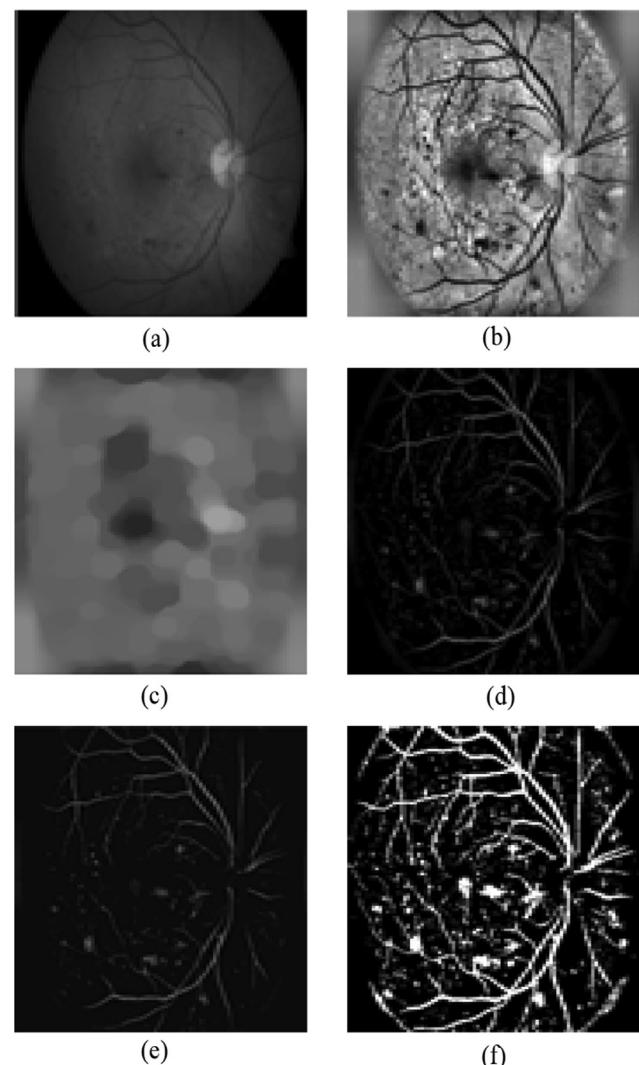


Fig. 5. (a) Green channel image (b) CLAHE image (c) ASF image (d) I_9 image (e) H_{min} operator image (f) Regional minima image.

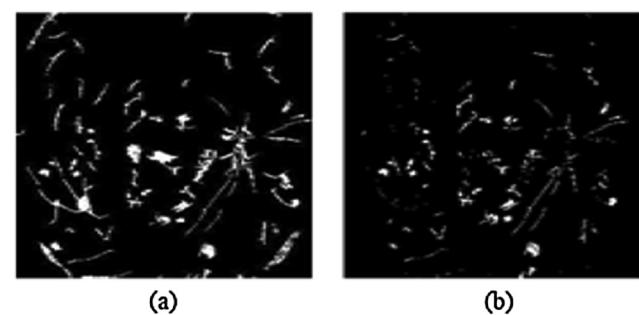


Fig. 6. (a) Residue after first time removal of blood vessels (b) Residue after one more iteration of removal algorithm of blood vessels.

3.3. Optic disc localization by Watershed segmentation method

Segmentation of optic disc plays essential role in detection of red lesions because some of the hemorrhages are very similar in size and color of optic disc. We have used Watershed Transform (WT) controlled with markers for optic disc segmentation from fundus photograph. The watershed transform is one of the methods in area of topography, widely used for morphological segmentation. It is marker based segmentation. Once marker is extracted precisely, segmentation can be

achieved correctly [28]. Watershed transform is fast, simple and intuitive method. It efficiently combines elements from similarity and discontinuity based methods. Conventional edge based methods most frequently form disconnected boundaries that need post-processing to produce closed regions. Watershed transform provides closed contour and it requires low computation time. It is able to accomplish complete partition of the image. Watershed algorithm is developed originally with grey scale image; hence it can be applied directly on pre-processed grey scale fundus image. WT for optic disc segmentation consists of many steps such as morphological gradient, marker-controlled watershed segmentation, erosion-based gray-scale image reconstruction, dilation -based gray-scale image reconstruction.

Morphological Gradient (MG): The gray scale image is changed into gradient image that specifies boundary quality of pixel. Limit is set with a specific goal, in order to convert every pixel to the edge point or non-edge point. A multiscale gradient algorithm is applied in order to achieve greater robustness to noise. The term multiscale refers to inspection of the images with different size structuring elements. The integration of MG in different scales, has strong tolerance to noise and it extracts a verity of finenesss of the boundaries [25].

$$MG(I) = (I \oplus b) - (I \ominus b) \quad (22)$$

where I is gray scale image and b is disc structuring element with 3 pixel radius.

Marker-controlled Watershed Segmentation: The marker controlled watershed transform is powerful and open-ended method for sectionalisation of entity with secured outlines. This method is very effective to reduce the over segmentation of gray scale image, if one knows how to place marker inside the image. The marker image employed in watershed segmentation is a binary image containing isolated points or large regions, where every attached marker is situated inside the object of concern. Every marker has coordinated relationship to watershed areas, henceforth, the quantity of markers will be equivalent to the last bit of watershed area [25].

After calculating morphological gradient of gray scale image, next step is creating internal (foreground) marker and external (background) marker. We used erosion-based gray scale image reconstruction followed by dilation-based gray scale reconstruction. the result is shown in Fig. 7(c).

Erosion-based gray-scale image reconstruction [17]

The morphological reconstruction of I_8 from the marker f with short flat structuring element D , is represented as

$$I_8 \Delta_D f = (f \ominus I_8 D)^\infty \quad (23)$$

where $(f \ominus I_8 D) = (f \ominus D) \vee I_8$ is called conditional (geodesic) erosion. Now the n conditional erosion

$$(f \ominus I_8 D)^n = (((f \ominus I_8 D) \ominus I_8 D) \ominus I_8 D) \dots \ominus I_8 D \quad (24)$$

Dilation-based gray scale reconstruction [17]

The morphological reconstruction of I_8 from the marker f with short flat structuring element D is

$$I_8 \Delta_D f = (f \oplus I_8 D)^\infty \quad (25)$$

where $(f \oplus I_8 D) = (f \oplus D) \wedge I_8$ is called conditional (geodesic) dilation. The n conditional dilations is given as.

$$(f \oplus I_8 D)^n = (((f \oplus I_8 D) \oplus I_8 D) \oplus I_8 D) \oplus I_8 D \dots \oplus I_8 D \quad (26)$$

where $I_8 \Delta_D f$ is infinite reconstruction of I_8 from the marker f , using the connectivity given by the structuring element D . $I_8 \hat{f}$ is pixel wise minimum between two images. $I_8 \vee f$ is pixel wise maximum between two images.

In the next step, *regional maxima* are applied on resulting image. Reconstruction by opening will take place in order to create the internal marker. Internal marker is superimposed on original image. The external marker is created around the optic disc of predefined radius centre. The internal marker image and superimposed image are shown

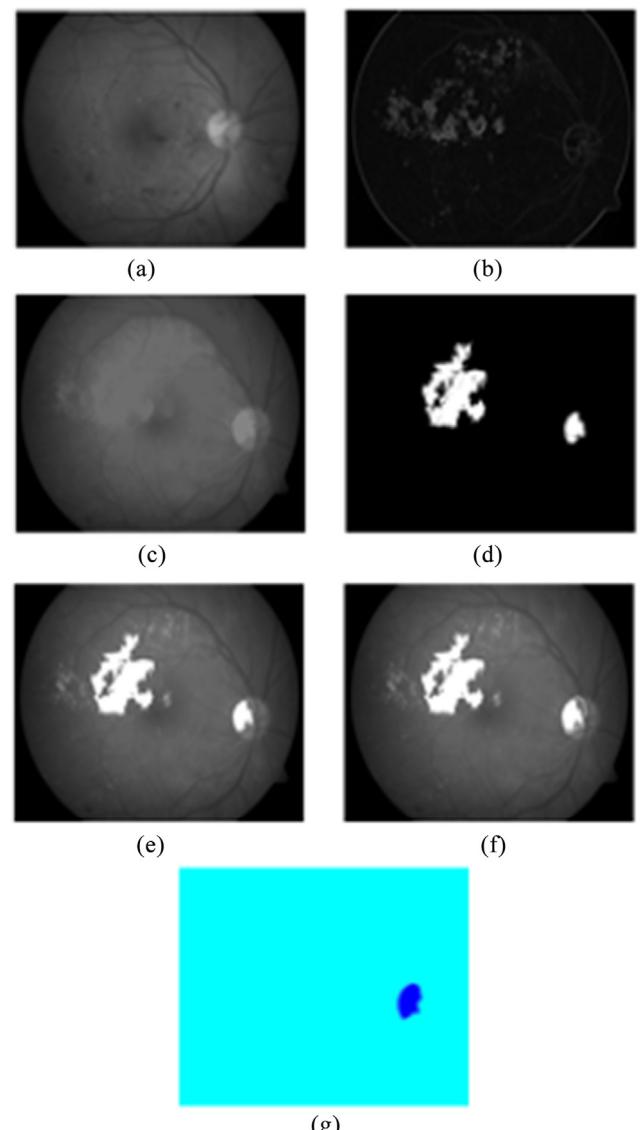


Fig. 7. (a) Gray scale image (b) Gradient image (c) Image after erosion based reconstruction followed by dilation based reconstruction (d) Regional maxima image (e) Regional maxima image superimposed on original image (f) Segmented optic disc (g) Segmented optic disc in coloured image.

on Fig. 7(d) and (e) respectively. Then watershed transform is applied on superimposed image for segmentation of optic disc. Only optic disc is segmented from grey scale image because of external marker. The optic disc is shown by watershed ridge lines in Fig. 7(f). Finally segmented optic disc in coloured image is shown in Fig. 7(g).

3.4. Fovea localization

Fovea is dark region inside the macula which is responsible for sharp central vision. Fovea has similar colour to some of DR lesions. It is difficult to differentiate fovea and some large haemorrhages, because of same colour. Hence, it is essential to segment fovea from the fundus image in order to detect diabetic retinopathy. Mathematical morphology is applied on gray scale image for fovea localization. Fovea extraction images are shown in Fig. 8.

The steps for detection of fovea are summarised as:

Step 1 *Opening by reconstruction* is tested on grey scale image with disc shape structuring element.

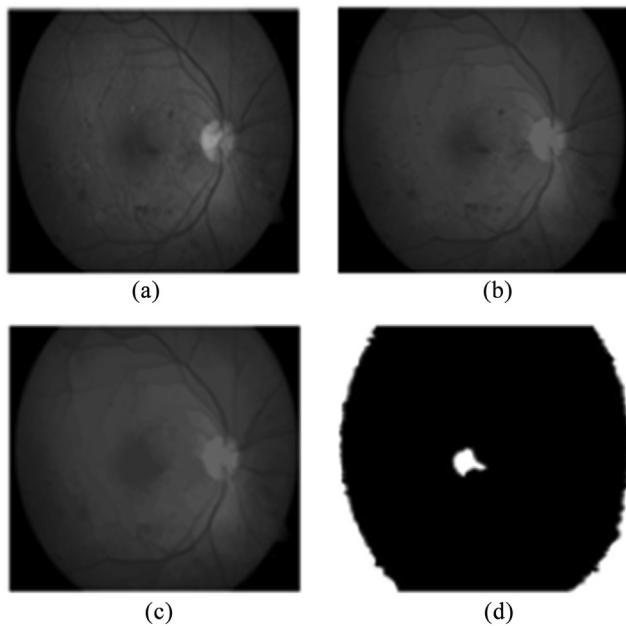


Fig. 8. (a) Gray scale image (b) Opening by reconstruction image (c) Opening closing by reconstruction image (d) Segmented fovea image.

Step II *Closing by reconstruction* is tested on resulting image of opening by reconstruction. These operations create the flat minima inside the image.

Step III Regional minima are applied on opening closing reconstruction image and erosion is performed on the resulting image with disc shape structuring element.

Now all the pathological structures, namely vascular tree, optic disc and fovea which may cause false positive, are removed from the fundus image. The remaining pathological structures are microaneurysm and haemorrhages. Extracted haemorrhages and MAs from colour fundus image are shown in Fig. 9.

3.5. Feature extraction of micoaneurysm and hemorrhages

After extracting microaneurysms and hemorrhages from fundus image, feature selection takes place in order to acquire relevant information from the given image. Since hemorrhages have particular properties, for example, round shape and red shading [29,30]. Feature extraction is very useful in artificial intelligence. The parameters of radial basis function neural network are trained by the features of the MAs and hemorrhages. Features such as area, perimeter, circularity,



Fig. 9. Extracted MAs and hemorrhages.

number of lesions, major axis length, minor axis length, and aspect ratio have been extracted [31,32]. Enhanced robustness of the proposed DR detection model is achieved by extracting more number of relevant features. Extracted features of MAs and hemorrhages, and their descriptions are shown in Table 1.

3.6. Classification of diabetic retinopathy using radial basis function neural network (RBF NN) [33]

RBF is a special case of single-hidden layer neural network for application to problems of supervised learning which is widely used for classification and regression. The model of RBF neural network is shown in Fig. 10. Radial Basis Function NN is a three-layer network, including input and output layer. Therefore, design of RBF NN is much simpler as compared to many other classifiers. Since, it has only one hidden layer, computational complexity of RBF classifier is lesser than feed forward NN based classifiers. RBF NN has strong tolerance to input noise, and online learning ability. Further, RBFNN has very good generalization properties. The properties of RBF networks make it very suitable to regression model as well as classification. The expression of RBFNN model with K centres, is described by Eq. (27). The network consists of three parameters, namely- output layer weights, centres and standard deviation. Output layer weights are linear parameters that regulate the height of the basis function and hence, RBFNN is known as linear model. Centres are non-linear parameters of the unrevealing layer neurons, regulate the location of the basis function. Standard deviation is nonlinear parameter of the hidden layer. It dictates the spread of the basis function. The number of centres regulate the architecture model because the number of neurons in hidden layer is equivalent to the number of centre. These all three parameters of network would be trained by the features extracted from the MAs and hemorrhages.

$$F(x) = \sum_{k=1}^K w_k \exp\left(-\frac{\|x_n - \mu_k\|^2}{\sigma_k^2}\right) \quad (27)$$

It is rewritten as

$$F(x) = \sum_{k=1}^K w_k \phi_k \quad (28)$$

$$\text{where } \phi(x) = \exp\left(-\frac{\|x_n - \mu_k\|^2}{\sigma_k^2}\right)$$

Choosing the centres: The centres of network are chosen by K-mean clustering. It iteratively minimizes the distance between data and centres as described in Eq. (29). K-mean clustering is an example of unsupervised learning. Lloyd algorithm for convergence of centre at local minima is described by the Eqs. (30) and (31) [34,35].

Iteratively minimized the objective function

$$\sum_{k=1}^K \sum_{x_n \in S_k} \|x_n - \mu_k\|^2 \quad (29)$$

with respect to S_K and μ_K . update centre (μ_K) and cluster (S_K)

$$\mu_k \leftarrow \frac{1}{|S_k|} \sum_{x_n \in S_k} x_n \quad (30)$$

$$S_k \leftarrow \{x_n : \|x_n - \mu_k\| \leq \text{all } \|x_n - \mu_k\|\} \quad (31)$$

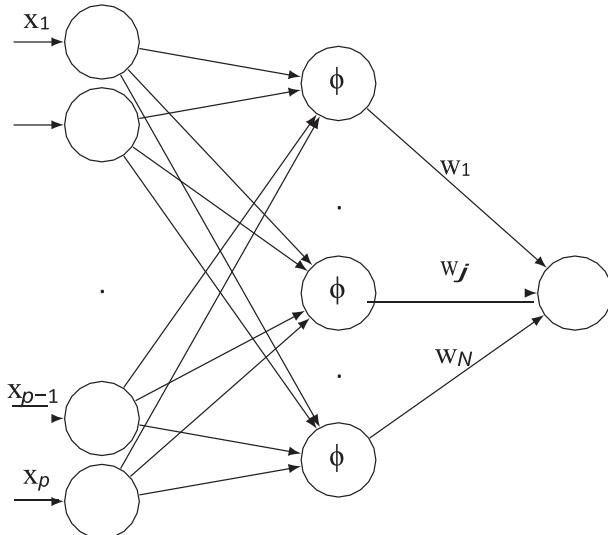
Finally, the Lloyd algorithm converges into local minima and optimum centres are achieved.

Choosing the weight: RBF network with K centres has been chosen for classification. The number of inputs are more than the number of weights of the network, as shown in Eq. (32). The number of input indicates the number of equations, which are more than number of unknowns (weights). Such systems are called as *over determined* systems, which do not have exact solution. Therefore, we go for the approximate solution by applying *Least Square* or *Pseudo Inverse* methods.

Table 1

Extracted features of MAs and hemorrhages for proposed DR detection.

Feature	Description
Area of object	The aggregate number of pixel in the area
Perimeter of object	Add up all the pixels on perimeter of object
Circularity	Roundness of the candidate
Number of objects	Total number of discontinuity in the region
Major axis length of the candidates	Total length of the major axis of the ellipse that has the identical normalized second central moment of the zone
Minor axis length of the candidates	Total length of the minor axis of the ellipse of the identical normalised second central moment of the region
Aspect ratio	Ratio of major axis length and minor axis length

**Fig. 10.** RBF NN model.

Least square solution minimizes the addition of all of the squares of errors produced in the result of each isolated equation. The solution of least square problem is achieved by normal equation, which is described in Eq. (33) [29,30].

$$y_n \approx \sum_{k=1}^K w_k \exp\left(-\frac{\|x_n - \mu_k\|^2}{\sigma_k^2}\right) \quad (32)$$

N equation in K \ll N unknowns. If $\phi^T \phi$ is invertible

$$w = [\phi^T \phi]^{-1} \phi^T y \quad (33)$$

Standard deviation (σ) is calculated using back propagation algorithm. The number and dimension of standard deviation (σ) are same as centres of the neural network. In this work, we have chosen eight centres, which are determined by trial and error method. Hence, the number of neurons in hidden layer are eight. The parameters of RBF network are trained by considering extracted features (Table 1) as *input* and *class* as output. Here, binary classification (DR and healthy) has been performed for the detection of DR.

4. Results

The results of proposed DR detection model from fundus image have been discussed in details in this section. The accuracy of result obtained from the suggested DR detection algorithm is based on evaluated sensitivity and specificity. One hundred thirty images (Healthy and DR images) have been taken from DIABET DB1 and Diaretbdb0 v 1 1 database. Out of these, sixty seven eye images have been used as training samples and sixty three images as testing samples. Number of samples in training set and test set are chosen randomly, which provides robustness to the proposed algorithm. The number of training samples are chosen more than the test samples, for better training of classification model. Table 2 shows the retinal fundus images under investigation and

the corresponding output images containing MAs and hemorrhages along with some blood vessel residues. RBFNN based classifier used in the proposed work, classifies as 'DR' and 'healthy' images. For the validation of the diagnostic results obtained from the proposed DR detection system, diagnostic opinion has been obtained from ophthalmologist. Comparison table of proposed method and ophthalmologist opinion is shown in Table 3. In this table, inference generated by proposed method and inference drawn by ophthalmologist are shown. Table 3 infers that proposed method has very good match with ophthalmologist opinion for early detection of DR. Out of 10 images investigated during the experiment, 6 were DR affected and 4 were healthy eyes. It is observed that DR affected retinal fundus image contains nearly circled bright spots namely haemorrhages & MAs, and some blood vessels residues. However, in case of healthy eye fundus image, only some vascular residues are present.

The number of true positive, true negative, false positive and false negative of the proposed DR detection method are twenty nine, twenty eight, four and two respectively. The sensitivity and specificity of proposed system have been computed as 87% and 93% respectively. The obtained results of the proposed method are compared with some existing methods. The comparison is shown in Table 4, which states that the proposed method is better than the considered existing methods.

5. Conclusion

This paper presented a scheme for the early detection of DR from colour fundus image for abundance screening of DR. The achieved values of sensitivity and specificity show that the proposed diagnostic system is better for non-proliferative diabetic retinopathy detection. The DR detection model was trained by seven features of haemorrhages and MAs, which made the proposed system robust and reliable. Future work of this paper is to propose a proliferative diabetic retinopathy detection system by considering cotton wools and abnormal blood vessels from colour fundus images. DR detection system could be extended to multi class diabetic retinopathy classification, to classify into healthy, mild NDPR, moderate NDPR, severe NDPR, and PDR. For classification, Feed Forward Neural Network, Probabilistic Neural Network (PNN), SVM and other emerging models can be examined.

Acknowledgement

Authors would like to acknowledge Young Faculty Research Fellowship (YFRF) scheme under Visvesvaraya PhD programme of Ministry of Electronics and Information Technology, Meity Government of India, for providing research grant. Authors would also like to thank Prof. M. K. Singh, HOD Ophthalmology Department, BHU Varanasi for validation of proposed algorithm.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.optlastec.2019.105815>.

Table 2
Fundus image and corresponding output.

S. No	Fundus Image	Extracted MAs and Hemorrhages
1		
2		
3		
4		
5		
6		
7		

(continued on next page)

Table 2 (continued)

S. No	Fundus Image	Extracted MAs and Hemorrhages
8		
9		
10		

Table 3

Comparison of the diagnostic conclusion made by the proposed system and ophthalmologist opinion.

S. No	Diagnostic decision based on features presence			Inference drawn by ophthalmologist
	Microaneurysm	Hemorrhages	Inference generated by proposed scheme	
1	Yes	Yes	DR presence	MAs, Hemorrhages
2	Yes	Yes	DR presence	MAs, Hemorrhages, exudate
3	Yes	Yes	DR presence	MAs, Hemorrhages, exudates and cotton wools
4	Yes	Yes	DR presence	MAs, Hemorrhages, exudate
5	Yes	Yes	DR presence	MAs, Hemorrhages, exudate and cotton wools
6	Yes	Yes	DR presence	MAs, Hemorrhages
7	Yes	Yes	DR presence	Hemorrhages, lasers
8	No	No	DR absence	Healthy eye
9	No	No	DR absence	Healthy eye
10	No	No	DR absence	Healthy eye

Table 4

Comparison result of proposed method with existing method.

	Sensitivity	Specificity
Kedir M. Adal et al. [6]	81.08%	92.3%
J.P. Bea et al. [9]	82.9%	
S. S. Rahim et al. [2]	80%	55%
Proposed Method	87%	93%

References

- [1] S.B. Junior, D. Welfer, Automatic detection of microaneurysms and hemorrhages in color eye fundus images, *Int. J. Comp. Sci. Inform. Technol.* 5 (5) (2013) 21.
- [2] S.S. Rahim, C. Jayne, V. Palade, J. Shuttleworth, Automatic detection of microaneurysms in colour fundus images for diabetic retinopathy screening, *Neural Comput. Appl.* 27 (5) (2016) 1149–1164.
- [3] Meindert Niemeijer, Bram Van Ginneken, Michael J. Cree, Atsushi Mizutani, Gwéolod Quellec, Clara I. Sánchez, Bob Zhang, et al., Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs, *IEEE Trans. Med. Imaging* 29 (1) (2009) 185–195.
- [4] D. Marin, M.E. Gegundez-Arias, A. Suero, J.M. Bravo, Obtaining optic disc center

and pixel region by automatic thresholding methods on morphologically processed fundus images, *Comput. Methods Programs Biomed.* 118 (2) (2015) 173–185.

[5] S.S. Rahim, V. Palade, J. Shuttleworth, C. Jayne, Automatic screening and classification of diabetic retinopathy fundus images, *International Conference on Engineering Applications of Neural Networks*, Springer, 2014, pp. 113–122.

[6] Kedir M. Adal, Désiré Sidibé, Sharib Ali, Edward Chaum, Thomas P. Karnowski, Fabrice Mériaudeau, Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning, *Comput. Methods Programs Biomed.* 114 (1) (2014) 1–10, <https://doi.org/10.1016/j.cmpb.2013.12.009>.

[7] R. Inbarathi, R. Karthikeyan, Detection of retinal hemorrhage in fundus images by classifying the splat features using svm, *Int. J. Innov. Res. Sci., Eng. Technol.* 3 (2014) 1979–1986.

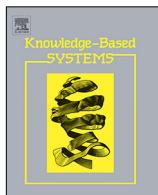
[8] P. Jitpakdee, P. Aimmanee, B. Uyyanonvara, A survey on hem- orrage detection in diabetic retinopathy retinal images, *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, IEEE, 2012, pp. 1–4.

[9] J.P. Bae, K.G. Kim, H.C. Kang, C.B. Jeong, K.H. Park, J.M. Hwang, A study on hemorrhage detection using hybrid method in fundus images, *J. Digit. Imaging* 24 (3) (2011) 394–404.

[10] A.D. Fleming, S. Philip, K.A. Goatman, J.A. Olson, P.F. Sharp, Automated microaneurysm detection using local contrast normalization and local vessel detection, *IEEE Trans. Med. Imaging* 25 (9) (2006) 1223–1232.

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

- [12] T. Walter, J.-C. Klein, P. Massin, A. Erginay, A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina, *IEEE Trans. Med. Imaging* 21 (10) (2002) 1236–1243.
- [13] R.P. Danis, M.D. Davis, *Proliferative diabetic retinopathy*, Springer, 2008, pp. 29–65.
- [14] K.B. Shah, D.P. Han, *Proliferative diabetic retinopathy*, *Int. Ophthalmol. Clinics* 44 (4) (2004) 69–84.
- [15] R. Welikala, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T.H. Williamson, S. Barman, Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification, *Comput. Methods Programs Biomed.* 114 (3) (2014) 247–261.
- [16] D.R.S.R. Group, et al., Photocoagulation treatment of proliferative diabetic retinopathy: clinical application of diabetic retinopathy study (drs) findings, drs report number 8, *Ophthalmology* 88 (7) (1981) 583–600.
- [17] E.R. Dougherty, R.A. Lotufo, *Hands-on Morphological Image Processing*, SPIE press, 2003.
- [18] J. Serra, P. Soille, *Mathematical Morphology and its Applications to Image Processing*, Springer Science & Business Media, 2012.
- [19] R. Radha, B. Lakshman, Identification of retinal image features using bitplane separation and mathematical morphology, 2014 World Congress on Computing and Communication Technologies (WCCCT), IEEE, 2014, pp. 120–123.
- [20] S.H. Rasta, M.E. Partovi, H. Seyedarabi, A. Javadzadeh, A comparative study on preprocessing techniques in diabetic retinopathy retinal images: illumination correction and contrast enhancement, *J. Med. Signals Sensors* 5 (1) (2015) 40.
- [21] S.K. Yadav, S. Kumar, B. Kumar, R. Gupta, Comparative analysis of fundus image enhancement in detection of diabetic retinopathy, 2016 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), IEEE, 2016, pp. 1–5.
- [22] V. Thirilogasundari, S.A. Janet, et al., Fuzzy based salt and pepper noise removal using adaptive switching median filter, *Procedia Eng.* 38 (2012) 2858–2865.
- [23] A. Hussain, M. Habib, A new cluster based adaptive fuzzy switching median filter for impulse noise removal, *Multimedia Tools Appl.* 76 (21) (2017) pp. 22 001–22 018.
- [24] D. Welfer, J. Scharcanski, C.M. Kitamura, M.M. Dal Pizzol, L.W. Ludwig, D.R. Marinho, Segmentation of the optic disk in color eye fundus images using an adaptive morphological approach, *Comput. Biol. Med.* 40 (2) (2010) 124–137.
- [25] K. Parvati, P. Rao, M. Mariya Das, Image segmentation using gray-scale morphology and marker-controlled watershed transformation, *Discrete Dynamics Nature Soc.* 2008 (2008).
- [26] S. Kumar, B. Kumar, Diabetic retinopathy detection by extracting area and number of microaneurysm from colour fundus image, 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), IEEE, 2018, pp. 359–364.
- [27] E. Ricci, R. Perfetti, Retinal blood vessel segmentation using line operators and support vector classification, *IEEE Trans. Med. Imaging* 26 (10) (2007) 1357–1365.
- [28] Niket Amoda, Ramesh K. Kulkarni, Efficient Image Segmentation Using Watershed Transform, *Int. J. Comp. Sci. Technol. (IJCST)* iv (ii), ver 2.
- [29] Lama Seoud, Thomas Hurtut, Jihed Chelbi, Farida Cheriet, J.M. Pierre Langlois, Red lesion detection using dynamic shape features for diabetic retinopathy screening, *IEEE Trans. Med. Imaging* 35 (4) (2016) 1116–1126.
- [30] Behdad Dashtbozorg, Jiong Zhang, Fan Huang, M. Bart ter Haar Romeny, Retinal microaneurysms detection using local convergence index features, *IEEE Trans. Image Process.* 27 (7) (2018) 3300–3315.
- [31] Wen Cao, Nicholas Czarnek, Juan Shan, Lin Li, Microaneurysm detection using principal component analysis and machine learning methods, *IEEE Trans. Nanobiosci.* 17 (3) (2018) 191–198.
- [32] Ramon Pires, Sandra Avila, Herbert F. Jelinek, Jacques Wainer, Eduardo Valle, Anderson Rocha, Beyond lesion-based diabetic retinopathy: a direct approach for referral, *IEEE J. Biomed. Health. Inf.* 21 (1) (2017) 193–200.
- [33] O. Nelles, *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*, Springer Science & Business Media, 2013.
- [34] R.S. Michalski, J.G. Carbonell, T.M. Mitchell, *Machine learning: An artificial intelligence approach*, Springer Science & Business Media, 2013.
- [35] C.M. Bishop, et al., *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.



Automated identification and grading system of diabetic retinopathy using deep neural networks



Wei Zhang^a, Jie Zhong^{b,*}, Shijun Yang^b, Zhentao Gao^a, Junjie Hu^a, Yuanyuan Chen^a, Zhang Yi^a

^a Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610065, China

^b The Department of Ophthalmology, Sichuan Academy of Medical Sciences and Sichuan Provincial Peoples Hospital, Chengdu 610072, China

HIGHLIGHTS

- We established a high-quality labelled dataset of DR medical images.
- We developed a novel and well-performing DR recognition and classification system.
- The optimal combination of ensemble model was explored experimentally.

ARTICLE INFO

Article history:

Received 31 October 2018

Received in revised form 14 March 2019

Accepted 15 March 2019

Available online 22 March 2019

Keywords:

Deep learning
Diabetic retinopathy
Ensemble learning
Fundus images
Image classification
Transfer learning

ABSTRACT

Diabetic retinopathy (DR) is a major cause of human vision loss worldwide. Slowing down the progress of the disease requires early screening. However, the clinical diagnosis of DR presents a considerable challenge in low-resource settings where few ophthalmologists are available to care for all patients with diabetes. In this study, an automated DR identification and grading system called DeepDR is proposed. DeepDR directly detects the presence and severity of DR from fundus images via transfer learning and ensemble learning. It comprises a set of state-of-the-art neural networks based on combinations of popular convolutional neural networks and customised standard deep neural networks. The DeepDR system is developed by constructing a high-quality dataset of DR medical images and then labelled by clinical ophthalmologists. We further explore the relationship between the number of ideal component classifiers and the number of class labels, as well as the effects of different combinations of component classifiers on the best integration performance to construct an optimal model. We evaluate the models on the basis of validity and reliability using nine metrics. Results show that the identification model performs best with a sensitivity of 97.5%, a specificity of 97.7% and an area under the curve of 97.7%. Meanwhile, the grading model achieves a sensitivity of 98.1% and a specificity of 98.9%. On the basis of the methods above, DeepDR can detect DR satisfactorily. Experiment results indicate the importance and effectiveness of the ideal number and combinations of component classifiers in relation to model performance. DeepDR provides reproducible and consistent detection results with high sensitivity and specificity instantaneously. Hence, this work provides ophthalmologists with insights into the diagnostic process.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Diabetic retinopathy (DR) is a chronic complication of diabetes that damages the retina. Notably, the risk of blindness in patients with DR is 25 times that in healthy people; thus, DR is a leading cause of blindness amongst people aged 20–65 years

worldwide [1]. The blindness caused by DR can be prevented through regular fundus examinations [2]. A widespread consensus regarding the benefits and cost-effectiveness of screening for DR has been formed amongst western nations [3–5]. Most DR studies use the international clinical disease severity scale to classify DR (Table 1) in accordance with the Early Treatment Diabetic Retinopathy Study (ETDRS). Other details are available in the latest American Association of Ophthalmology Clinical Guidelines: Diabetic Retinopathy (2016 Edition, 2017 Updated) [6]. Nowadays, diabetes screening is common in developed countries;

* Corresponding author.

E-mail addresses: zweiscu@gmail.com (W. Zhang), jiezhang1968maya@163.com (J. Zhong), odalisyang@gmail.com (S. Yang), gaozhentao@stu.scu.edu.cn (Z. Gao), hujunjiescu@gmail.com (J. Hu), chenyuanyuan@scu.edu.cn (Y. Chen), zhangyi@scu.edu.cn (Z. Yi).

patients with diabetes are screened from the general population and transferred to DR specialists. Follow-up examinations are performed by these specialists, and medical intervention is implemented when necessary; therefore, the incidence of severe DR in developed countries is low.

However, the situation in China is not as promising. (1) Currently, ophthalmologist-to-patient ratio is nearly 1:1000 in China. (2) The rate of DR screening is less than 10%. (3) The vast majority of patients with DR in China do not know the risks associated with DR; hence, they may not realise they have the disease. Furthermore, a large proportion of diabetic patients ignore this serious complication. Patients with DR in China often undergo late invasive treatment and exhibit serious illness, resulting in poor prognosis and high medical expenses. Therefore, the incidence of severe proliferative DR is much higher in China than in developed countries. Moreover, the eventual blindness from DR is irreversible, thereby placing a heavy burden on Chinese families and the society. The automatic screening and grading of DR is a pressing demand because it can help to solve the abovementioned problems.

Deep neural networks (DNNs), also called deep learning by brain-inspired systems [7,8], can automatically learn numerous abstract high-level features or representations of attribute categories directly from original big data to ascertain a distributed representation of data. A widely used type of DNN is the recurrent neural network, which has shown unprecedented success in academia and industries, including in the areas of speech recognition and machine translation [9–11]. With regard to the characteristics of the spatial coherence of images, convolutional neural networks (CNNs) are preferred because they are highly specialised in views for image recognition, analysis and classification [12]. In recent years, CNNs have also provided insights into various medical studies. Furthermore, they have abilities rivalling those of medical experts [13], especially when applied to skin cancer [14] and breast cancer classification [15] and lung cancer [16] and retinopathy of prematurity detection [17,18].

Nevertheless, challenges remain in the use of CNNs in medical studies. First, sufficient real-world medical images, especially those for some specialised diseases, are difficult to obtain. Furthermore, the availability of labelled medical data is typically limited. Second, DR features are so complex that they are likely to cross-effect with various other lesions, and the minute lesions of DR cannot be detected if images quality is poor. According to medical journals, fundus photographs are labelled by a manual operation process, which is prone to subjectivity. Third, high disease-detection accuracy is difficult to attain effectively by training a single model with a limited scale of medical image data and inevitable image noise. Therefore, two important strategies are used in deep learning: transfer learning [19,20] and ensemble learning [21–23]. The primary concept of transfer learning is knowledge reuse: the migration of big data to small data fields to resolve the problem of data and knowledge scarcity in small data fields. The major conception of classifier ensemble learning is the combination of a series of component classifiers with different learning preferences to resolve the same predictive problem. These ensemble methods enable increased generalisation that outperforms that of each individual component.

In the current work, we developed an automated system called DeepDR for DR screening via deep learning. DeepDR is a complex process composed of three steps: judgment of the existence of retinal lesion characteristics via screening of fundus photographs, evaluation of the severity of DR if lesion features are detected and reporting of the detection of clinical DR. Thus far, DeepDR has been used in some local hospitals to aid primary hospitals in remote areas or clinical communities that lack retinal specialists or appropriate equipment.

Table 1
International clinical DR disease severity scale.

Severity level	Findings observable upon dilated ophthalmoscopy
No DR	No abnormalities.
Mild DR	Microaneurysms only.
Moderate DR	More than just microaneurysms but less than severe NPDR.
Severe DR	Any of the following and no signs of proliferative retinopathy: (1) severe intraretinal hemorrhages and microaneurysms in each of four quadrants; (2) definite venous beading in two or more quadrants; (3) prominent IRMA in one or more quadrants.
Proliferative DR	One or both of the following: (1) neovascularisation; (2) vitreous/preretinal hemorrhage.

IRMA = intraretinal microvascular abnormalities; NPDR = nonproliferative diabetic retinopathy; PDR = proliferative diabetic retinopathy Note :(1) Any patient with two or more of the characteristics of severe NPDR is considered to have very severe NPDR. (2) PDR may be classified as high-risk and non-high-risk.

Table 2
The grading scale of DR in the study.

Grade	Disease severity level
NORMAL	No DR
NPDR	Mild NPDR to moderate NPDR
NPDR2PDR	Severe NPDR to non-high-risk PDR
PDR	High-risk PDR

The contributions of this work are as follows.

(1) We establish a high-quality labelled dataset of DR medical images.

(2) We develop a novel DR identification and grading system. The system performs well in comparison with human evaluation metrics.

(3) We explore the relationship between the number of ideal component classifiers and the number of class labels. Furthermore, the effects of different combination methods on the best integration performance are discussed.

In Section 2, we analyse the related works. In Section 3, we detail the dataset. In Section 4, we describe two novel ensemble models for the two respective tasks. In Section 5, we show the experiments on the two tasks. In Section 6, we provide a discussion of the entire study and future work. Finally, In Section 7, we draw the conclusions.

2. Related works

In the past few decades, the development of automated DR pathology screening has made encouraging progress. From an application perspective, computer-aided detection (CADe) algorithms and computer-aided diagnosis (CADx) algorithms can be viewed as typical representatives in the field. CADe detects lesions at the pixel level with manual segmentations [24]. On the basis of the detected lesions, CADx detects pathologies at the image level [25].

2.1. Traditional methods for DR diagnosis

From a methodological perspective, automated screening for DR has long focused on pattern recognition or traditional machine learning algorithms. Walter et al. [26] created efficient algorithms for detecting optic disc and exudates; these algorithms yielded a mean sensitivity of 92.8% and a mean predictive value of 92.4% on 30 images. Niemeijer et al. [27] developed a machine learning system capable of detecting exudates, cotton wool spots and drusen; their system can differentiate amongst 300 colour images, and its reporting performance approaches that of retinal

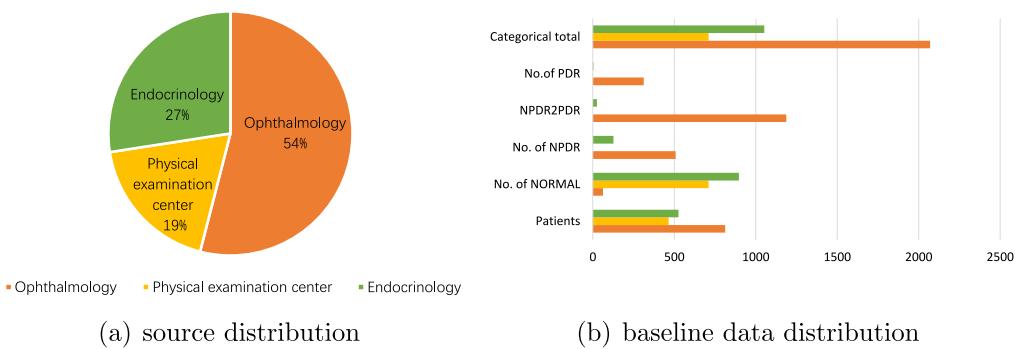


Fig. 1. Characteristics of the dataset for DR.

experts. Faust et al. [28] contributed an important and valuable review of algorithms used for extractions of these features from retinopathy images; they also discussed some reported classification systems. Akram et al. [29] presented a hybrid classifier comprising classifiers that could be used to detect all types of non-proliferative DR (NPDR) lesions and grade different stages of NPDR on the basis of a Gaussian mixture model and m-Medoids; whereas the area under the curve (AUC) values of m-Medoids and Gaussian mixture model are 97.7% and 96.3%, respectively, that of the hybrid classifier achieves reaches 98.1%. Gowda et al. [30] employed the fuzzy c-means clustering method on a dataset consisting of 100 images to identify the exact region of DR and reported an accuracy of 99.01%, sensitivity of 98.38% specificity of 96.36%.

Some obvious shortcomings of these techniques include the following. First, they focus on feature engineering, such that the extraction of features must be specified by experts; fulfilling such requirement is a time-consuming process and increases the burden on clinicians. Second, they show limited scope because the studies present results that were derived from small databases; thus, the generalisation and robustness of systems are limited to a certain extent. Third, the early clinical symptoms of DR are not always obvious, and the sizes of some lesions are insufficient to attract the attention of some graders. Hence, questions arise with regard to the accurate extraction of lesion features and comprehensive diagnosis. This approach is thus transformed into a complex vision issue, although this type of error is relatively understandable. Therefore, one of the ultimate goals of artificial intelligence is to automate this process of feature engineering as much as possible.

2.2. Deep learning for DR diagnosis

Deep learning, especially CNNs, provides powerful support to alleviate the aforementioned problems. Models trained by deep learning can discern subtle local features directly from retinopathy images without human effort or specific domain knowledge.

Litjens et al. [31] contributed an important survey regarding the use of deep learning in image classification, object detection, segmentation, registration and other tasks. They summarised over 300 contributions to the field, most of which appeared in 2016. All top algorithms in the Kaggle machine learning competition [32] in 2015 used CNNs to support an automated DR detection system. Benson et al. [33] repurposed an existing deep network for DR screening via transfer learning and other pre-processing techniques on 979 clinical cases; the repurposed deep network achieves a 95% AUC for identifying severe DR with equal sensitivity and specificity of 90%. Gargya et al. [34] developed a data-driven deep learning algorithm capable of classifying fundus images as healthy or not; this algorithm identifies relevant cases with high reliability. Gulshan et al. [35] created

a deep learning-trained algorithm for detecting referable diabetic retinopathy (RDR) in two separate datasets of 9963 and 1748 images; the algorithm achieved high sensitivity (97.5% and 96.1%) and specificity (93.4% and 93.9%) when applied to the two datasets. A study by researchers at Stanford University (Tamkin et al., [36]) used the InceptionV3 model via transfer learning techniques; the approach achieves 72.96% accuracy in detection of RDR and 92.59% accuracy in detection of stage PDR. Pratt et al. [37] designed a CNN network that enables a classifier to predict the exact DR stage of the sample for a five-class DR detection task. The proposed technique achieves 75% accuracy, 30% sensitivity and 95% specificity.

3. Dataset

3.1. Materials

In our study, macula-centred retinal fundus images were taken from the Sichuan Academy of Medical Sciences and Sichuan Provincial Peoples Hospital between September 2017 and May 2018. The original data comprising 13,767 images of 1872 patients were collected from three sources: ophthalmology, endocrinology and physical examination centres (Fig. 1). In general, almost all patients from the ophthalmology department were diagnosed with DR, and nearly two-thirds of the patients from the endocrinology department had DR; the data from the physical examination centre showed no DR symptoms amongst patients. As almost all patients from ophthalmology department had DR, two types of images, namely, retinal colour fundus photographs and fluorescein angiography fundus photographs (pharmacological pupil dilation) were required, and two or more photographs with a 45° view were captured per eye. This project aims to screen preoperative fundus retinal images and to diagnose the degree of lesions. Therefore, all fluorescence contrast images and postoperative fundus images, that is, 9934 images in 1229 patients, were excluded. For the patients from the endocrinology and physical examination centres, only fundus photos of each eye were taken. Further detailed statistics of the dataset are shown in Table 3.

3.2. Grading standard

Although ETDRS (Table 1) has indispensable reference value for our labelling work, some challenges remained in the grading of DR. First, the interpretation for the reference criteria primarily depends on the ophthalmologist's experience of reading images with reference to the guidelines and is thus qualitative rather than quantitative; thus, the assessment of severity has a degree of subjectivity. Second, most patients with DR in China often neglect this disease, and thus fail to secure timely interventions;

Table 3

Summary of training and validation datasets for DR.

Source	Camera	Assessors	Patients	Images	Images/eye	Normal	Npdr	Npdr2pdr	Pdr	Total ^a
Training										
OPH ^b	Canon	(1,2) ^c	613	1669	2~6 ^d	50	421	953	245	1669
PEC ^e	Canon	(1,2)	379	575	1~2 ^f	575	0	0	0	575
END ^g	Kowa	(1,2)	409	818	2	696	100	19	3	818
Total ^h			1401	3062		1321 (0.43)	521 (0.17)	972 (0.32)	248 (0.08)	3062
Validation										
OPH	Canon	(1,2)	199	401	2~6	12	87	234	68	401
PEC	Canon	(1,2)	86	136	1~2	136	0	0	0	136
END	Kowa	(1,2)	117	234	2	200	27	6	1	234
Total _v ⁱ			402	771		348 (0.45)	114 (0.15)	240 (0.31)	69 (0.09)	771
Total ^j			1803	3833		1619	635	1212	317	3833

^aTotal = Total of images in the department.^bOPH = Ophthalmology.^c(1,2) = 2 ophthalmologists, 1 retinal specialist.^d2~6 = Number of fundus photographs taken per eye per patient.^ePEC = Physical Examination Centre.^f1~2 = some patients had only 1 images per eye.^gEND = Endocrinology.^hTotal_t = Categorical total of training dataset.ⁱTotal_v = Categorical total of validation dataset.^jTotal = Categorical total of training and validation.

as a result, their cases become aggravated, with symptoms being frequently considered as intermediate to severe NPDR and early PDR. Third, because the transition period above has similar clinical manifestations, the recommended treatment for them, according to the ETDRS, is the same. This situation is also reflected in cases ranging from mild NPDR to moderate NPDR [6].

Therefore, we classified severity into four levels: normal, NPDR, NPDR2PDR and PDR. This annotation strategy was used in our four-class classification task. Table 2 shows the specific classifications.

3.3. Manual grading

All images of the dataset were assessed in stages as the data volume accumulated. The graders included one retinal specialist with more than 27 years of experience in DR research and two seasoned ophthalmologists with more than 5 years of experience in clinical diagnosis and treatment. The entire grading process was divided into three steps. First, the annotators indicated whether a given image was of sufficient quality for grading. Second, the quality of the image was deemed insufficient when it became difficult or impossible to make a confident assessment regarding the characteristics of DR. Then, the image was categorised as normal (absence of DR) and abnormal (presence of DR). Third, the severity of DR in the abnormal images was annotated.

The grading reliability of each image was measured by cross-validation (checking others' grading results per image for every grader). First, almost all patients from the ophthalmology department were found with DR; hence, we used two types of images per eye: retinal colour fundus photographs and fluorescein angiography fundus photographs (pharmacological pupil dilation). The fundus images labelled with inconsistency were corrected using the original diagnosis reports with corresponding photographs of fluorescein angiography because accuracy ultimately came from fluorescent photography. Meanwhile, only retinal colour fundus photos were taken from the patients from the endocrine and physical examination centres. Disagreements were re-examined and resolved via discussion. If no consensus was reached, arbitration was performed by the retinal expert to generate final grading.

3.4. Pre-processing of retinal images

Given the complexity of the retina structure, the characteristics of DR were easily confused with those of other eye diseases. Moreover, we observed a range of imaging noise, such as black space on either side of the eye, low contrast, blurred lens or insufficient light. As a result, some minute lesions could not be accurately identified within the poor photographs. Therefore, some pre-processing was necessary. Fig. 2(b) shows some examples of poor-quality data.

First, an algorithm was devised to remove the invalid areas of black space by cropping a fixed number of pixels from each of the four sides of each image whilst avoiding a large amount of computational overhead caused by the black space. Second, the images exhibited resolutions in the range of 1631×1879 to 1823×1650 ; we standardised the resolution by downsizing all images to a uniform size in accordance with the input requirements of specific models. Furthermore, for measuring the light intensity at each pixel in a single image, we converted all images to greyscale. For the images with excessively bright or dark background and foreground, we used histogram equalisation (HE) to improve the visual effects of the images and to discover hidden messages. To improve local contrast and enhance edge definition in each image region, we used adaptive HE, which was initially proposed by Stark [38,39], as a part of the pre-processing steps. Mathematically, the adaptive HE can be written as follows:

$$f_c(u, v) = q(u - v, \alpha) - \beta q(u - v, 1) + \beta u q(d, \alpha) \quad (1)$$

$$q(d, \alpha) = \frac{1}{2} \sin(d) |2d|^\alpha,$$

where $0 \leq \alpha \leq 1$ and $\beta \leq 1$, and we set $\alpha = \beta$. f_c is called an accumulation function, and HE is given if $\alpha = 0$. These equations are explained in detail in [39].

For dark images, we provided a contrast stretching algorithm to enhance the contrast effect of each area of the image. This algorithm was performed by using the following equation:

$$I(x, y) = \frac{I(x, y) - I_{min}}{I_{max} - I_{min}} \times 255, \quad (2)$$

where $I(x, y)$ is the grey value of a pixel in the original image and I_{min} and I_{max} are the real minimum and maximum greyscale values of the original image, respectively.

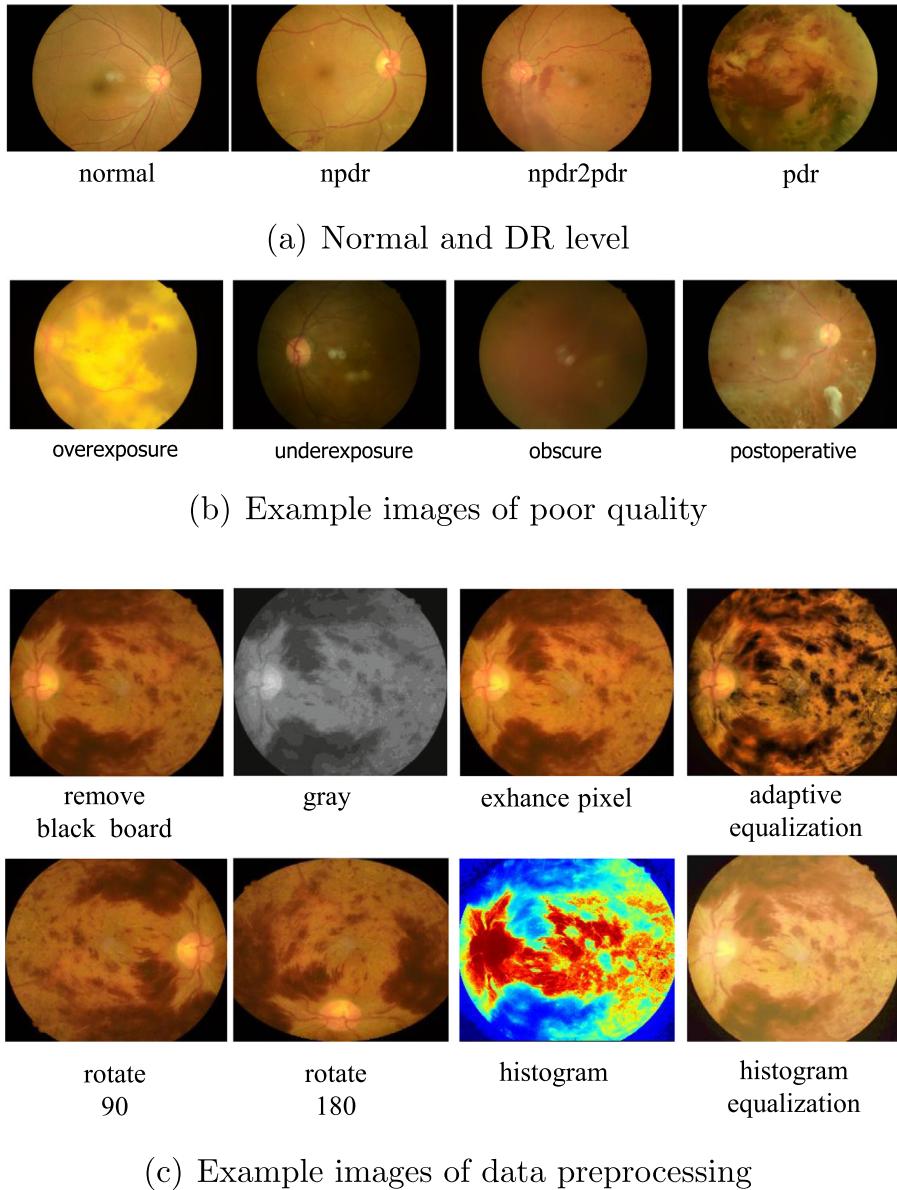


Fig. 2. Example images of poor quality and preprocessing in the dataset.

3.5. Performance comparison

Fig. 2(c) shows some examples of the pre-processing in our dataset. The comparative experiment was executed via Xception. The model performed well in the binary task. The model was trained via resizing only and no other pre-processing. Then, it was compared with the models trained with all the pre-processing methods. The accuracy of the model without pre-processing reached 94.79% until 300 epochs. This model's accuracy did not exceed such level even after fine tuning. By contrast, the model with pre-processing converged well after 220 epochs and achieved an accuracy of 95.68%. This accuracy rate improved to 97.15% after fine tuning (**Fig. 5**). **Fig. 3** and **Table 7** provide further details.

3.6. Data augmentation

The amount of PDR only accounted for 9% of the total data (**Table 3**), and the inter-grader variability was serious in the pathological features. The model had difficulties in learning the

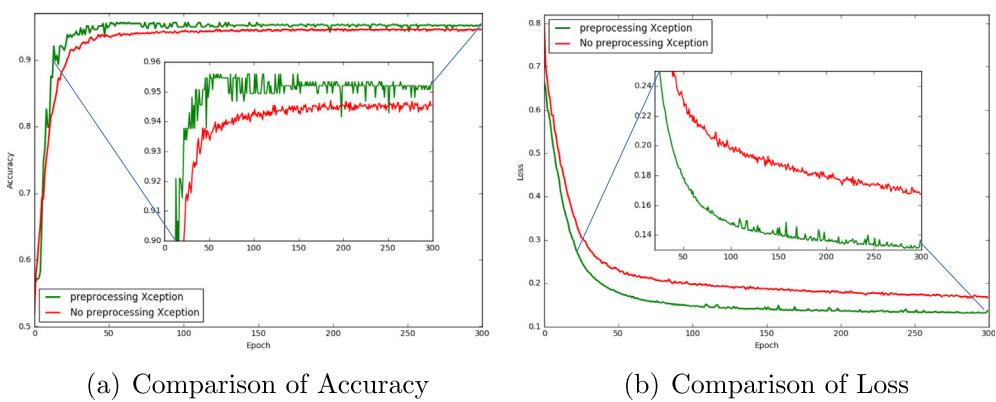
characteristics of PDR, and PDR was over identified as NPDR2PDR. Therefore, we used data augmentation technology, such as random rotation by 0° – 180° , random horizontal or vertical flips, to enhance the size of the training set in real time. Generally, augmented images retained the major features from their original images. Therefore, the technologies ensured that the training set was expanded whilst the images were not copied completely.

4. Model and methodology

4.1. Aim and objective

The following two aims motivated this study and were realised with corresponding ensemble models that were designed in this work.

(1) To build an early DR automatic screening system. This aim is a binary classification task to identify the presence of DR. Currently, the manual DR screening method is labour intensive and suffers from inconsistencies across sites [40]; moreover, the number of people who can master DR treatment skills is still

**Fig. 3.** The comparison before and after preprocessing of the identification system.**Table 4**
Hyper-parameter configuration.

Configuration	Value
Optimisation function	RMSprop
Epoch	300 (training), 100 (fine tuning)
Batch size	32
Learning rate	2.00E–04
Drop out	5.00E–01
Class_weight	auto
ReduceLROnPlateau	monitor = 'val_acc', factor = 0.5, patience = 10, epsilon = 0.0001
EarlyStopping	monitor = 'val_loss', patience = 15
ModelCheckpoint	monitor = 'val_acc', mode = 'auto', period = 1

Table 5
Class distribution and classifications report of the identification system.

	Training	Validation	Testing				
				Precision	Recall	F1_score	Support
Normal	1189	132	348				
Abnormal	1567	174	423				
Total	2756	306	771				
Normal	0.97	0.98	0.97	348			
Abnormal	0.98	0.97	0.98	423			
ave/total	0.98	0.98	0.98	771			

small amongst most grassroots health workers. Therefore, the first task was primarily applied to communities at the grassroots level because patients with DR may be identified timely and referred to ophthalmology for further diagnosis and treatment whilst potentially alleviating the burden for ophthalmologists.

(2) To build an automatic grading system. This aim is a multi-classification task to predict the level of DR severity. The grading system was primarily used in hospitals to provide ophthalmologists with auxiliary diagnostic references whilst avoiding human subjectivity.

4.2. Architecture and strategy of ensemble model

In this study, 'learner', 'basic learner', 'component classifier' and 'component' refer to an independent neural network used in an ensemble.

Several components comprised a corresponding ensemble model. Each component was a two-part neural network. The first part was a feature extractor comprising a pre-training model that was initialised via transfer learning which involved the removal of the top layer. The second part was a classifier that made predictions on the basis of the aforementioned features; it was realised by a customised stand deep neural network with training from scratch.

4.3. Ensemble strategy

Many studies have largely realised ensemble methods with many components at the expense of time and memory. Therefore, the number of component classifiers to be included in the ensemble is an important issue in integration model experiments [41]. The issue further triggers deep thinking, namely, the choice of combination method and the training mechanism of components, which largely characterise the ensemble method [42–44]. The former indicates which learners can be combined, whereas the latter describes how to combine them.

For 'which learners', extensive literature has shown that the performance of the overall classifier can be enhanced if a not-so-weak learner is used as a base learner [45]. Meanwhile, the predictive power of a learner is closely related to its ability to extract high-quality features; thus, a high-performance pre-model must be selected as the feature extractor in the bottom part of the component.

For 'how to', because DNNS has high variance and low bias, the variance of models can be dramatically reduced via averaging if the models are independent [46,47]. In the current work, we averaged the softmax scores of all models to solve the second issue above because these probabilities from different models might have varying output magnitudes. Comparative experiments between the averaging and max methods applied to the ensemble models were added for objectivity. The results demonstrated that the averaging method was more effective in our study (Tables 8 and 9).

4.4. Transfer learning at the first part

Several standout CNNs that remove top layers made up the first stage of different ensembles. We used these CNNs to produce a compact feature vector representation in a DR image. In consideration of different tasks, we performed some analyses of respective CNN characteristics as follows to facilitate our selection of a strong feature extractor (learner) for each component of the ensemble.

To enhance the speed of the calculation, we used ResNet50 [48] and InceptionV3 [49] as alternatives. InceptionResNetV2 [50] was also attempted because it can make the network deep and fast by mitigating the problem of gradient disappearance using jumper connections. Similarly, Xception [51] was considered as one improvement to InceptionV3 due to its ability to improve the effects of the model without increasing the complexity of the network. DenseNets [52] were attractive because they can fully use the features and further reduce gradient disappearance; furthermore, they performed well in our experiments.

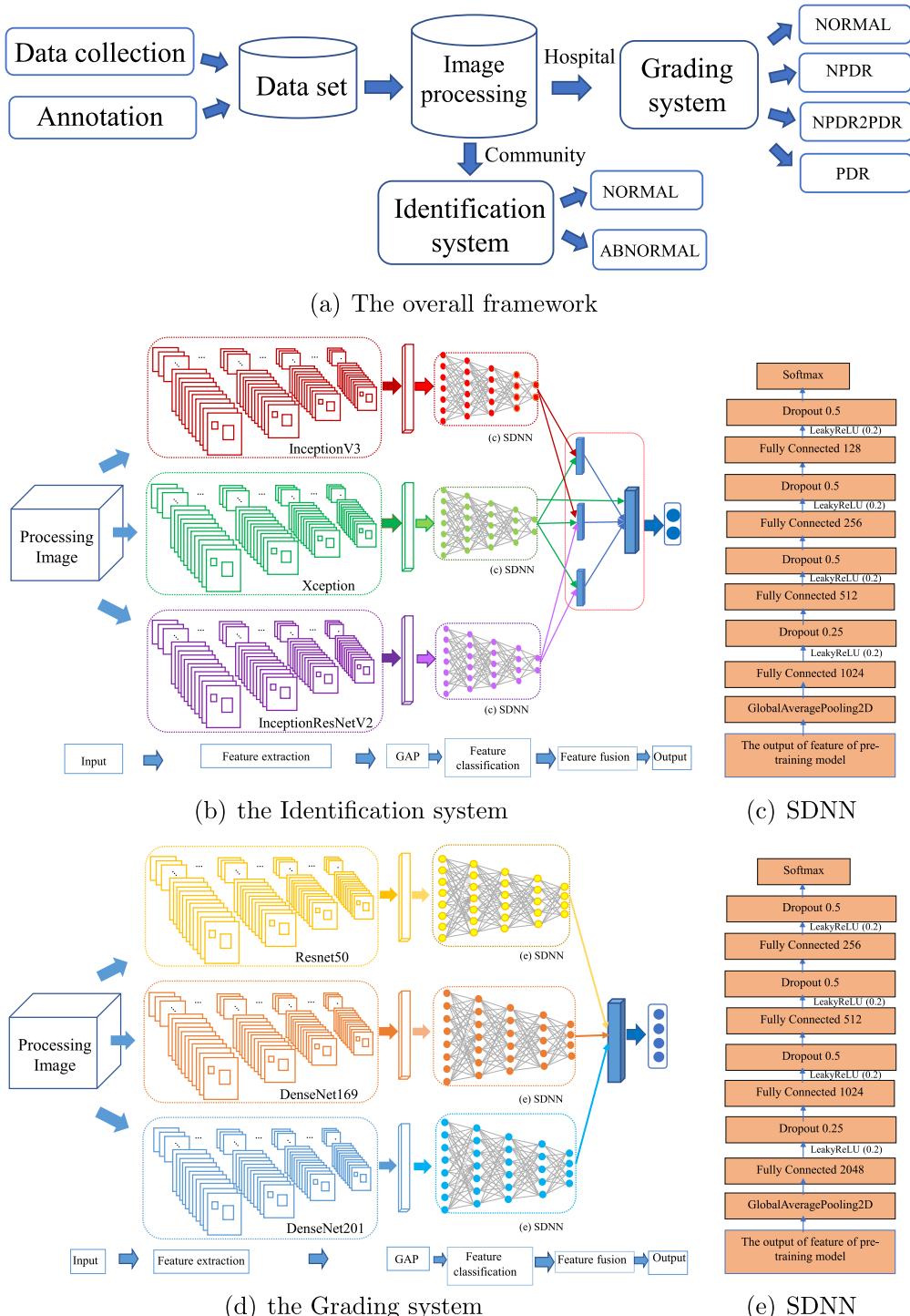


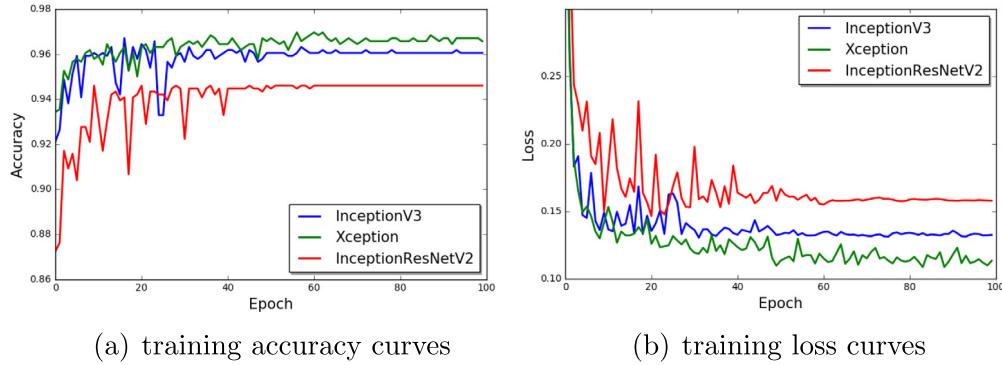
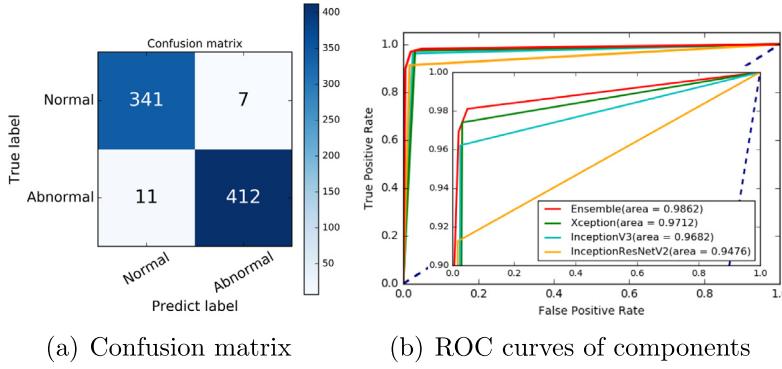
Fig. 4. The whole system.

4.5. Design of customised standard deep neural network (SDNN) at the second part

On the basis of the specific data distributions and difficulties of the two tasks, two respective types of SDNNs were defined as component classifiers at the second stage (Fig. 4). Input to the SDNNs was given by the output of the feature extractors. SDNNs have the same network depth but different parameters. This faint distinction led to important changes in the prediction performance: through experiments and observations, we found

that adding or removing a layer could reduce the learning capacity of the network regardless of the task. Given the similarity of designs of SDNN frameworks, we showed only the development of the SDNN model in the four-class classification task.

Notably, the feature maps from the forward layer are spatial, but they can be normalised via global average pooling (GAP). Thus, we designed a GAP layer as the first layer of the SDNN. The second layer was a fully connected layer with 2048 hidden neurons. Proper nonlinearity is an important factor in the incremental performance of a classification model, especially for a

**Fig. 5.** Training curves of the components of the identification system.**Fig. 6.** Confusion matrix and ROC curves of the identification system.

limited dataset. Thus, we closely monitored the effects of non-linearity between the rectified linear unit function (ReLU) and its variant, leaky rectified linear unit function (leaky ReLU), on SDNN performance. In the same condition, we conducted multiple comparison experiments and found that leaky ReLU was significantly faster than ReLU with respect to convergence and shortening of training time. According to prior literature [43–45], the difference may be caused by the potential disadvantage of ReLU during optimisation: the gradient is 0 whenever the neuron is dead (saturated and not active). This occurrence may cause the unit to remain inactive because the gradient-based optimisation algorithm does not adjust the weight of a dead neuron. Therefore, the speed of training ReLU networks is slow when gradients remain at zero. By contrast, leaky ReLU slightly adjusts the weight of dead neurons into small and non-zero gradients. On the basis of the above analysis, we applied leaky ReLU layers to the output of all inner fully connected layers, except the output layer, to achieve nonlinearity in the SDNN. The two functions of ReLU and leaky ReLU are defined in Eqs. (3) and (4), respectively.

$$h^{(i)} = \max(w^{(i)(T)}x, 0) = \begin{cases} w^{(i)(T)}x, & \text{if } w^{(i)(T)}x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

$$h^{(i)} = \max(w^{(i)(T)}x, 0) = \begin{cases} w^{(i)(T)}x, & \text{if } w^{(i)(T)}x > 0 \\ 0.01w^{(i)(T)}x, & \text{otherwise} \end{cases} \quad (4)$$

The formulas above show the similarities and differences between the two functions. The latter function may achieve a more robust gradient than the former by sacrificing sparsity.

A dropout layer was added after each dense layer. The addition can effectively omit many neurons of hidden layers during training and ensure the validity of the data; it can also mitigate or prevent data overfitting if the network shows excessive reliance on certain nodes in one layer. We updated each node with probability $p = 0.5$ whilst updating each layer. Then, we

left it unchanged with probability $1-p$. Following these multiple layers, the final layer was a standard softmax classifier with cross entropy as the loss function. The softmax function took an N -dimensional vector of arbitrary real values and produced another N -dimensional vector with real values in the range (0,1), thereby adding up to 1.0; each value of the output vector represented the probability that the sample belonged to each class. Cross entropy served as a loss function that revealed the distance or degree of closeness between the true labels and the predicted labels of the network. It is defined as follows:

$$L_j = -\log\left(\frac{e^{s_j}}{\sum_{k=1}^N e^{s_k}}\right), \quad (5)$$

where N indicates the number of classes, s_j is the score for the sample label j , and s_k is the score for a particular label k . Softmax ensured that the prediction probabilities exhibited a proper probability distribution.

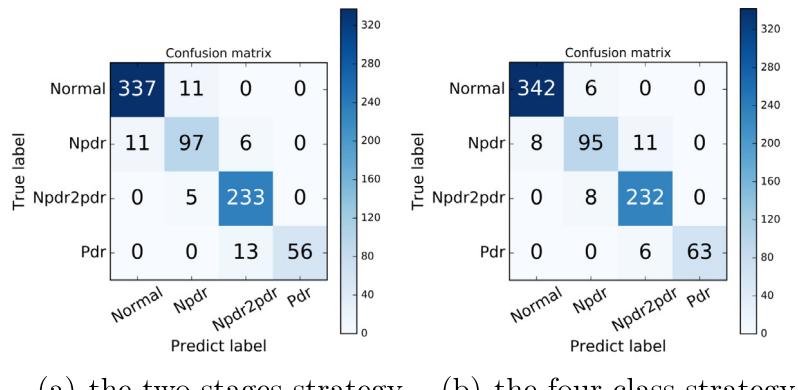
5. Experiments

5.1. Configuration

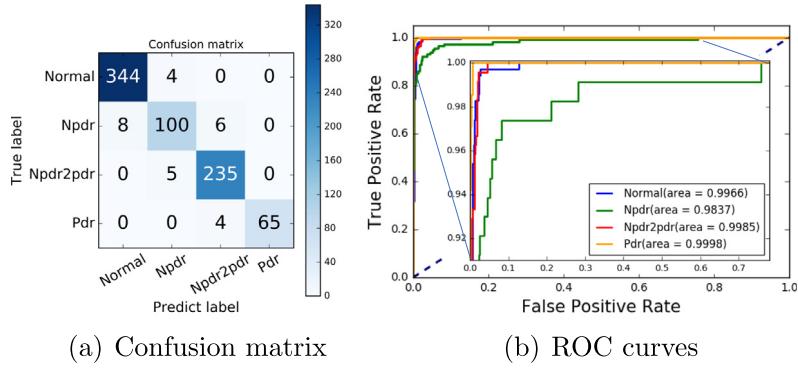
The algorithms were implemented using Keras (<http://keras.io/>). All experiments were performed on a high-end workstation with an Intel Xeon E5-2620 CPU and NVIDIA Tesla K40 GPU with 64 GB of RAM. The dataset was split into 70% for training, 10% for validation and 20% for testing. The configuration of the hyper-parameter and the class distributions of the two tasks are shown in Tables 4–6.

5.2. Strategy

The experiment process consisted of six steps: input data, data pre-processing, single-model feature extraction, single-model



(a) the two-stages strategy (b) the four-class strategy

Fig. 7. Comparison of confusion matrices between two strategies of the grading system.**Fig. 8.** Confusion matrix and ROC curves of the grading system.**Table 6**
Class distribution and classifications report of the grading system.

	Training	Validation	Testing	
Normal	1189	132	348	
npdr	469	52	114	
npdr2pdr	875	97	240	
pdr	224	25	69	
Total	2756	306	771	
	Precision	Recall	F1_score	Support
Normal	0.98	0.99	0.98	348
npdr	0.92	0.88	0.9	114
npdr2pdr	0.96	0.98	0.97	240
pdr	1	0.94	0.97	69
ave/total	0.96	0.96	0.96	771

feature classification, multi-model feature fusion and output results (Fig. 4). During one component development, the first phase focused on the separate pre-training of the SDNNs; the key points were the number of layers needed in each SDNN with corresponding parameters to achieve optimal performance. The second phase was to combine each pre-processing model with the corresponding SDNN model with retraining. Lastly, the component was fine-tuned by determining the suitable layer to be frozen. For convenience, each component classifier was named as the components' internal pre-training model.

Models were designed in parallel and independently whilst taking them as alternative component classifiers. The choice of the most suitable component classifier and the size of the ensemble have a great impact on the accuracy of the prediction results of the model. Therefore, we attempted rich experiments to explore the relationship amongst the ideal size of the ensemble model (the number of basic component classifiers), the optimal

width of the ensemble model (the combination method of component classifiers) and the number of class labels for the two systems on the basis of our dataset.

5.3. Metrics

We utilised reliability and validity to evaluate the models and their ensembles. Reliability measures the degree of stability obtained by repeated tests under the same conditions. It can be assessed using the Kappa value. Kappa > 0.8 indicates excellent consistency. Validity reflects the degree of conformity between measured and actual values, as indicated by the seven metrics below.

Accuracy indicates the proportion of samples classified correctly. Precision is the proportion of positives correctly predicted. For early screening systems, sensitivity and specificity are important reference indicators of the referral decisions of screening options that directly indicate the effectiveness of the system. Sensitivity measures the proportion of positives correctly predicted, whereas specificity is the proportion of true negatives correctly identified as such. Given a typical balance between the two measures, a receiver operating characteristic curve with AUC can represent this balance graphically. The F1_score is the harmonic average of precision and recall. An F1_score close to 1 indicates good performance. F_β is the weighted harmonic average of F1_score. As the Youden's index approaches 1, the screening system exhibits enhanced authenticity; the reverse relationship is equally valid. These measures are presented in the following:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (6)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

Table 7

Metrics of the identification system ensembles.

Model	Accuracy	Precision	Sensitivity	Specificity	Auc	Kappa	F1_score	F_β _score	Youden's index
Xception ^a	0.9521	0.9491	0.9321	0.9782	0.9551	0.9032	0.9523	0.9503	0.9101
Baseline:Resnet 50	0.9342	0.9333	0.9332	0.9351	0.9342	0.9231	0.9322	0.9223	0.9223
0:Xception ^b	0.9715	0.9711	0.9740	0.9683	0.9712	0.9424	0.9715	0.9712	0.9424
1:InceptionV3	0.9676	0.9667	0.9622	0.9741	0.9682	0.9676	0.9669	0.9346	0.9363
2:IncepresV2 ^c	0.9442	0.9435	0.9125	0.9828	0.9476	0.9444	0.9433	0.8883	0.8953
Ens(01) ^d	0.9767	0.9759	0.9716	0.9828	0.9772	0.9529	0.9767	0.9761	0.9544
Ens(02) ^e	0.9689	0.9677	0.9574	0.9828	0.9701	0.9373	0.9689	0.9680	0.9402
Ens(12) ^f	0.9598	0.9585	0.9456	0.9770	0.9613	0.9191	0.9599	0.9588	0.9226
3:Ens(012) ^g	0.9702	0.9691	0.9622	0.9799	0.9710	0.9340	0.9702	0.9694	0.9421
Ens(012,01) ^h	0.9689	0.9678	0.9598	0.9799	0.9698	0.9373	0.9689	0.9681	0.9397
Ens(012,02) ⁱ	0.9662	0.9652	0.9574	0.9770	0.9672	0.9321	0.9663	0.9655	0.9345
Ens(0,01,02) ^j	0.9753	0.9750	0.9764	0.9742	0.9752	0.9503	0.9753	0.9750	0.9505
Ens(0,2,01,02) ^k	0.9702	0.9691	0.9621	0.9799	0.9710	0.9399	0.9710	0.9694	0.9420
Ens(0,1,01,02) ^l	0.9740	0.9732	0.9693	0.9799	0.9746	0.9477	0.9741	0.9735	0.9492
Ens(0,01,02,3) ^m	0.9767	0.9760	0.9740	0.9799	0.9769	0.9529	0.9767	0.9762	0.9539
Ens(0,2,01,02,3) ⁿ	0.9701	0.9691	0.9622	0.9799	0.9710	0.9399	0.9702	0.9694	0.9421
Ens(0,1,01,02,3) ^o	0.9741	0.9732	0.9693	0.9799	0.9745	0.9477	0.9741	0.9735	0.9492
Ens (0,01,02,012) ^p	0.9767	0.9760	0.9764	0.9800	0.9862	0.9530	0.9769	0.9762	0.9540

^aXception = Comparison of performances before preprocessing of Xception.^bXception = Comparison of performances after preprocessing of Xception.^cIncepresV2 = InceptionResNetV2.^dEns(01) = Average of Xception and InceptionV3.^eEns(02) = Average of Xception and InceptionResNetV2.^fEns(12) = Average of InceptionV3 and InceptionResNetV2.^gEns(012) = Average of Xception, InceptionV3 and InceptionResNetV2.^hEns(012,01) = Average of Ens(012) and Ens(01).ⁱEns(012,02) = Average of Ens(012) and Ens(02).^jEns(0,02,02) = Average of Xception, Ens(02) and Ens(02).^kEns(0,2,02,12) = Average of Xception, InceptionResNetV2, Ens(02) and Ens(12).^lEns(0,1,01,02) = Average of Xception, InceptionV3, Ens(01) and Ens(02).^mEns(2,02,12,3) = Average of Xception, Ens(01), Ens(02) and Ens(012).ⁿEns(0,2,01,02,3) = Average of Xception, InceptionResNetV2, Ens(01), Ens(02), and Ens(012).^oEns(0,1,01,02,3) = Average of Xception, InceptionV3, Ens(01), Ens(02) and Ens(012).^pEns(0,01,02,012) = Average of Xception, Ens(01), Ens(02) and Ens(012).

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (8)$$

$$\text{F1_score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (9)$$

$$F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{recall}}{\beta^2 \times \text{Precision} + \text{recall}} \quad (10)$$

$$\begin{aligned} \text{Youden's index} &= \text{Sensitivity} + \text{Specificity} - 1 \\ &= \frac{\text{TP} \times \text{TN} - \text{FN} \times \text{FP}}{(\text{TP} + \text{FN}) \times (\text{TN} + \text{FP})}. \end{aligned} \quad (11)$$

Where, the samples can be divided into true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) in accordance with the combination of the real class and the classifier prediction category.

5.4. Identification system

The situation of classification imbalance was not serious. To reduce the computational cost, we selected three models (InceptionV3, InceptionResNetV2 and Xception) as alternative feature extractors, with ResNet50 as the baseline. These models largely represent the best learning abilities available. Whilst making multiple combinations of these models, we expected the identification system to achieve the best performance.

After training these models independently (Fig. 5), we marked the models from strong to weak: Xception, InceptionV3 and InceptionResNetV2 as 0, 1 and 2, respectively. First, the models were combined directly to an ensemble named Ens(012) by averaging their softmax scores. However, the performance did not

match that of Xception. We speculated that the ideal number of component classifiers for binary classification problems may not simply be a result of the direct integration of the three basic models above. Thus, ensemble models consisting of two basic models were generated. The experiments showed that Ens(01) performed the best amongst all ensembles. We widen the ensemble models by combining these basic models with Ens(01) and Ens(02). Notably, Ens (0,01,02) exceeded well the previous best model Ens(01), and the result indicated that the optimal width of the ensemble may sometimes be useful. Therefore, the ensembles were further widened by combining the basic models with other ensemble models, such as Ens (0,2,02,01). The new results were almost better than those of Ens(012), but they were not as good as those of Ens(01). We replaced these basic models with the ones that performed well in the above experiments. Ens (0,01,02,012) showed a standout performance; hence, it was taken as the ensemble model of the task, for which it yielded the highest performance (Fig. 4(b)). Fig. 6 and Table 5 show the performance of the final ensemble. Table 7 presents the primary ensembles attempted and their performances under as many combinations as possible.

5.5. Grading system

5.5.1. Two alternative strategies

Regarding the grading system, we considered two alternative implementation strategies. The first strategy was a two-stage system: the first stage was a binary classification that distinguished between abnormal images and normal images via Xception, and the second stage was a ternary classification process to predict the level of DR severity on the basis of the abnormal data above

Table 8

Metrics of the grading system ensembles: the first group, the input is (299×299) .

Model	Accuracy	Recall	Precision	Sensitivity	Specificity	Kappa	F1_score	F_{β} _score	Youden's index
0: InceptionV3	0.9455	0.9260	0.9376	0.9787	0.9655	0.8811	0.9449	0.9349	0.9442
1: IncepresV2	0.9416	0.9072	0.9072	0.9764	0.9741	0.9123	0.9410	0.9306	0.9505
2: Xception	0.9209	0.8809	0.9001	0.9622	0.9828	0.8811	0.9201	0.9959	0.9449
Ens(01) ^a	0.9442	0.9101	0.9408	0.9788	0.9770	0.9436	0.9436	0.9336	0.9460
Ens(01) _m	0.9442	0.9101	0.9408	0.9788	0.9770	0.9162	0.9436	0.9336	0.9557
Ens(02) ^b	0.9507	0.9332	0.9431	0.9811	0.9684	0.9262	0.9503	0.9409	0.9495
Ens(02) _m	0.9468	0.9267	0.9396	0.9787	0.9684	0.9203	0.9461	0.9367	0.9471
Ens(12) ^c	0.9429	0.9094	0.9387	0.9787	0.9741	0.9143	0.9424	0.9318	0.9529
Ens(12) _m	0.9416	0.9072	0.9377	0.9764	0.9741	0.9123	0.9411	0.9306	0.9505
3: Ens(012) ^d	0.9520	0.9701	0.9677	0.9811	0.9770	0.9280	0.9517	0.9423	0.9581
Ens(012) _m	0.9442	0.9101	0.9408	0.9787	0.9770	0.9162	0.9436	0.9336	0.9557
Ens(3,0) ^e	0.9455	0.9262	0.9377	0.9787	0.9626	0.9184	0.9449	0.9350	0.9414
Ens(3,1) ^f	0.9416	0.9072	0.9376	0.9764	0.9741	0.9123	0.9410	0.9306	0.9505
Ens(3,2) ^g	0.9494	0.9217	0.9406	0.9764	0.9770	0.9241	0.9489	0.9365	0.9534
Ens(3, 2) _m	0.9416	0.9072	0.9377	0.9764	0.9741	0.9123	0.9410	0.9306	0.9505
Ens(3,01) ^h	0.9546	0.9282	0.9529	0.9811	0.9799	0.9318	0.9542	0.9473	0.9610
Ens(3, 01) _m	0.9533	0.9260	0.9518	0.9810	0.9799	0.9299	0.9528	0.9459	0.9610
Ens(3,02) ⁱ	0.9494	0.9281	0.9432	0.9787	0.9713	0.9242	0.9488	0.9398	0.9500
Ens(3, 02) _m	0.9494	0.9296	0.9423	0.9787	0.9684	0.9242	0.9490	0.9396	0.9471
Ens(01,02) ^j	0.9494	0.9310	0.9421	0.9787	0.9684	0.9242	0.9489	0.9396	0.9471
Ens(01, 02) _m	0.9520	0.9324	0.9463	0.9787	0.9741	0.9281	0.9513	0.9432	0.9529
Ens(3, Ens(01, 02) _m) ^k	0.9533	0.9306	0.9489	0.9787	0.9741	0.9300	0.9528	0.9449	0.9529
Ens(3, Ens(01, 02) _m) _m	0.9520	0.9310	0.9465	0.9787	0.9741	0.9280	0.9515	0.9431	0.9529

^aEns(01) = Average of InceptionV3 and IncepresV2; Ens(01)_m = max of them.

^bEns(02) = Average of InceptionV3 and Xception; Ens(02)_m = max of them.

^cEns(12) = Average of InceptionV3, IncepresV2 and Xception; Ens(12)_m = max of them.

^dEns(012) = Average of IncepresV2 and Xception; Ens(012)_m = max of them.

^eEns(3,0) = Average (max) of Ens(012) and InceptionV3.

^fEns(3,1) = Average (max) or max of Ens(012) and IncepresV2.

^gEns(3,2) = Average of Ens(012) and Xception; Ens(3, 2)_m = max of them.

^hEns(3,01) = Average of Ens(012) and Ens(01); Ens(3, 01)_m = max of them.

ⁱEns(3,02) = Average of Ens(012) and Ens(02); Ens(3, 02)_m = max of them.

^jEns(01,02) = Average of Ens(01) and Ens(02); Ens(01, 02)_m = max of them.

^kEns(3, Ens(01, 02)_m) = Average of Ens(012), Ens(01, 02)_m; Ens(3, Ens(01, 02)_m)_m = max of them.

via Resnet50. The alternative strategy was a quaternary classification model for predicting the level of DR severity on the basis of all testing images via Resnet50. The accuracy of the former was 94.1%, and the accuracy of the latter was 95.2%. The intuitive comparison (Fig. 7) showed that the strategy of the four-class classification worked the best.

5.5.2. Four-class classification

The seven models discussed above were used in the experiments because of the thin granularity of multiple classification and the small amount of PDR data. The four-class experiments were divided into two groups by the models on the basis of their different input sizes (299×299 , 224×224). From strong to weak, three models, i.e. InceptionV3, InceptionResNetV2 and Xception, were marked as 0, 1 and 2, respectively, in group one; the others, i.e. ResNet50, DenseNet169, DenseNet201 and DenseNet121, were marked as 4, 5, 6, 7, respectively. First, we evaluated various ensembles in each group. Second, the outstanding models from the two groups were combined as much as possible to determine the final ensemble framework of the grading system.

Galton's theory states that combining many simple predictions is a force for accurate predictions. Hence, we combined all basic models directly. This step yielded Ens (0124567) with a remarkable accuracy rate of 96.36% relative to the other base learners above. Notably, the sensitivity and specificity of Ens (0124567) were 98.10% and 98.56%, respectively; however, it costs more than the pre-trained models do. Inspired by the binary experiments above, we reduced the size of the ensembles to four. Moreover, the optimal number of component classifiers should be similar to the number of class labels in some studies [42–44]. The results showed that Ens (4567) had a high accuracy of 96.23% under the condition involving the same number of base

learners. Sequentially, we reduced the size of the ensembles to three. The top three single models (0,4,5) were integrated into Ens (045); however, its accuracy was only 94.94%. Subsequently, we replaced the third-ranked model 2: InceptionV3 with the fourth-ranked model 5: DenseNet201; this step yielded Ens (456) with an accuracy of 96.50% and other relatively high indicators. A variety of integrations were further attempted, and similar performances were achieved; however, the results did not exceed the current level. Therefore, our judgment was correct. In other words, an optimal number of component classifiers exists and might be near the number of class tags with subtle adjustments depending on the specific task. Fig. 4(d) shows the Ens (456) framework, and Fig. 8 shows its performance. Further information about the evaluation metrics is shown in Tables 8–10.

5.6. Analysis of experiments

The UK National Institute for Clinical Excellence guidelines state that a DR screening test should have at least a sensitivity of 80% and a specificity of 95%. In our work, the identification model performed well with a sensitivity of 97.5%, a specificity of 97.7% and an accuracy of 97.7%. By contrast, the grading model achieved a sensitivity of 98.1%, a specificity of 98.9% and an accuracy of 96.5%. Therefore, the models achieved satisfactory performance on our dataset.

From the experiments, we found that the stronger the base learner was, the higher the performance was generally; moreover, the effects of the ensembles with multiple-ensemble classifiers were stronger than those of the dull ensemble models in some cases. We noticed that some results degenerated after the ensemble phase in the four-class experiments. We conjectured this degeneration was primarily caused by the model selection strategy in the ensemble phase: various base learners of different

Table 9

Metrics of the grading system ensembles: the second group, the input is (224 × 224).

Model	Accuracy	Recall	Precision	Sensitivity	Specificity	Kappa	F1_score	F_{β} _score	Youden's index
4: Resnet50	0.9494	0.9240	0.9451	0.9811	0.9828	0.9241	0.9491	0.9404	0.9638
5: DN_169	0.9469	0.9195	0.9466	0.9551	0.9914	0.9199	0.9464	0.9406	0.9465
6: DN_201	0.9429	0.9140	0.9333	0.9811	0.9741	0.9144	0.9421	0.9289	0.9552
7: DN_121	0.9364	0.9162	0.9222	0.9740	0.9741	0.9052	0.9372	0.9203	0.9481
Ens(45) ^a	0.9507	0.9261	0.9463	0.9811	0.9828	0.9261	0.9504	0.9418	0.9638
Ens(46) ^b	0.9546	0.9295	0.9534	0.9811	0.9828	0.9318	0.9540	0.9480	0.9638
Ens(47) ^c	0.9507	0.9273	0.9456	0.9811	0.9826	0.9261	0.9506	0.9415	0.9638
Ens(56) ^d	0.9494	0.9190	0.9450	0.9787	0.9856	0.9239	0.9485	0.9390	0.9644
Ens(57) ^e	0.9494	0.9241	0.9450	0.9645	0.9885	0.9240	0.9491	0.9403	0.9530
Ens(67) ^f	0.9481	0.9186	0.9428	0.9787	0.9798	0.9220	0.9472	0.9373	0.9586
8: Ens(456) ^g	0.9650	0.9467	0.9635	0.9811	0.9885	0.9475	0.9647	0.9599	0.9696
Ens(456) _m	0.9546	0.9295	0.9534	0.9811	0.9828	0.9318	0.9540	0.9480	0.9638
Ens(457) ^h	0.9571	0.9319	0.9561	0.9787	0.9885	0.9357	0.9567	0.9508	0.9508
Ens(467) ⁱ	0.9611	0.9402	0.9572	0.9834	0.9828	0.9417	0.9610	0.9535	0.9662
Ens(567) ^j	0.9611	0.9347	0.9589	0.9764	0.9914	0.9415	0.9605	0.9535	0.9508
9: Ens(4567) ^k	0.9623	0.9412	0.9616	0.9787	0.9856	0.9435	0.9620	0.9572	0.9644
Ens(4567) _m	0.9546	0.9295	0.9534	0.9811	0.9828	0.9318	0.9540	0.9480	0.9638
Ens(8,9,467) ^l	0.9611	0.9405	0.9595	0.9787	0.9828	0.9416	0.9609	0.9554	0.9615
Ens(8,9,567) ^m	0.9637	0.9434	0.9639	0.9763	0.9885	0.9454	0.9632	0.9594	0.9649
Ens(7,56,9,567) ⁿ	0.9650	0.9387	0.9635	0.9764	0.9913	0.9473	0.9645	0.9580	0.9678

^aEns(45) = Average of Resnet50 and DN_169.^bEns(46) = Average of Resnet50 and DN_201.^cEns(47) = Average of Resnet50 and DN_121.^dEns(56) = Average of DN_169 and DN_201.^eEns(57) = Average of DN_169 and DN_121.^fEns(67) = Average of DN_201 and DN_121.^gEns(456) = Average of Resnet50, DN_169 and DN_201; Ens(456)_m = max of them.^hEns(457) = Average of Resnet50, DN_169 and DN_121.ⁱEns(467) = Average of Resnet50, DN_201 and DN_121.^jEns(567) = Average of DN_169, DN_201 and DN_121.^kEns(4567) = Average of Resnet50, DN_169, DN_201 and DN_121; Ens(4567)_m = max of them.^lEns(456,8,467) = Average of Ens(456), Ens(4567), Ens(467).^mEns(456,8,567) = Average of Ens(456), Ens(4567), Ens(567).ⁿEns(7,56,8,567) = Average of DN_121, Ens(56), Ens(4567) and Ens(567).

depths can implicitly learn different levels of semantic image representation. On the basis of model complementarity, the posterior probabilities of these weak learners can be fused to predict the modalities of unseen images. Compared with the max method in Table 8, the averaging method was effective in our work because it reduced the variances of the components substantively [47,53].

6. Discussion

During the design of the two ensemble models, we made several considerations in the following aspects.

(1) Combined strategy of components: The frameworks of the two classification tasks searched for the ideal combination of component classifiers on the basis of the number of class tags in the dataset as a guide. We assumed that the basic components used in our experiments were all independent. In the experiments, we found that arbitrarily increasing or decreasing the number of component classifiers would reduce the performance of the model. Moreover, different combinations of methods of the component classifiers were important in achieving the best integration performance. However, the diversity of the real-world dataset, the complexity of the task and the degree of independence of existing component classifiers under the constraint of computational resource requirements did not guarantee this assumption in most cases; thus, determining the integration framework of a given set classifier on real data remains a challenging problem.

(2) Model optimisation: In view of the limited dataset and the deep models used in the system, we should note the problem of gradient disappearance. On the basis of the unsupervised layer-wise training [54], we proposed a supervised block-wise training strategy. Each SDNN of a component classifier was briefly

used as an independent model, and it was trained with high-level features as input via the corresponding pre-trained feature extractor model. After the SDNN was separately trained, it could be connected to the corresponding feature extractor to form a component classifier; the feature extractor module could be initialised with the pre-trained weight, and the optimal weight obtained after the SDNN independent training could be used as its pre-trained weight to initialise itself. Subsequently, the entire component classifier could be trained using fine-tuning after the training. This training arrangement increased the speed of the convergence of the whole classification component. In sum, the entire component was divided into several blocks for training; a good setting of weight parameters was found for each block, and the component was then globally optimised in accordance with the local optimal weight of each block whilst minimising the training cost.

(3) Several obvious advantages: First, the reproducibility and consistency of diagnostic results could provide clinicians with insights into the diagnostic process. Additionally, the two different systems could be used to match different application requirements. When screening populations with substantial diseases, achieving high sensitivity and high specificity is critical to minimise false-positive and false-negative results. Finally, a quick reporting of auxiliary diagnostic results could improve clinicians' efficiency.

(4) Some limitations: The annotation work was based on the clinical experience of the ophthalmologist graders. Therefore, the algorithm may perform differently when used in images with subtle findings that a majority of clinicians could not identify. Another fundamental limitation arises from the black box, which is the nature of deep networks. The network automatically learns the features from the images and associated grade; however, the

Table 10

Metrics of the grading system ensembles: integration.

Model	Accuracy	Recall	Precision	Sensitivity	Specificity	Kappa	F1_score	F_{β} _score	Youden's index
Ens(04) ^a	0.9559	0.9292	0.9716	0.9828	0.9885	0.9336	0.9552	0.9506	0.9544
Ens(045) ^b	0.9494	0.9242	0.9448	0.9645	0.9885	0.9240	0.9491	0.9403	0.9530
Ens(0456) ^c	0.9494	0.9242	0.9448	0.9645	0.9885	0.9240	0.9491	0.9403	0.9530
10:Ens(0124567) ^d	0.9637	0.9411	0.9647	0.9810	0.9856	0.9454	0.9632	0.9594	0.9667
Ens(01567) ^e	0.9624	0.9404	0.9634	0.9812	0.9856	0.9435	0.9618	0.9582	0.9667
Ens(01,56) ^f	0.9624	0.9406	0.9632	0.9811	0.9856	0.9435	0.9615	0.9577	0.9667
Ens(8,10) ^g	0.9624	0.9406	0.9632	0.9811	0.9856	0.9435	0.9615	0.9577	0.9667
Ens(Ens(3,01),8,10) ^h	0.9611	0.9353	0.9627	0.9811	0.9856	0.9414	0.9605	0.9564	0.9667
Ens(3,8,567) ⁱ	0.9611	0.9353	0.9627	0.9811	0.9856	0.9414	0.9605	0.9564	0.9667
Ens(Ens(3,01),9,567) ^j	0.9637	0.9412	0.9644	0.9834	0.9856	0.9454	0.9632	0.9564	0.9690
Ens(9,567)	0.9611	0.9428	0.9584	0.9787	0.9856	0.9416	0.9608	0.9550	0.9644

^aEns(04) = Average of InceptionV3 and Resnet50.^bEns(045) = Average of InceptionV3, Resnet50, DN_169.^cEns(0456) = Average of InceptionV3, Resnet50, DN_169, and DN_201.^dEns(0124567) = Average of all models.^eEns(01567) = Average of InceptionV3, InceptionV3Resnet50, DN_169, and DN_201.^fEns(01,56) = Average of Ens(01) and Ens(56).^gEns(8,10) = Average of **Ens(456)** and Ens(01567).^hEns(Ens(3,01),8,10) = Average of Ens(3,01), Ens(456) and Ens(01567).ⁱEns(3,8,567) = Average of Ens(012), Ens(4567) and Ens(567).^jEns(Ens(3,01),9,567) = Average of Ens(012), Ens(4567) and Ens(567).

specific features by which the networks are formed are unknown. Understanding the aspects used by deep neural networks to make predictions is an active area of research.

In the future, a large training dataset containing tens of thousands of abnormal cases must be collected from other hospitals via various types of cameras to improve the models' generalisation. Second, the visualisation of the aspects learned by CNNs is important as it can improve the interpretability of diagnostic results by identifying the source regions of features associated with a specified classification result, as well as the magnitude of the feature intensity. Additionally, doctors can make an accurate diagnosis on the basis of visualisation results. Third, in the case of a medical dataset of limited scale, further discussing the design and research of ensemble frameworks from a theoretical perspective is necessary.

7. Conclusion

In conclusion, a high-quality labelled medical imaging DR dataset was built, and an identification and grading system of DR called DeepDR was proposed. The relationship between the number of ideal component classifiers and the number of class labels was verified and explored. Using nine medical metrics, we evaluated the models in terms of validity and reliability. The results demonstrated that DeepDR worked satisfactorily.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 61432012 and U1435213.

References

- [1] R. Klein, B.E. Klein, S.E. Moss, M.D. Davis, D.L. DeMets, The wisconsin epidemiologic study of diabetic retinopathy: II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years, *Arch. Ophthalmol.* 102 (4) (1984) 520–526.
- [2] Y.H. Ma, Diabetic retinopathy screening rate of less than 10% in China, *Chin. J. Med. Sci.* 3 (2016).
- [3] J.C. Javitt, L.P. Aiello, Y. Chiang, F.F. Rd, J.K. Canner, S. Greenfield, Preventive eye care in people with diabetes is cost-saving to the federal government. implications for health-care reform, *Diabetes Care* 17 (8) (1994) 909–917.
- [4] D.E. Singer, Screening for diabetic retinopathy, *J. Intern. Med.* 240 (1) (1996) 45.
- [5] J.K. Kristinsson, E. Stefánsson, F. Jónasson, I. Gíslason, S. Björnsson, Systematic screening for diabetic eye disease in insulin dependent diabetes, *Acta Ophthalmol.* 72 (1) (2010) 72–78.
- [6] Diabetic Retinopathy, American Academy of Ophthalmology <https://www.aao.org/preferred-practice-pattern/diabetic-retinopathy-ppp-updated-2017/>, 2016.
- [7] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1798–1828.
- [8] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436.
- [9] V. Sze, Y.-H. Chen, T.-J. Yang, J.S. Emer, Efficient processing of deep neural networks: A tutorial and survey, *Proc. IEEE* 105 (12) (2017) 2295–2329.
- [10] Z. Yi, Foundations of implementing the competitive layer model by lotka-volterra recurrent neural networks, *IEEE Trans. Neural Netw.* 21 (3) (2010) 494–507, <http://dx.doi.org/10.1109/TNN.2009.2039758>.
- [11] L. Zhang, Z. Yi, S.-i. Amari, Theoretical study of oscillator neurons in recurrent neural networks, *IEEE Trans. Neural Netw. Learn. Syst.* (2018).
- [12] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 2012 (2012) 1097–1105.
- [13] R. Fakoor, F. Ladakh, A. Nazi, M. Huber, Using deep learning to enhance cancer diagnosis and classification, in: Proceedings of the International Conference on Machine Learning, vol. 28, 2013.
- [14] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (7639) (2017) 115.
- [15] D. Wang, A. Khosla, R. Gargya, H. Irshad, A.H. Beck, Deep learning for identifying metastatic breast cancer, 2016, arXiv preprint [arXiv:1606.05718](https://arxiv.org/abs/1606.05718).
- [16] A.M. Rossetto, W. Zhou, Deep learning for categorization of lung cancer CT images, in: 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologie, CHASE, 2017, pp. 272–273.
- [17] J. Wang, R. Ju, Y. Chen, L. Zhang, J. Hu, Y. Wu, W. Dong, J. Zhong, Z. Yi, Automated retinopathy of prematurity screening using deep neural networks, *Ebiomedicine* (2018).
- [18] J. Hu, Y. Chen, J. Zhong, R. Ju, Z. Yi, Automated analysis for retinopathy of prematurity by deep neural networks, *IEEE Trans. Med. Imaging* (2018).
- [19] S.J. Pan, Q. Yang, et al., A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* 22 (10) (2010) 1345–1359.
- [20] R.S. Michalski, A theory and methodology of inductive learning, in: *Machine Learning*, Springer, 1983, pp. 83–134.
- [21] L. Rokach, Ensemble-based classifiers, *Artif. Intell. Rev.* 33 (1–2) (2010) 1–39.
- [22] R. Polikar, Ensemble based systems in decision making, *IEEE Circuits Syst. Mag.* 6 (3) (2006) 21–45.
- [23] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, *J. Artificial Intelligence Res.* 11 (1999) 169–198.
- [24] R.J. Winder, P.J. Morrow, I.N. Mcritchie, J.R. Bailie, P.M. Hart, Algorithms for digital image processing in diabetic retinopathy, *Comput. Med. Imaging Graph.* 33 (8) (2009) 608–622.

- [25] M.D. Abràmoff, M.K. Garvin, M. Sonka, Retinal imaging and image analysis, *IEEE Rev. Biomed. Eng.* 3 (2010) 169–208.
- [26] T. Walter, J.-C. Klein, P. Massin, A. Erginay, A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina, *IEEE Trans. Med. Imaging* 21 (10) (2002) 1236–1243.
- [27] M. Niemeijer, B. van Ginneken, S.R. Russell, M.S. Suttorp-Schulten, M.D. Abramoff, Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis, *Invest. Ophthalmol. Vis. Sci.* 48 (5) (2007) 2260–2267.
- [28] O. Faust, R. Acharya, E.Y.-K. Ng, K.-H. Ng, J.S. Suri, Algorithms for the automated detection of diabetic retinopathy using digital fundus images: A review, *J. Med. Syst.* 36 (1) (2012) 145–157.
- [29] M.U. Akram, S. Khalid, A. Tariq, S.A. Khan, F. Azam, Detection and classification of retinal lesions for grading of diabetic retinopathy, *Comput. Biol. Med.* 45 (2014) 161–171.
- [30] L. Gowda, K. Viswanatha, Automatic diabetic retinopathy detection using FCM, 2018.
- [31] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.* 42 (2017) 60–88.
- [32] I. Kaggle, Diabetic Retinopathy Detection, American Academy of Ophthalmology, 2015, URL: <https://www.kaggle.com/c/diabetic-retinopathy-detection>.
- [33] J. Benson, H. Carrillo, J. Wigdahl, S. Nemeth, J. Maynard, G. Zamora, S. Barriga, T. Estrada, P. Soliz, Transfer learning for diabetic retinopathy, in: *Image Processing*, 2018, p. 70.
- [34] R. Gargya, T. Leng, Automated identification of diabetic retinopathy using deep learning, *Ophthalmology* 124 (7) (2017) 962–969.
- [35] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *JAMA* 316 (22) (2016) 2402–2410.
- [36] A. Tamkin, I. Usiri, C. Fufa, Deep CNNs for diabetic retinopathy detection.
- [37] H. Pratt, F. Coenen, D.M. Broadbent, S.P. Harding, Y. Zheng, Convolutional neural networks for diabetic retinopathy, *Procedia Comput. Sci.* 90 (2016) 200–205.
- [38] J.A. Stark, Adaptive image contrast enhancement using generalizations of histogram equalization, *IEEE Trans. Image Process.* 9 (5) (2002) 889–896.
- [39] J.A. Stark, W.J. Fitzgerald, Model-based adaptive histogram equalization, *Signal Process.* 39 (1–2) (1994) 193–200.
- [40] J.K.H. Goh, C.Y. Cheung, S.S. Sim, P.C. Tan, G.S.W. Tan, T.Y. Wong, Retinal imaging techniques for diabetic retinopathy screening, *J. Diabetes Sci. Technol.* 10 (2) (2016) 282–294.
- [41] G. Tsoumakas, I. Partalas, I. Vlahavas, A taxonomy and short review of ensemble selection, in: *Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications*, 2008, pp. 1–6.
- [42] D. Hernández-Lobato, G. Martínez-Muóz, A. Suárez, How large should ensembles of classifiers be? *Pattern Recognit.* 46 (5) (2013) 1323–1336.
- [43] H.R. Bonab, F. Can, A theoretical framework on the ideal number of classifiers for online ensembles in data streams, 2016, pp. 2053–2056.
- [44] H.R. Bonab, F. Can, Less is more: A comprehensive framework for the number of components of ensemble classifiers, *CoRR* (2017).
- [45] T.G. Dietterich, Ensemble methods in machine learning, in: *Multiple Classifier Systems*, Springer, Berlin, Heidelberg, 2000, pp. 1–15.
- [46] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [47] C. Ju, A. Bibaut, M. van der Laan, The relative performance of ensemble methods with deep convolutional neural networks for image classification, *J. Appl. Stat.* (2018) 1–19.
- [48] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (9) (2015) 1904–1916, <http://dx.doi.org/10.1109/TPAMI.2015.2389824>.
- [49] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: *The IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, 2016.
- [50] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, 2016.
- [51] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.
- [52] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [53] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [54] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, *Adv. Neural Inf. Process. Syst.* (2007) 153–160.

Journal Pre-proof

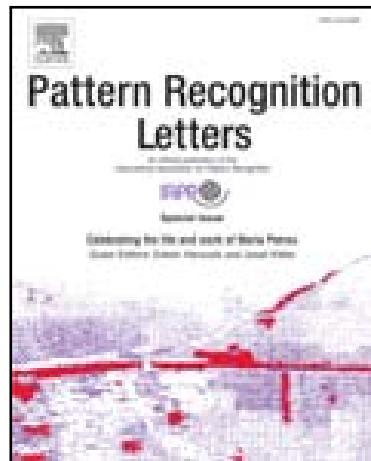
Deep Learning Frameworks for Diabetic Retinopathy Detection with Smartphone-based Retinal Imaging Systems

Recep E. Hacisoftaoglu , Mahmut Karakaya , Ahmed B. Sallam

PII: S0167-8655(20)30129-X

DOI: <https://doi.org/10.1016/j.patrec.2020.04.009>

Reference: PATREC 7856



To appear in: *Pattern Recognition Letters*

Received date: 1 October 2019

Revised date: 26 March 2020

Accepted date: 8 April 2020

Please cite this article as: Recep E. Hacisoftaoglu , Mahmut Karakaya , Ahmed B. Sallam , Deep Learning Frameworks for Diabetic Retinopathy Detection with Smartphone-based Retinal Imaging Systems, *Pattern Recognition Letters* (2020), doi: <https://doi.org/10.1016/j.patrec.2020.04.009>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier B.V.



Deep Learning Frameworks for Diabetic Retinopathy Detection with Smartphone-based Retinal Imaging Systems

^aRecep E. Hacisoftaoglu, ^aMahmut Karakaya*, ^bAhmed B. Sallam

^aDept. of Computer Science, University of Central Arkansas, Conway, AR, 72035, USA

^bJones Eye Institute, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA

ABSTRACT

Diabetic Retinopathy (DR) may result in various degrees of vision loss and even blindness if not diagnosed in a timely manner. Therefore, having an annual eye exam helps early detection to prevent vision loss in earlier stages, especially for diabetic patients. Recent technological advances made smartphone-based retinal imaging systems available on the market to perform small-sized, low-powered, and affordable DR screening in diverse environments. However, the accuracy of DR detection depends on the field of view and image quality. Since smartphone-based retinal imaging systems have much more compact designs than a traditional fundus camera, captured images are likely to be the low quality with a smaller field of view. Our motivation in this paper is to develop an automatic DR detection model for smartphone-based retinal images using the deep learning approach with the ResNet50 network. This study first utilized the well-known AlexNet, GoogLeNet, and ResNet50 architectures, using the transfer learning approach. Second, these frameworks were retrained with retina images from several datasets including EyePACS, Messidor, IDRiD, and Messidor-2 to investigate the effect of using images from the single, cross, and multiple datasets. Third, the proposed ResNet50 model is applied to smartphone-based synthetic images to explore the DR detection accuracy of smartphone-based retinal imaging systems. Based on the vision-threatening diabetic retinopathy detection results, the proposed approach achieved a high classification accuracy of 98.6%, with a 98.2% sensitivity and a 99.1% specificity while its AUC was 0.9978 on the independent test dataset. As the main contributions, DR detection accuracy was improved using the deep transfer learning approach for the ResNet50 network with publicly available datasets and the effect of the field of view in smartphone-based retinal imaging was studied. Although a smaller number of images were used in the training set compared with the existing studies, considerably acceptable high accuracies for validation and testing data were obtained.

2019 Elsevier Ltd. All rights reserved.

1. Introduction

Based on data from the World Health Organization, 422 million people have diabetes in 2014 around the world, and the number is predicted to be 552 million by 2030 [1]. The US Department of Health and Human Services National Diabetes Statistics Report [2] demonstrates that an estimation of 30.5 million in the US population (10.5 percent) has diabetes in 2020, with 7.3 million people undiagnosed, among all age groups. Individuals with diabetes are at high risk of diabetic eye diseases such as Diabetic Retinopathy (DR), Diabetic Macular Edema (DME), and Glaucoma. DR, the most suffered disease among all others, is caused by the damaging of blood vessels in the retina. The signs of DR can be listed as including but not limited to the existence of microaneurysms, vitreous hemorrhage, hard exudates, and retinal detachment. Fig. 1 shows retina images with different DR levels such as (a) normal, (b) mild, (c) moderate, (d) severe, and (e) proliferative.

It is projected that 14 million people will have DR in the US by 2050 [3]. If the detection of DR is not conducted at earlier stages, it may result in various degrees of vision impairment and even blindness. Therefore, a diabetic person must have an annual eye screening. Since developing countries suffer from high DR percentages, the lack of equipment is the main barrier to early

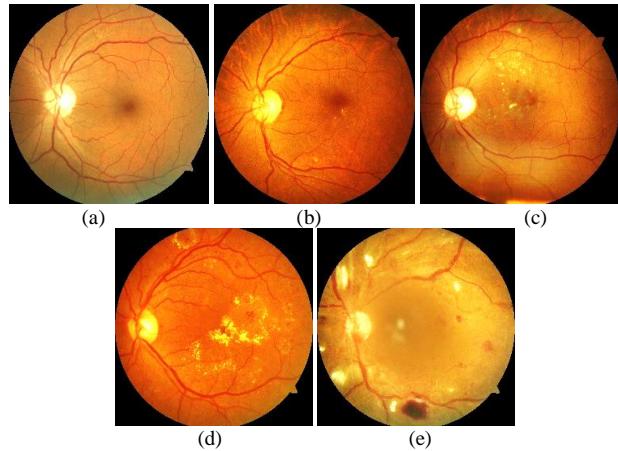


Fig. 1. Retina images from the UoA-DR dataset with different DR levels, (a) normal, (b) mild, (c) moderate, (d) severe, and (e) proliferative.

diagnosis of DR. Besides, patients in rural areas may not have access to the state-of-the-art diagnosis devices, such as fundus cameras. Even if they have enough equipment, image analysis can take 1-2 days by an ophthalmologist. Hence, there is a growing demand for portable and inexpensive smartphone-based devices and automation of detecting such eye diseases.

* Corresponding author. E-mail: mkarakaya@uca.edu

TABLE I
SPECIFICATIONS OF SMARTPHONE-BASED RETINAL IMAGING SYSTEMS



Fig. 2. Smartphone-based retinal imaging systems available in the market, (a) iExaminer (b) D-Eye, (c) Peek Retina, and (d) iNview.

Recent advances in computing and imaging technologies have enabled scientists to design small-sized, low-power, and affordable biomedical imaging devices using smartphones. These devices are capable of imaging, onboard processing, and wireless communication. Since they make existing systems small and portable, smartphone-based systems are widely used in several applications, ranging from health care to entertainment. Due to their large size, heavy weight, and high price, traditional fundus cameras are a good candidate to be transformed into a portable smartphone-based device to perform fast DR screening. The development of smartphone-based portable retinal imaging systems is an emerging research and technology area that attracts several universities and companies.

Holding a 20D lens in front of a smartphone camera is the simplest smartphone-based design to capture retina images [4]. Welch Allyn developed the iExaminer [5] system by attaching a smartphone to a PanOptic ophthalmoscope as shown in Fig. 2(a). These systems are built by attaching a smartphone to an existing medical device. There already exist several standalone designs for smartphone-based retinal imaging in the market including D-Eye, Peek Retina, and iNview. D-Eye [6] is the smallest retinal imaging system to capture retina images as an attachment to a smartphone as shown in Fig. 2(b). It illuminates the retina using the reflection of the smartphone's flashlight next to the camera without requiring additional external light and power sources. Its optics design allows it to capture images at 20 degrees in angle for dilated eyes. To simplify the design and to have evenly distributed illumination, the Peek Retina system [7] uses a circular placed multiple-LED light source to illuminate the retina as shown in Fig. 2(c). The iNview [8] was developed by Volk Optical as a new wide-angle smartphone-based retinal imaging system as shown in Fig. 2(d). For illumination, since iNview uses the reflection of the smartphone's flashlight, it does not require external light. Also, iNview can visualize the entire posterior pole in a single image by capturing 50 degrees of retinal view. Table I summarizes the hardware specifications of the publicly available smartphone-based imaging systems. Also, iExaminer, D-Eye, and iNview have Food and Drug Administration (FDA) approval. However, Peek Retina is currently waiting for its approval. Although these smartphone-based systems can capture retina images, none of them offers a solution to evaluate disease by analyzing the images with machine learning and image

processing methods.

Since deep learning techniques, especially Convolutional Neural Networks (CNNs), are an emerging research area, different research communities have already applied CNNs for several applications, including DR detection [9]. Deep learning is widely used for image classification tasks using neural networks that calculate hundreds of mathematical equations with millions of parameters. Recent works in the literature related to DR detection have mainly focused on designing new algorithms for traditional fundus images that are primarily affected by occlusion, refraction, variations in illumination, and blur. Kaggle competition is one of the important breakthroughs for DR detection where the EyePACS retina image dataset was presented with 35,126 training and 53,576 testing images. It attracted researchers and data scientists all over the world where several deep learning solutions were presented to detect DR.

Abramoff et al. [10-11] developed the Iowa Detection Program using their dataset and Messidor-2 dataset for training and testing. They have presented a variety of DR definitions such as referable Diabetic Retinopathy (rDR), vision-threatening Diabetic Retinopathy (vtDR), and referable Diabetic Macular Edema (rDME). They also reported high detection performance for rDR and vtDR. Gulshan et al. also developed CNN based deep learning frameworks for DR detection [12]. They trained the Inception-v3 architecture [13] with 128,175 images from EyePACS and Messidor-2 datasets and achieved high sensitivity and specificity. Gargya et al. [14] used a customized CNN architecture to classify images into two categories: healthy vs. others with any DR stage. They trained their network with 75,137 fundus images from their dataset, tested with Messidor-2 and E-Ophtha datasets, and achieved high accuracy.

Instead of training the CNNs from scratch, the transfer learning approach was used for pretrained deep learning frameworks [15-19]. Lam et. al. [15] proposed using pretrained CNN-based deep learning frameworks to detect DR using various classification models including but not limited to 2-ary, 3-ary, and 4-ary. They investigated the transfer learning approach for AlexNet [16] and GoogLeNet [17] using images in EyePACS and Messidor-1 datasets. They suggested using image pre-processing to increase validation accuracy, especially for the detection of mild DR. They augmented the retina images to increase the number of images in the training set and to prevent overfitting. Their results showed high sensitivity and specificity. Pires et al. [18] also proposed using transfer learning techniques for rDR detection. For training, they applied data augmentation, multi-resolution, and feature extraction to images in EyePACS dataset. They tested the network with Messidor-2 dataset and showed high rDR detection accuracy. Besides, Li et al [19] presented the binary and multi-class DR detection methods using the transfer learning for the Inception-v3 network. They trained the network with 19,233 images from their dataset and tested with Messidor-2 dataset. Their high accuracy results were comparable with the accuracy of three independent experts.

EyeArt is a cloud-based retina image assessment tool to detect DR using deep learning. It is capable of image description, image normalization, image rejection, region of interest detection, and descriptor computation. Solanki et al. [20] tested EyeArt with Messidor-2 dataset and achieved high accuracy. Rajalakshmi et al. [21] presented an early work to detect DR using EyeArt at retina images captured by Fundus On Phone (FOP) device. FOP proves the concept of smartphone-based designs and shows the technological and economic feasibility of the portable retinal imaging systems. Although all these related works achieved superior performance with high-quality fundus images, there were some limitations for smartphone-based retinal images. Due to their fewer controllable parameters and inexpensive lenses, smartphone-based systems have a smaller field of view and lower image quality compared to the fundus camera and FOP. Also, some existing methods [10–14] trained the CNNs from scratch that required very large labeled retina images and an extremely long time for the training process. Therefore, the existing approaches could not be applied directly to the retina images captured with smartphone-based systems because the field of view and image quality play important roles at the accuracy of the deep learning frameworks.

To address the above challenges and maximize the clinical utility of smartphone-based systems, this study explored the deep transfer learning frameworks for automatic DR detection. Our motivation in this paper is to develop an automatic DR detection model for smartphone-based retinal images using the deep learning approach with the pretrained networks. The main contributions of this article are two-fold: (i) to improve DR detection accuracy using the deep transfer learning approach for the pretrained networks with publicly available datasets and (ii) to study the effect of the Field of Views (FoVs) of smartphone-based retinal imaging devices. This study, with its high accuracy, high sensitivity, and high specificity, could help to design affordable and portable retinal imaging systems attached to smartphones that can be used by a variety of professionals ranging from ophthalmologists to nurses. It allows distributing quality eye care to virtually any location with the lack of access to eye care. Since recent patients are more involved in the monitoring and care of their diseases, there is an increasing trend in at-a-distance or telemedicine efforts to provide health care services for individuals living in far rural areas. For example, the teleophthalmology program based on the Joslin Vision Network was designed for DR screening and showed that it is a less costly and more effective strategy to examine the DR than conventional clinical-based screening [22]. This is clear evidence that smartphone-based retinal imaging systems will improve the technical capability and clinical practice for DR screening, increase the rate of access to DR imaging, and will help to decrease blindness due to DR even for individuals at distant locations from the health care facilities.

2. Methods

This section presented the general structure of the utilized deep learning architectures using transfer learning approach. Deep learning is capable of learning those structures by extracting the required information from the network using training images. It does not require extracting vein structures and identifying lesions such as exudates, microaneurysms, and hemorrhages at the retina for diabetic retinopathy detection. Therefore, training is an essential part of any deep learning system where the network needs to be feed with thousands of images to learn from their pixel values and edges. Since none of the publicly available datasets have enough retina images to train such a big network from scratch, this study utilized pretrained

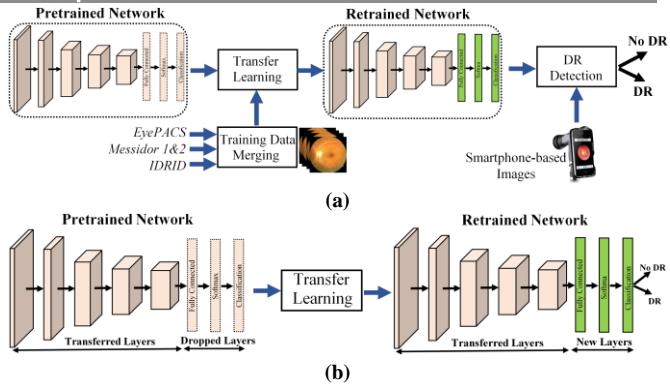


Fig. 3. (a) Flow chart of the proposed smartphone-based DR detection method and (b) Transfer learning approach for deep learning architectures.

networks using the transfer learning approach. The flowchart of the major steps in the proposed DR detection approach for smartphone-based images is shown in Fig. 3(a).

AlexNet, GoogLeNet, and ResNet50 are the well-known CNN architectures used for classification tasks. They were trained on ImageNet dataset [23] with millions of images to classify them into 1,000 different classes such as a keyboard, mouse, and several species of animals with a very low error rate. AlexNet, GoogLeNet, and ResNet50 consist of 25, 144, 177 layers in MATLAB, respectively. For the memory efficiency in training, they start with traditional deep learning fashion using convolutional layers followed by activation layers and max-pooling layers to extract the low-level features. AlexNet is the shallowest network with five convolutional layers to extract low-level features. GoogLeNet stacked nine inception modules upon each other with a block encapsulation where different sizes of filters (1x1, 3x3, and 5x5) are used for capturing both low-level and high-level spatial features at different scales. ResNet50 [24] is the deepest network where skip connection was introduced to feed the input from the previous layer to the next layer without any modification. Both GoogLeNet and ResNet50 use 1x1 convolutional layers to reduce the computational complexity by preventing feeding a large number of inputs from the previous layer to the next. This dimension reduction method helps to reduce the model size and to decrease the number of parameters from 138 million to 4 million as an increasing depth of architecture with more layers and units at each stage. The extracted features in the last fully connected layers are fed into a classifier such as Naïve Bayes, Random Forest, and Support Vector Machines to make decisions. Finally, the softmax layer classifies images into different classes based on the highest probability. For training, the weights and biases are updated at each iteration.

This study adapted the transfer learning approach for pretrained networks including AlexNet, GoogLeNet, and ResNet50. For transfer learning, the last three layers from the pretrained networks were replaced with new fully-connected, softmax, and classification layers as shown in Fig. 3(b). The classification layer has two classes since images are separated into two classes: DR and No DR. To speed up network training and prevent overfitting, the first 110 layers of transferred GoogLeNet and ResNet50 networks are frozen by setting their learning rates to zero. The parameters and weights of the remaining layers are allowed to update during training. The new network is retrained with the retina images using Stochastic Gradient Descent (SGD) algorithm with a learning rate of $1e-5$, a momentum of 0.9, and a minibatch size of 8, 16, and 32 examples. The number of max epoch in experiments was set to 32, 64, and 128, depending on images in the training set.

TABLE II
RETINA IMAGE DATASETS WITH DR SEVERITY LABELS

Datasets	Label0	Label1	Label2	Label3	Label4	Total
<i>EyePACS</i>	25810	2443	5292	873	708	35126
<i>EyePACS-u</i> *	9895	899	2175	568	317	13624
<i>Messidor</i>	547	149	240	251	-	1187
<i>Messidor-2</i>	1017	270	347	75	35	1748
<i>IDRiD</i>	168	25	168	93	62	516
<i>UoA-DR</i>	56	9	50	55	30	200

**EyePACS-u*: EyePACS-Updated

TABLE III
LABEL ASSIGNMENTS FOR DR SEVERITY

Datasets	Label0	Label1	Label2	Label3	Label4
<i>EyePACS</i>	No	Mild	Moderate	Severe	Prolif.*
<i>Messidor</i>	No	Mild	Mod-Sev.+	Prolif.*	-
<i>Messidor-2</i>	No	Mild	Moderate	Severe	Prolif.*
<i>IDRiD</i>	No	Mild	Moderate	Severe	Prolif.*
<i>UoA-DR</i>	No	Mild	Moderate	Severe	Prolif.*

*Mod-Sev: Moderate and Severe, *Prolif.: Proliferative

3. Experimental setup and datasets

This study was carried out using several publicly available retina image datasets, including EyePACS [25], Messidor [26], Messidor-2 [27], IDRiD [28], and University of Auckland Diabetic Retinopathy (UoA-DR) [29-30]. EyePACS is the largest publicly available dataset that was offered during Kaggle competition with 35,126 retina images that includes five different DR severity labels. Messidor DR dataset contains 1,187 images with four labels and DME grades. Messidor-2 dataset is an extension of Messidor dataset that includes 1,058 images from Messidor dataset and 690 new images. A total of 1,748 images in Messidor-2 are graded into five labels by a panel of three retina specialists. Indian Diabetic Retinopathy Image Dataset (IDRiD) has 271 retinal images and its DR severity assigned to five classes. UoA-DR dataset has 200 retina images and provides detailed DR and DME severity scales as well as information about neovascularization, hemorrhage, and microvascular abnormalities. Using this information, UoA-DR dataset was categorized into five DR classes. Table II shows the number of images in each data label for these datasets.

Retina images in EyePACS, Messidor-2, IDRiD, and UoA-DR datasets are graded according to the International Clinical DR scale [31]. This scale classifies the retina images into five classes including None, Mild DR, Moderate DR, Severe DR, and Proliferative DR as shown in Fig. 1. However, Messidor dataset is graded into four labels based on the existence of neovascularization and the number of microaneurysms and hemorrhages. Table III shows the available data labels in each dataset. Originally, there are five different DR labels in each dataset except the Messidor dataset. When the grader classifies images into several groups, it is very common to make an incorrect grading decision, especially for the mild and moderate DR images. To remove the inconsistencies in grading and transfer the problem into an easier domain, Abramoff proposed to group the images into two labels based on the referable Diabetic Retinopathy (rDR) and vision-threatening Diabetic Retinopathy (vtDR) standards. In the rDR approach, images with moderate, severe, and proliferative DR labels are merged into a single label (rDR) and compared with normal (No DR) retinas. The rDR label also includes referable DME and ungradable images. The vtDR is another approach where it drops out moderate DR and ungradable images from the rDR and classifies retina images into

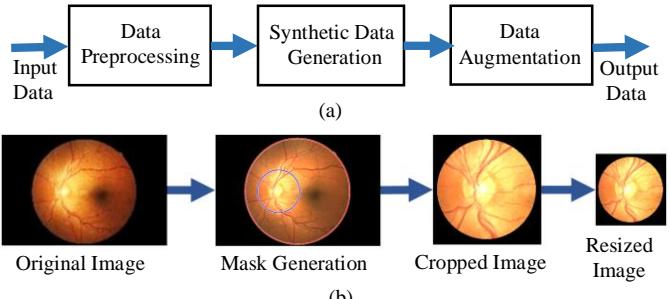


Fig. 4. (a) Workflow of synthetic retina image generation for smartphone-based retinal imaging systems and (b) Steps of synthetic data generation with masking, cropping, and resizing the input images.

normal and vtDR classes. Similar approaches were used in our work by classifying retina images into two classes. Therefore, this study tested several approaches tested including (1) normal retina vs. proliferative DR, (2) normal retina vs. severe and proliferative DR, (3) referable diabetic retinopathy (rDR), and (4) vision-threatening diabetic retinopathy (vtDR).

To investigate the DR detection accuracy for smartphone-based retinal imaging systems and compare them with traditional fundus imagery, two sets of experiments were conducted using original and synthetic retina images. Fig. 4(a) shows the flow chart of synthetic retina image generation for smartphone-based retinal imaging systems. Data preprocessing is required before using retina images in experiments because images in each dataset are captured by different image acquisition devices such as Canon, Centervue DRS, Optovue iCam, and Topcon NW cameras. Pupil dilation levels also might be different for each image. Also, some images include darkness, reflections, lack of contrasts, and even lack of optic nerve. For the data preprocessing step, the images were removed from the dataset when the optic disk is not visible in the image and there is an imbalanced classification problem due to the different number of images in each label in the training. Also, the resolution of retina images in the dataset varies since they were captured by different fundus cameras. Therefore, 21,502 images were removed from the EyePACS dataset.

To train and test the deep learning networks requires retina images from smartphone-based imaging systems. However, there is no publicly available data captured by any smartphone-based retinal imaging devices. Besides, pretrained frameworks require the inputs to have a certain size as color images. Therefore, synthetic retina images were generated by simulating the field of view (FoV) for different smartphone-based devices using the original retina images from UoA-DR dataset. Fig. 4(b) shows the steps of synthetic data generation where input images are masked, cropped, and resized for the required size. First, a circular mask was created around the center of the optic disc based on the different ratios of FoVs ranging from 20% to 90% with a step size of 10%, compared with the original images. The mask radius is calculated by multiplying the radius of the original image boundary and the percentage of the radius of FoV. Fig. 5(a) shows each circular mask representing the different FoVs to compare the difference in smartphone-based systems. The dotted yellow line represents the 20% FoV and the solid green line shows the 90% FoV. Finally, the original image was cropped at the mask center as a square. Then, the cropped square image was down-sampled into the required size. Examples of generated smartphone-based synthetic images for different FoVs are shown in Fig. 5(b-i).

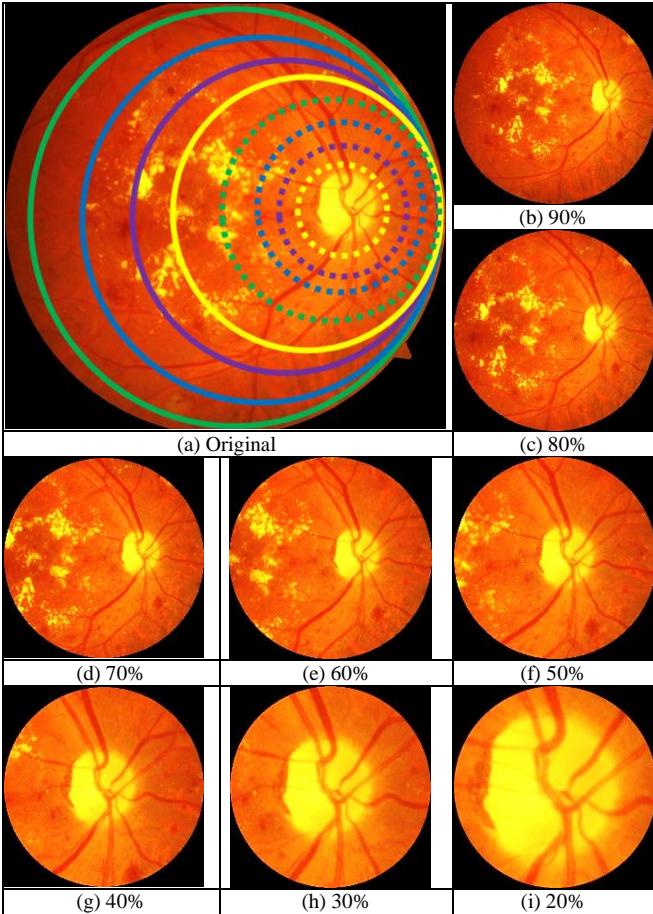


Fig. 5. (a) Comparison of the FoV of synthetic images with different percentages w.r.t the original image where the solid green, blue, purple, and yellow lines represent 90%, 80%, %70, and %60 FoV, respectively. The dotted green, blue, purple, and yellow lines represent 50%, 40%, 30%, and 20% FoV, respectively. The corresponding synthetic images for (b) 90%, (c) 80%, (d) 70%, (e) 60%, (f) 50%, (g) 40%, (h) 30%, (i) 20%.

Data augmentation is a very useful technique to prevent overfitting and bias, especially for the deep learning networks that require large datasets. Therefore, data augmentation is crucial for a small number of training images. With the data augmentation, the user acquires more data by applying an affine transformation to the existing images. Some of the data augmentation operations include, but not limited to: filling value, random rotation, reflection, scaling, shearing, and translation. These operations might be essential for some object recognition tasks for searching the different locations of the images. However, having the entire retina image is necessary for more accurate DR detection, especially the surroundings of the optic disc, fovea, and macula. Therefore, in our experiments, the only vertical flip was used to get a mirror image and both original and mirror images were included in the experiments. After data preprocessing, synthetic data generation, and data augmentation, retina images in each dataset were split into training and validation sets with a ratio of 0.9. The training sets include a maximum of 4,500 images.

For DR detection performance analysis of the deep learning frameworks, several experiments were designed using seven combinations of retina datasets for training and validation: (1) train and validation with only EyePACS, (2) train and validation with only Messidor, (3) train with EyePACS and validation with Messidor, (4) train with Messidor and validation with EyePACS, (5) merged datasets (EyePACS and Messidor), (6) rDR detection with the merged dataset (EyePACS, Messidor, Messidor-2, and IDRiD), and (7) vtDR detection with Merged dataset (EyePACS,

TABLE IV
ACCURACY FOR DR DETECTION OF DEEP LEARNING FRAMEWORKS

Networks / Related Works	VALIDATION			TESTING		
	ACC, %	SEN, %	SPE, %	ACC, %	SEN, %	SPE, %
AlexNet	95.6	92.8	98.3	91.4	97.6	82.5
GoogLeNet	93.6	90.7	96.4	94.5	99.7	86.8
ResNet50	96.2	93.9	98.4	98.6	98.2	99.1
Abramoff [11]	-	-	-	-	96.8	87.0
Gulshan [12]	-	90.3	98.1	-	87.0	98.5
Lam [15]	-	95.0	96.0	74.5	-	-
Li [19]	-	-	-	98.6	99.3	98.5

Messidor, and IDRiD). In this paper, the deep neural network was trained, validated, and tested with images from single, crossed, and merged datasets.

For all our experiments, algorithms were developed in MATLAB 2019 using the MatConvNet [32], deep learning, and image processing toolboxes. The experiments were run on a SkyTech Prism workstation with 8 core Intel i9 9900K processor at 3.6GHz, NVIDIA GeForce RTX 2080 with 11GB GPU, and 16GB memory. For example, training time for transfer learning of ResNet50 using the merged dataset (EyePACS, Messidor, and IDRiD) was around 877 seconds for 2,840 images with 64 epochs. The testing time per image was around 0.032 seconds.

4. Results and Discussion

This section first presented the results of our pretrained networks for the original fundus camera images to investigate their strengths and weaknesses by comparing them with the published works to support the novelty of our proposed approach. Second, we investigated the effect of using retina images from the single, cross, and merged datasets in training and validation. Third, these results were also compared with the smartphone-based synthetic retina images to explore the effect of FoVs for smartphone-based retinal imaging systems on the DR detection accuracy.

First, AlexNet, GoogLeNet, and ResNet50 models were trained on the merged dataset (EyePACS, Messidor, and IDRiD) and tested with UoA-DR dataset for vtDR detection. Table IV shows the overall accuracy, sensitivity, and specificity of our proposed networks from the validation and testing and compared with similar existing works in the literature. AlexNet showed better performance for validation compared with GoogLeNet. However, its accuracy for test images dropped and became lower than GoogLeNet because AlexNet is the shallowest network among others. Besides, ResNet50 reached the highest accuracy of 98.6%, the sensitivity of 98.2%, and specificity of 99.1% for test images since it is the deepest network with a larger number of layers than others. These results are comparable with the results of recently published related works [11, 12, 15, and 19] to authenticate the contribution of the proposed method. Abramoff et al [11] and Gulshan et al. [12] trained their CNNs from scratch using a very large dataset and their sensitivity was 96.8% and 87%, and specificity was 87% and 98.5% for testing, respectively. Also, Lam et al. [15] used transfer learning to retrain AlexNet and GoogLeNet where they achieved a sensitivity of 95% and specificity of 96% for validation and accuracy of 74.5% for testing. Li et al [19] explored the deep transfer learning method using the Inception-v3 network. For their testing, the accuracy was 98.6%, the sensitivity was 99.3%, and specificity was 98.5%. The comparison with the result of these existing studies also proves the effectiveness and efficiency of our proposed ResNet50 framework by showing state-of-the-art accuracy levels.

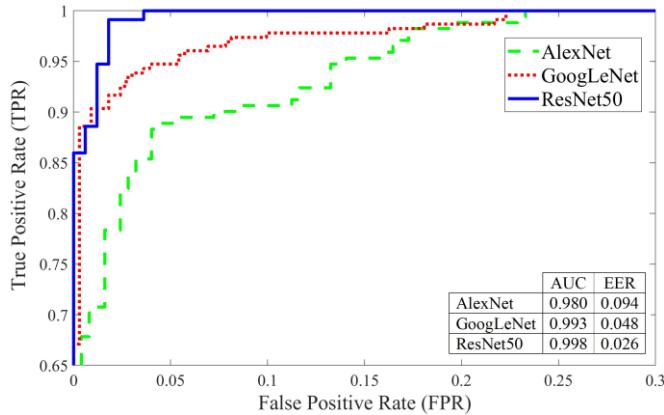


Fig. 6. Performance analysis using ROC for AlexNet, GoogLeNet, and ResNet50 frameworks from original retina images.

For better visualization of the performance analysis, the ROC curves of AlexNet, GoogLeNet, and ResNet50 models were presented in Fig. 6. A ROC curve was plotted by calculating the true positive rate (TPR) and the false positive rate (FPR) for different threshold values at the probability output of deep learning networks. TPR is the probability of detecting healthy (No vtDR) images as healthy. FPR is the probability of a false alarm where a healthy retina image is categorized as a disease (vtDR). The area under the curve (AUC) values of these ROC curves were 0.980, 0.993, and 0.998 for AlexNet, GoogLeNet, and ResNet50, respectively. As shown in ROC curves, ResNet50 marked with a solid blue line showed the best performance. It was also observed that the network accuracy depends on the network depth of pretrained frameworks and it decreases as the number of layers becomes smaller. Since ResNet50 framework showed better performance compared with AlexNet and GoogLeNet, it was used for the rest of our experiments.

This study also explored the performance of CNNs by training and testing with different types of images from different datasets to show the network effectiveness with different training sets. Therefore, the data fusion capabilities of CNNs from different datasets were addressed to improve recognition performance. Besides, the network behavior was investigated for untrained retina images from different datasets. The following subsections first presented the results for the original fundus camera images. Then, the results for synthetically generated smartphone-based images were presented to show the effect of FoVs on DR detection accuracy.

4.1 Results for Original Fundus Camera Images

In our first set of experiments, seven combinations of datasets in training and validation were tested using the ResNet50 framework. The deep learning results were shown in Table V. Initially, the first and second sets of experiments were performed using single datasets. When the network was trained and validated with images from the same datasets, the overall DR detection accuracies of the network were 92.1% and 99.1% for EyePACS-u and Messidor datasets, respectively. Our third and fourth set of experiments presented results for cross datasets where the network is trained with retina images from one dataset and tested with images from another dataset. The accuracy of the network dropped to 69.7% for training with Messidor images and testing with EyePACS images. However, better accuracy results of 81.5% for training with EyePACS and validation with Messidor were observed. The main reason for performance drop is that training images in EyePACS dataset have lower quality images compared with images in Messidor dataset due to the

TABLE V CLASSIFICATION ACCURACY OF DEEP LEARNING FRAMEWORKS				
Datasets	Type	ACC, %	SEN, %	SPE, %
(1) EyePACS-u	Single	92.1	86.5	96.3
(2) Messidor	Single	99.1	98.3	100
(3) EyeP_Mess	Cross	69.7	34.5	97.4
(4) Mess_EyeP	Cross	81.5	57.7	99.3
(5) Mess_EyeP	Merged	94.6	88.5	98.3
(6) rDR	Merged	91.2	92.2	91.2
(7) vtDR	Merged	98.6	98.2	99.1

TABLE VI
ACCURACY FOR RDR AND VTDR DETECTION USING TWO OPERATING POINTS

	RDR			VTDR		
	ACC, %	SEN, %	SPE, %	ACC, %	SEN, %	SPE, %
High Sensitivity	91.5	93.3	90.0	94.9	97.0	93.5
High Specificity	91.7	90.0	92.5	96.9	92.3	100

reflections, dark regions, and low contrasts. Furthermore, there exist several inconsistencies in labeling in EyePACS images. Therefore, training with only EyePACS images and testing with other datasets resulted in lower accuracy. When these two datasets were merged for training and validation for the fifth experiment, the DR detection accuracy reached 94.6%. Finally, rDR and vtDR detections with merged dataset were tested in the sixth and seventh experiments. The detection accuracy for rDR was 91.2% with a sensitivity of 92.2% and a specificity of 91.2%. The vtDR detection accuracy was 96.2%, sensitivity was 93.9%, and specificity was 98.4%. It was observed that training deep networks with diverse images from different datasets improves DR detection accuracy.

The proposed method classifies images into two different classes based on the highest probability calculated in the softmax layer. Since there are only two classes, the image is classified as a healthy retina if its probability is higher than 0.5. However, equal probability might not provide the best performance. Therefore, ROC curves were used to make performance analysis in our experiments where accuracy is plotted based on the various thresholds. Based on the ROC curves, two operating points were selected. The first operating point was set for the best sensitivity and the second one for the best specificity. The sensitivity is the most important factor for medical research where it shows the rates of the successfully detected unhealthy retinas. Table VI shows the accuracy of rDR and vtDR detection using two operating points for high sensitivity and high specificity. Note that, EyePACS (Label 0 vs. 3-4), Messidor (Label 0 vs. 2), Messidor-2 (Label 3-4), and IDRiD (Label 0 vs. 2-3-4) images were used for rDR; and EyePACS (Label 0 vs. 3-4), Messidor (Label 0), Messidor-2 (3-4), and IDRiD (Label 0 vs. 3-4) images were used for vtDR. For the high sensitivity, it was observed that the rDR sensitivity reached 93.3% with a specificity of 90% and the vtDR sensitivity reached 97% with a specificity of 93.5%. For the high specificity, the rDR specificity increased to 92.5% with a sensitivity of 90% and vtDR specificity increased 100% with a sensitivity of 92.3%.

4.2 Results for Smartphone-based Images

The second set of experiments investigated the effect of FoVs on smartphone-based synthetic retina images. Based on the previous baseline results, the high DR detection performance was received for vtDR detection at the merged datasets. Therefore, deep learning network was trained with retina images from EyePACS, Messidor, and IDRiD datasets as the seventh

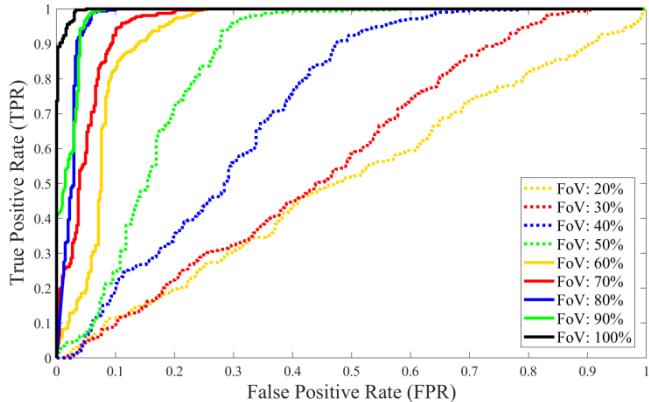


Fig. 7. Performance analysis using ROC for different testing images from original and synthetic smartphone-based retinal images with various percentage of FoV compared with original retina images.

experiment in the previous subsection. Also, to address the cross datasets issues in deep learning, the trained network was tested with smartphone-based synthetic images generated from a completely new dataset, UoA-DR with different FoVs ranging from 20% to 90% with a 10% step-size. To include smartphone-based synthetic images from PanOptic, D-Eye, Peek Retina, and iView systems, images from 32%, 40%, 45%, and 94% FoVs were also tested based on the calculations in this work [33]. Using images from UoA-DR dataset allows us to test the cross datasets without overlap between training and testing images.

Table VII presented the results of the vtDR detection performance of the ResNet50 framework for original and synthetic images with different FoVs. Since the network is trained with the original images, it shows the highest overall accuracy for testing with original images, as expected. The vtDR detection accuracy decreases from 98.6% to 51.2% as the FoV of the smartphone-based synthetic images gets smaller. It is observed that the sensitivity reduced slowly from 98.2% to 77.9% while FoVs decrease. However, the specificity declined very fast from 99.1% to 9.4%. The FoV affected the specificity of DR detection more aggressively than sensitivity because smaller FoVs covered only the optic disc and its surroundings. However, having a lesion in the retina is more likely to be close to the fovea than the optic disc. Therefore, it might be better to capture the surroundings of the fovea when the smartphone-based retinal imaging systems have a small FoV. The performance analysis of the ResNet50 deep network for original and synthetic images using ROC curves was presented in Fig. 7. Since TPR and FPR values change according to selected thresholds, any threshold value can be selected to lower the false alarm rate or increase the detection accuracy based on the specific system requirement. Equal error rate (EER) is the error value where false positive and false negative rates are equal to each for a specific threshold value in a ROC. For lower EER, the overall accuracy is higher. For original and synthetic images, calculated EERs range from 0.026 to 0.528 for the threshold values from 0.22 to 0.977. The area under the curve (AUC) values for the ROC curves changed from 99.8% to 50.6%. As shown in ROC curves, original retina images marked with a black line show the best result compared with smartphone-based images. Since the network is trained with original images that include all retinal structures such as the optic nerve, fovea, macula, and blood vessels, the deep network showed the best performance for images from datasets with larger FoVs. However, smartphone-based systems generally have narrower FoVs that cover smaller areas. Therefore, the network accuracy depends on the FoV and it decreases as the FoV becomes smaller.

TABLE VII
TEST RESULTS OF SMARTPHONE-BASED DEVICES AND FOV WITH UOA-DR

Datasets	ACC, %	SEN, %	SPE, %	AUC	EER	THR
Original	98.6	98.2	99.1	0.998	0.026	0.320
94% (IN)	95.7	93.8	98.1	0.988	0.038	0.220
90%	95.8	94.0	98.1	0.987	0.038	0.252
80%	95.9	95.6	96.2	0.985	0.038	0.384
70%	89.3	92.7	84.9	0.963	0.094	0.795
60%	83.5	91.2	73.6	0.929	0.113	0.956
50%	75.2	83.8	64.2	0.859	0.189	0.946
45% (PR)	68.6	79.4	54.7	0.817	0.245	0.825
40% (DE)	61.2	79.4	37.7	0.736	0.359	0.906
32% (PO)	54.6	77.9	24.5	0.574	0.491	0.938
30%	54.6	77.9	24.5	0.574	0.491	0.938
20%	51.2	83.8	9.4	0.463	0.528	0.977

TABLE VIII
ACCURACY OF VTDR DETECTION FOR DIFFERENT SIZE OF IMAGE BLURS

Size of Filter	Average Blur			Gaussian Blur ($\sigma = 3$)		
	ACC, %	SEN, %	SPE, %	ACC, %	SEN, %	SPE, %
No Blur	95.9	95.6	96.2	95.9	95.6	96.2
3x3	95.9	97.1	94.3	95.9	97.1	94.3
5x5	92.6	95.6	88.7	92.6	95.6	88.7
7x7	86.0	88.2	83.0	86.8	86.8	86.8
9x9	73.6	69.1	79.3	82.6	85.3	79.3

As an image quality assessment, the effect of the image blur on the accuracy of the deep network was investigated by adding average blur and Gaussian blur to smartphone-based synthetic retina images with 80% FoV. First, average filters and Gaussian filters were applied to test images at different sizes such as 3x3, 5x5, 7x7, and 9x9. For Gaussian filters, the standard deviation was fixed at 3 ($\sigma = 3$). Then, these blurred images were fed into the retrained ResNet50 architecture for vtDR detection. Table VIII shows the vtDR detection accuracy for different amounts of blur and compares them with the original synthetic image. It was observed that detection accuracies decrease as the blur increases for larger filters.

To design an accurate smartphone-based retinal imaging system for DR detection, this paper suggests capturing the retina images using a device with an FoV as large as possible and training the deep network with diverse images from different datasets. Moreover, further improvement might be possible using semi-automated systems and multi-modal classifier fusion. There exist several decades of experience in designing DR detection algorithms using traditional feature extraction methods and expertise of ophthalmologists to make the final decision. Semi-automated systems enable manual inputs of professionals with required medical education and solid experience in computerized systems [34]. This process will provide valuable feature extraction and ground truth information for the improvement of the DR detection accuracy. Multi-modal classification systems [35] might be another alternative approach to improve the accuracy and robustness of DR detection by fusing the combinations of different data and classifiers. Since the multimodality concept uses the complementarity between the different data and classifiers where each modality provides additional types of information to the system, their combination may show better results compared with using them separately. Therefore, the fusion of information from hand-crafted shape and texture features and convolutional neural networks trained with multiple datasets will improve the accuracy.

5. Conclusion

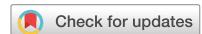
This paper presented the utility of CNN-based AlexNet, GoogLeNet, and ResNet50 frameworks to improve the performance of DR detection in smartphone-based and traditional fundus camera retina images. This study allowed us to compare the deep learning frameworks and to study the effect of FoVs in smartphone-based retinal imaging systems on their DR detection accuracy. Based on our results, the proposed ResNet50 approach showed the highest accuracy, sensitivity, and specificity for validation and test images compared with other frameworks and recently published related works. This also proves the effectiveness and efficiency of our proposed methods by showing state-of-the-art accuracy levels. DR detection accuracy was also improved by training networks with publicly available merged datasets. Although a smaller dataset was used in the training, considerably acceptable high accuracies were obtained. Also, the proposed ResNet50 model tested with different smartphone-based synthetic retina images from the UoA-DR dataset that were generated by simulating the different FoVs. It was observed that the DR detection accuracy increases as the FoVs get larger and deep networks are trained with images from different datasets. Since the FoV affected the specificity of DR detection more aggressively than sensitivity for images covering around the optic disc, capturing the surroundings of the fovea might be helpful for better sensitivity when the smartphone-based systems have a smaller FoV. However, there also exist several challenges for smartphone-based imaging systems due to the limitations of computational power, battery capacity, and camera properties in smartphones. For example, images captured with smartphone-based systems have lower quality and a narrower field of view compared with the traditional fundus camera because of the fewer controllable parameters, more sensitivity to illumination changes, and inexpensive lenses used in the design. Therefore, it is necessary to consider all challenging issues when designing algorithms for smartphone-based imaging systems.

Acknowledgments

This project was made possible by support from The Arkansas IDEa Network of Biomedical Research Excellence program with Award P20GM103429 from the National Institutes of Health/ National Institute of General Medical Sciences.

References

- [1] Bourne, R.R., et al., 2013. Causes of vision loss worldwide, 1990-2010: a systematic analysis. *Lancet Glob Health*, 6, 339-49.
- [2] Centers for Disease Control and Prevention, 2020. National Diabetes Statistics Report. US Department of Health and Human Services.
- [3] Kempen, J.H., et al., 2004. The prevalence of diabetic retinopathy among adults in the United States. *Arch. Ophthalmol.* 122(4), 552-63.
- [4] Haddock, L.J., et al., 2013. Simple, inexpensive technique for high-quality smartphone fundus photography in human and animal eyes. *Journal of Ophthalmology*, 518479-5.
- [5] Petruskin, H., et al., 2012. Optic disc assessment in the emergency department: a comparative study between the PanOptic and direct ophthalmoscopes. *Emergency Medicine Journal*, 29, 1007-1008.
- [6] Russo, A., et al., 2015. A novel device to exploit the smartphone camera for fundus photography. *Journal of Ophthalmology*, 823139-5.
- [7] Bastawrous, A., et al., 2016. Clinical validation of a smartphone-based adapter for optic disc imaging in Kenya. *JAMA Ophthalmology*, 134(2), 151-158.
- [8] User's manual for Volk iNView Retinal Camera, 2016. Volk Optical Inc. Mentor, OH: IM-088, Revision C.
- [9] Rajalakshmi, R., et al., 2015. Validation of smartphone based retinal photography for diabetic retinopathy screening. *PLoS one*, 10(9).
- [10] Abràmoff, M.D., et al., 2013. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmology*, 131(3), 351-357.
- [11] Abràmoff, M.D., et al., 2016. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative Ophthalmology & Visual Science*, 57(13), 5200-5206.
- [12] Gulshan, V., et al., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402-2410.
- [13] Szegedy, C., et al., 2016. Rethinking the inception architecture for computer vision. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2818-2826.
- [14] Gargya, R. and Leng, T., 2017. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 124(7), 962-969.
- [15] Lam, C., et al., 2018. Automated detection of diabetic retinopathy using deep learning. *AMIA Summits on Translational Science Proceedings*, 147-155.
- [16] Krizhevsky, A., et al., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 1097-1105.
- [17] Szegedy, C., et al., 2015. Going deeper with convolutions. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 1-9.
- [18] Pires, R., et al., 2019. A data-driven approach to referable diabetic retinopathy detection. *Artificial Intelligence in Medicine*, 96, 93-106.
- [19] Li, F., et al. 2019. Automatic Detection of Diabetic Retinopathy in Retinal Fundus Photographs Based on Deep Learning Algorithm. *Translational Vision Science & Technology*, 8(6):4, 1-13.
- [20] Solanki, K., et al., 2015. EyeArt: automated, high-throughput, image analysis for diabetic retinopathy screening. *Investigative Ophthalmology & Visual Science*, 56(7), 1429-1429.
- [21] Rajalakshmi, R., et al., 2018. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*, 32:6-1138.
- [22] Whited, J.D., et al., 2006. A modeled economic analysis of a digital teleophthalmology system as used by three federal health care agencies for detecting proliferative diabetic retinopathy. *Telemedicine Journal and e-Health*, 11(6), 641-651.
- [23] Deng, J., et al., 2009. Imagenet: A large-scale hierarchical image database. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 248-255.
- [24] He, K. et al. 2016. Deep Residual Learning for Image Recognition. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 770-778.
- [25] Graham, B., 2015. Kaggle Diabetic Retinopathy Detection Competition Report. Tech. Rep., University of Warwick.
- [26] Decencière, E., et al., 2014. Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology*, 33(3), 231-234.
- [27] Quellec, G., et al., 2008. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Transactions on Medical Imaging*, 27(9), 1230-1241.
- [28] Porwal, P., et al., 2018. Indian diabetic retinopathy image dataset (IDRid). *IEEE Dataport*, doi: 10.21227/H25W98.
- [29] Chalakkal, R.J., et al., 2017. Comparative analysis of University of Auckland diabetic retinopathy database. In Proc. of the 9th International Conference on Signal Processing Systems, 235-239.
- [30] Chalakkal, R.J. and Abdulla, W., 2017. Automatic segmentation of retinal vasculature. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 886-890.
- [31] Wilkinson, C.P., et al., 2002. International Clinical Diabetic Retinopathy Disease Severity Scale. In Proc. of American Academy of Ophthalmology Annual Meeting. Orlando, FL.
- [32] Vedaldi, A. and Karel, L., 2015. Matconvnet: Convolutional neural networks for Matlab. *Proceedings of the 23rd ACM International Conference on Multimedia*, 689-692.
- [33] Hacisofaoglu, R.E., and Karakaya, M., 2019. Field of view of portable ophthalmoscopes for smartphones. In Proc. of SPIE, Smart Biomedical and Physiological Sensor Technology XV, 110200X.
- [34] Chakraborty, S., et al., 2014. A Semi-automated System for Optic Nerve Head Segmentation in Digital Retinal Images. *International Conference on Information Technology*, Bhubaneswar, 112-117.
- [35] Benzebouchi, N. E., et al., 2019. Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis. *Journal of Experimental & Theoretical Artificial Intelligence*, 31(6), 841-874.



OPEN

Deep learning-based hemorrhage detection for diabetic retinopathy screening

Tamoor Aziz¹✉, Chalie Charoenlarpnopparut¹ & Srijidtra Mahapakulchai²

Diabetic retinopathy is a retinal compilation that causes visual impairment. Hemorrhage is one of the pathological symptoms of diabetic retinopathy that emerges during disease development. Therefore, hemorrhage detection reveals the presence of diabetic retinopathy in the early phase. Diagnosing the disease in its initial stage is crucial to adopt proper treatment so the repercussions can be prevented. The automatic deep learning-based hemorrhage detection method is proposed that can be used as the second interpreter for ophthalmologists to reduce the time and complexity of conventional screening methods. The quality of the images was enhanced, and the prospective hemorrhage locations were estimated in the preprocessing stage. Modified gamma correction adaptively illuminates fundus images by using gradient information to address the nonuniform brightness levels of images. The algorithm estimated the locations of potential candidates by using a Gaussian match filter, entropy thresholding, and mathematical morphology. The required objects were segmented using the regional diversity at estimated locations. The novel hemorrhage network is propounded for hemorrhage classification and compared with the renowned deep models. Two datasets benchmarked the model's performance using sensitivity, specificity, precision, and accuracy metrics. Despite being the shallowest network, the proposed network marked competitive results than LeNet-5, AlexNet, ResNet50, and VGG-16. The hemorrhage network was assessed using training time and classification accuracy through synthetic experimentation. Results showed promising accuracy in the classification stage while significantly reducing training time. The research concluded that increasing deep network layers does not guarantee good results but rather increases training time. The suitable architecture of a deep model and its appropriate parameters are critical for obtaining excellent outcomes.

The International Diabetes Federation (IDF) estimated that 700 million people will be living with Diabetes mellitus (DM) by 2045¹. DM develops fat and cholesterol in blood vessels that obstruct the flow of blood and nutrients required by human organs. The physiologic autoregulatory response to this progression increases the intracranial pressure of blood vessels²—This change in the retina yields rupture of small arteries that compiles Diabetic retinopathy (DR). DR is a retinal compilation that damages blood vessels, and DM is one of the causes of its development. DR cannot be cured permanently, but its progression rate can be reduced significantly by effective control of serum glucose, blood pressure and timely treatment.

DR is broadly categorized into non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). NPDR is classified by visible pathological features like microaneurysms (MAs), hemorrhages (HEs), exudates, and microvascular abnormalities. In comparison, PDR is diagnosed by the emergence of new blood vessels called neovascularization³. Figure 1 depicts two stages of DR.

MAs and HEs are grouped into red lesions. MAs are the earliest clinically manifest of DR and appear as small circular red dots in fundus images. Causes of MAs development are endothelial cell dysfunction, hypercoagulability, and atherosclerosis⁴. When MAs are weakened enough to rupture and swell, they rise to HEs. These dot-HEs are indistinguishable from MAs. HEs are disastrous for eyesight, predominantly when emerging in the macular region. Therefore, they often lead to significant and perpetual ocular impairment⁵.

Fluorescein angiography (FA), optical coherence tomography (OCT), and ophthalmoscopy are eminent methods of DR diagnosis. FA is a standard gold method and is capable of assessing microvascular changes. It helps to determine the origin of the leakage. Conversely, FA is time-consuming, invasive, and requires injection of intravenous fluorescein dye that may cause adverse side effects⁶. OCT is advanced for microvascular evaluation and captures a cross-section of the retina. OCT is non-invasive and does not require fluorescent dye injection.

¹Sirindhorn International Institute of Technology, Thammasat University, Bangkok, Thailand. ²Kasetsart University, Bangkok, Thailand. ✉email: tamoor.azi@dome.tu.ac.th

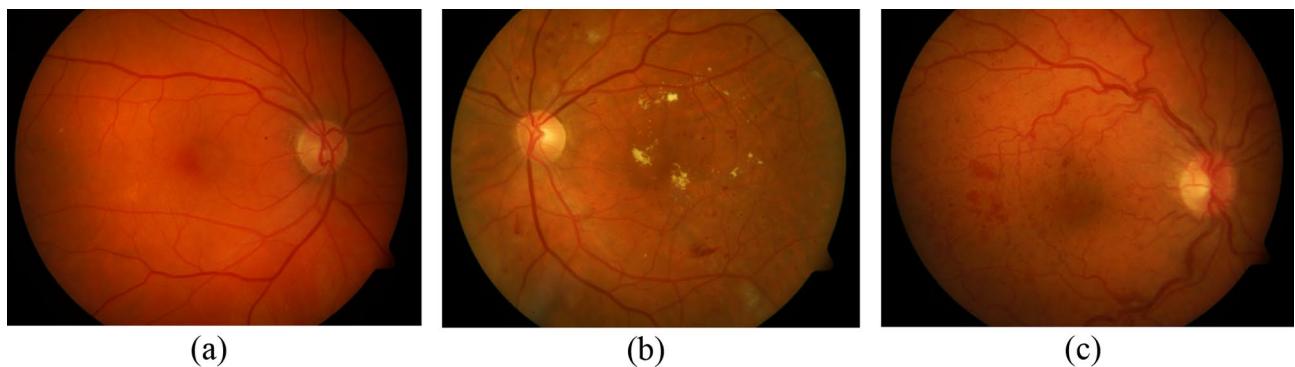


Figure 1. Stages of diabetic retinopathy in fundus images **(a)** Normal image **(b)** NPDR **(c)** PDR.

OCT assists in determining the progression rate of compilation and efficacy of therapy⁷, but complexities and risks are evident. Therefore, it necessitates computationally efficient, affordable, and non-invasive diagnostic method. Retinal ophthalmoscopy can be an alternative to address these problems⁸. It allows examination of retinal structures and can identify clinical pathological symptoms. This method does not constitute any risks; therefore, it is used for regular eye examinations.

Fundus images are often blurry and poorly illuminated. Intensity profiles of HEs appear similar to the macula, blood vessels, and dark regions caused by the lighting conditions. Therefore, the classification of HEs from other objects is a challenging task. Furthermore, the lack of a standardized description of HEs locations obstructs comparing various prognosis approaches. Localization of hemorrhage is a critical subject when prescribing treatment by an ophthalmologist. For instance, macular hemorrhage is more disastrous than other HEs⁹ because it deteriorates central vision. Therefore, the detection of HEs with their precise locations is exceptionally imperative. A fundus image may have many HEs with different sizes, and the intensity profiles of their surrounding regions are inconsistent. Hence, it demands an adaptive segmentation process to deal with differing sizes and inconsistent regions for effective HEs identification. Effective implementation of deep models requires sufficient data for training a model. A deep network provides obscure results that are difficult to interpret when trained on insufficient training samples. Therefore, these challenges compel the development of an efficient computer-based technique for HEs recognition that may act as a second interpreter. It may assist medical experts in prescribing appropriate treatment due to the heterogeneous nature of treatment modalities.

In this study, the impediments mentioned above to HEs detection are incorporated. Modified adaptive gamma correction by employing gradient information adjusts the image's brightness level for better contrast. The algorithm localizes prospective candidates with accurate locations using empirical image processing. This step eliminates redundant information and expedites the detection process by considering only those objects that are likely to be HEs. Besides, this localization process provides sufficient data by generating windows/patches for the training of the classification model. It helps to design a computer-aided design (CAD) that works efficiently for limited images. The HEs can be located anywhere with random sizes in the retinal region and surrounded by areas with different intensity profiles. Therefore, the novel smart window-based adaptive thresholding incorporates regional diversity and segments HEs regardless of their sizes and encompassing regions. Removing unrequired intensity information from the image makes the HEs classification task simple. Therefore, the shallow Hemorrhage net (HemNet) is designed to learn from deep features in the training stage and differentiates HEs from other retinal structures. Despite HemNet containing fewer deep layers, it is competitive with other deep models that are extensively deep. Its generalization capability is proven through synthetic experimentation, which shows that HemNet takes less training time and has higher evaluation metrics than other models.

The research work is organized as follows. Section one is the introduction and importance of the research problem. It illustrates the application of computer-aided design in ophthalmology, the impediments to detecting the HEs, and how difficulties are addressed to identify HEs automatically. Section two is the review of relevant methods proposed by the research community. Section three explains the propounded preprocessing stage for quality enhancement and discusses the estimation of prospective HEs candidates. Then the segmentation technique is explained thoroughly to generate sufficient windows/patches for training a novel HemNet. “[Experimental results](#)” section is about experimentation design and compares deep models using evaluation metrics. It also includes synthetic experimentation to validate the performance of HemNet. Section five is the conclusion and contains the points for future consideration.

Related Work

Li et al.¹⁰ comprehensively review fundus photography's deep learning applications. It has been reported that training convolution neural networks (CNN) is time-consuming. Khojasteh et al.¹¹ use probability maps provided by the SoftMax layer for retinal abnormalities classification. Enhancement and segmentation are performed, and then annotated patches train CNN. A trained classifier analyzes testing data where the probability map identifies the DR symptoms. In this study, 18,882 HEs-related examples are utilized for network training, which requires a long time to train CNN. A fully automated HEs detection method is presented by Lehmiri and Shmuel¹². Variational mode decomposition processes retinal fundus images to obtain high-frequency components. Four sets of texture features train a classifier that discriminates healthy from unhealthy images. Short processing time

with higher accuracy is reported. Orlando et al.¹³ detected red lesions using hand-crafted and deep features. The augmented ensemble vector and random forest (RF) classifier identify red lesions in fundus images. Three RF classifiers are trained using ensemble vector, hand-crafted, and deep features, and their performances are compared. Son et al.¹⁴ develop a method for the simultaneous detection of retinal pathologies that include HEs. A large amount of data trains CNN, and results are compared with ophthalmologists' manual annotations of retinal fundus images. CNN architecture provides probability and a heatmap, which is a normalized single-channel low-resolution image. This image contains lesions identified by the algorithm and is superimposed onto the original image to highlight lesions. Most importantly, the classification result correlates with the gray values of the heatmap to identify the locations. Gayathri et al.¹⁵ devised a novel CNN architecture to extract features from the retinal fundus images. Extracted deep features train various classifiers like support vector machine (SVM), AdaBoost, Naïve Bayes, RF, and J48. These classifiers are evaluated using specificity, precision, recall, false positive rate (FP), Kappa-score, and accuracy. The performance of the J48 classifier with the proposed feature extraction model is the best among all classifiers. Hacisofaoglu et al.¹⁶ developed an automated diagnostic model specially designed for low-quality images captured using smartphones' small field of view (FOV). The transfer learning approach employs well-known architectures of Alexnet, Googlenet, and Resnet50. These CNN models evaluate the effects of single, cross, and multiple datasets. The proposed Resnet50 model is applied to smartphone-based synthetic images, evaluated on an independent dataset, and yields promising results. Qureshi et al.¹⁷ propounded a label-efficient CNN called active deep learning (ADL-CNN) using expected gradient length (EGL). ADL-CNN selects critical samples by using ground truth labels for feature extraction. Then, retinal pathologies are segmented and graded according to severity levels. Hemanth et al.¹⁸ propounded a compound of image processing and a deep learning-based hybrid method for DR diagnosis. Image quality is enhanced using histogram equalizations. Each channel of a color image is enhanced and concatenated to improve quality. The sizes of images are normalized to 150×225 for the CNN classification stage. Four hundred retinal fundus images of the MESSIDOR dataset are used for validation, and the algorithm reports promising results.

Training CNN is time-consuming, and obtaining high accuracy is a challenging task. Some techniques increase classification metrics and reduce training time using various approaches. The first approach estimates required objects that are more likely to be objects of interest. The second approach includes and excludes training examples based on their contribution to classification. For instance, a two-stream red-lesions detection is proposed by Asiri et al.¹⁹. Regions of prospective candidates are extracted using vessel segmentation and morphological operations to reduce computational complexity. This preprocessing step yields better results because it explores prospective candidates, enhancing accuracy in the classification stage. Pre-trained visual geometry group (VGGNet) is tuned for vessel and potential candidates' segmentation. These candidates are classified using Faster RCNN, which provides promising results. Grinsven²⁰ proposed a technique for HEs detection to expedite the training process. Training examples are heuristically sampled and misclassified negative samples are dynamically selected. Performances of trained CNNs using selective sampling and without selective sampling are evaluated. The method reduces a substantial number of epochs and provides promising classification results.

The proposed detection scheme ventures similar approaches presented earlier to reduce training complexity. First, prospective HEs candidates are estimated to eliminate irrelevant objects. It reduces the number of examples that are to be explored by CNN and expedites the detection process. Secondly, windows of fundus images are segmented to remove unrelated information before being provided to a CNN model. This step reduces intensities from the windows, and CNN extracts the features related to the HEs. Therefore, this research evaluates various CNN models trained on data from a small dataset.

Method

Dataset description. Two datasets of DIARETDB1²¹ and DIARETDB0²² are employed for experimentation. The first dataset contains 89 fundus images, of which five are standard, and the rest have various retinal pathological symptoms. The second dataset contains 130 images, of which 110 have DR signs. These images are captured by the 50-degree field of view using a fundus camera under various illumination conditions. Figure 2 depicts multiple steps of the propounded detection scheme.

Methodology. *Quality enhancement.* Qualities of retinal fundus images vary due to different illumination conditions. A rigorous visual inspection of images reveals that excess light provides over-saturation in some regions, the edges of required objects are blurry, and insufficient light produces dark regions. Therefore, the qualities of fundus images are enhanced to reduce the effects of low-quality characteristics of digital images. First, the contrast of an image is enhanced using contrast-limited adaptive histogram equalization (CLAHE)²³.

An adaptive process is required to adjust brightness levels because some images have adequate brightness levels while some are dark. The brightness level is adjusted using gamma correction^{24,25}, and the gradient value is used to produce the adaptivity. This process is applied to the individual color channel. Let φ be a threshold value that separates smooth regions from the edges of the Sobel gradient image and is considered as the brightness interpretation of an image. Low correction is required when φ is large, which yields an adequate brightness level of the image. Conversely, high correction applies when φ is small, which reveals a low brightness level of the image. Therefore, γ is calculated by adjusting φ as:

$$\gamma = \alpha * \frac{\lfloor \varphi \times 100 \rfloor}{10} \quad (1)$$

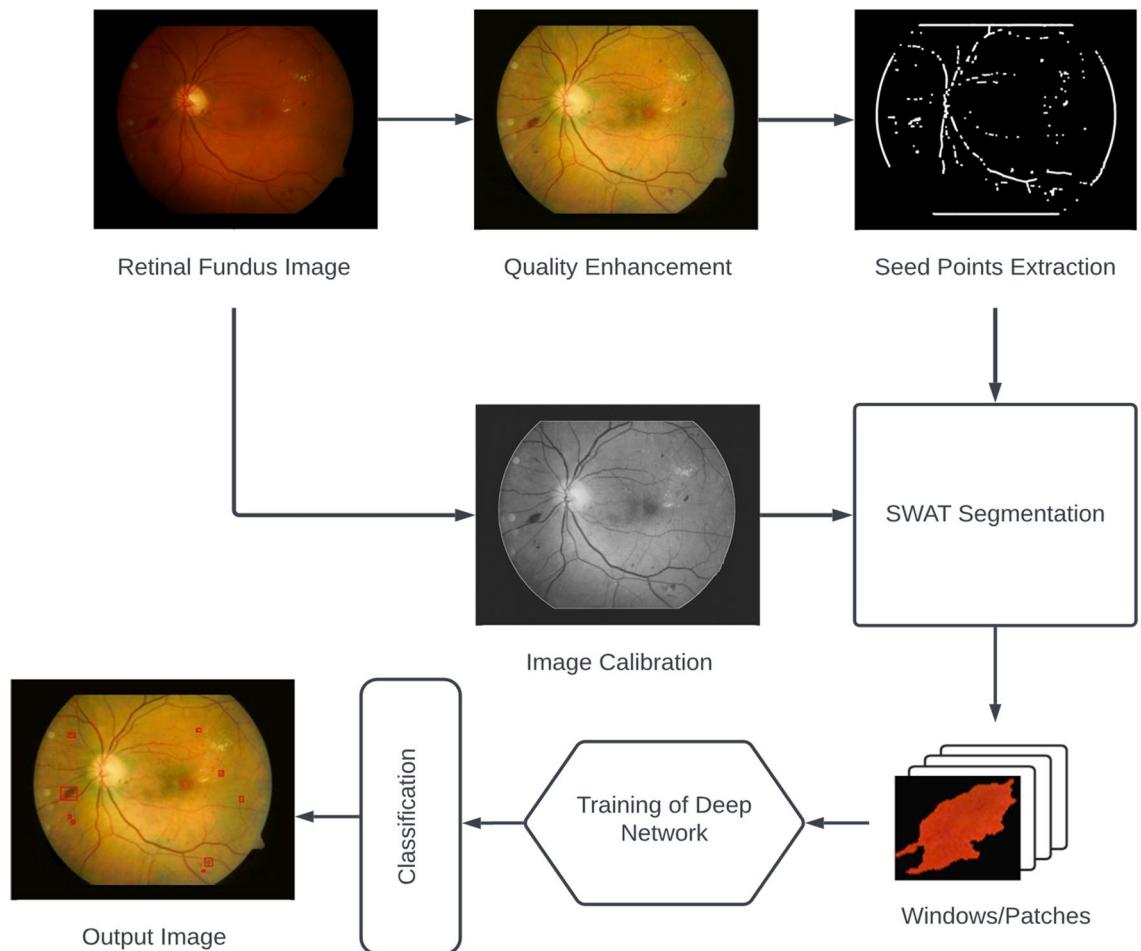


Figure 2. Illustration of the proposed detection technique.

$$V(x, y) = I_{\max} \left(\frac{I}{I_{\max}} \right)^{\gamma} \quad (2)$$

where α is the brightness adjustment coefficient and used as $\alpha = 2$. Equation (2) adjusts the brightness level of the image, where $I \in \{0, I_{\max}\}$ and I_{\max} is the maximum intensity of an image.

Blurriness reduces texture and edge information in the image. The green component is sharpened using non-linear unsharp masking to improve edge information. This method computes the intensity difference using the fuzzy relationship between focused and neighboring pixels in a 3×3 window. Pixels are sharpened using non-linear relationships depending on the luminance difference of the adjacent pixels²⁶. The sample of the quality enhancement method is given in Fig. 2.

Seed points extraction. The intended research work automatically detects HEs using seed points. Seed points highlight the locations of prospective HEs candidates. HEs share intensity information with blood vessels because of their similar appearances, and they are dark objects surrounded by bright regions. These characteristics lead to the development of an inverted Gaussian-based matched filter²⁷. The kernels of the matched filter are applied to enhance dark objects, including HEs. Kernels with 0° and 90° angles can be depicted in Fig. 3.

The matched filter effectively enhances HEs and blood vessels due to the strong correlation. However, it provides a low response to the intensity variations of the image due to a weak correlation. GLCM-based local cross-entropy thresholding²⁸ removes the low response of the matched filter. This method estimates the intensities into background to the background (BB), foreground to foreground (FF), background to the foreground (BF), and foreground to background (FB), given a threshold value t . BB and FF represent smooth regions, while BF and FB contain the information of edges. The edges do not contain substantial information; therefore, the entropies of BB and FF quadrants are used to find the optimum threshold value t^* , successively.

The formation of HEs starts from the leakage of blood vessels, so some of them can be attached to the blood vessels. Their isolation is critical for the early detection of DR. Mathematical morphology analyzes the spatial structures. The morphological opening is applied to break the larger objects than structuring elements. A square structure element of size 11×11 is used in our experiment. The image containing the seed points for subsequent segmentation is shown in Fig. 2.

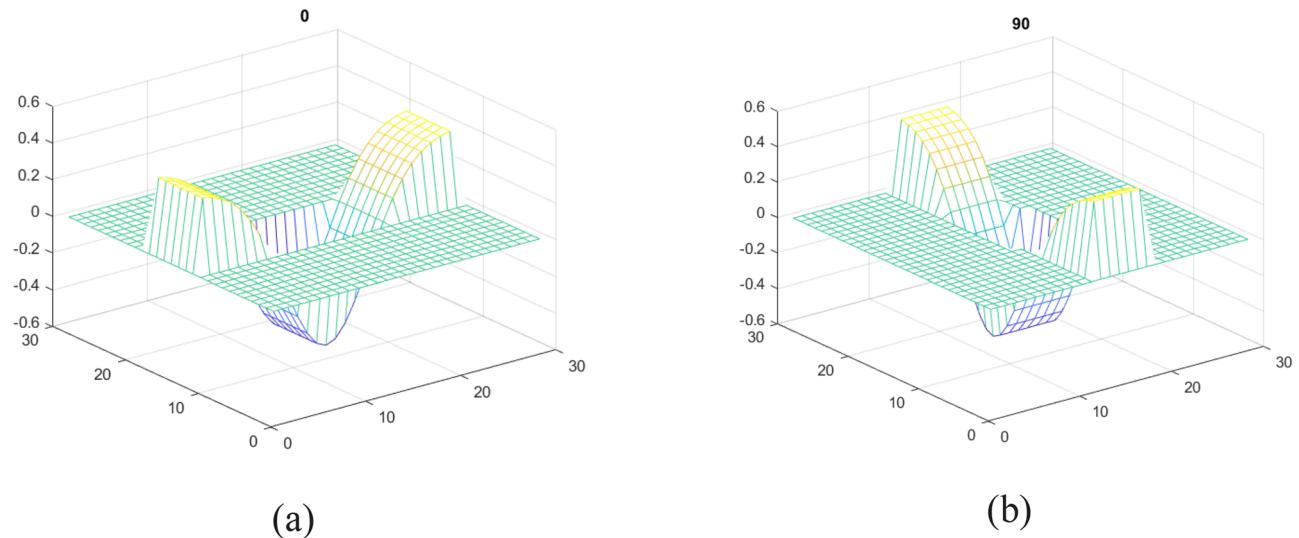


Figure 3. Rotation of matched filters **(a)** kernel with 0° angle **(b)** kernel with 90° angle.

Image calibration. HEs emerge at the retina's vitreous humour, and the black background does not contain any information. Black background misleads the detection process for the HEs at the retina's rim. Therefore, the illumination of the black background assists in automatic detection and reduces the search space. The median filter is applied on the green channel to suppress the random intensity variations on the black background then the image is binarized to get the retinal mask. The eroded mask is subtracted from the retinal mask to achieve the exact retinal boundary. The hand-crafted image is obtained by adding complimented retinal mask multiplied by the average gray value, retinal periphery, and the enhanced green channel. Figure 2 shows the calibrated image with a bright background where the retinal information is not undermined and is used for feature extraction.

SWAT segmentation. Segmentation of two types of HEs is challenging and requires a highly strenuous and intelligent method. First, the HEs that are located at the retinal periphery blended with the dark background. Secondly, those are attached to the blood vessels. The black background has been illuminated using image calibration that aids in performing segmentation at the retina's border. Smart window-based adaptive thresholding (SWAT) can sense gray dissimilarity between HEs and blood vessels. It also provides adaptivity in segmentation to deal with the HEs encompassed by various regions. SWAT uses Otsu's method for thresholding and finds the effectiveness value η using:

$$\eta = \frac{\sigma_B^2(\tau^*)}{\sigma_T^2} \quad (3)$$

where η is the ratio of inter-region variance $\sigma_B^2(\tau^*)$ to the total variance σ_T^2 of the image. The value of $\sigma_B^2(\tau^*)$ depends upon the selection of an appropriate number of regions q in a window $W(x, y)$. SWAT finds the optimum number of regions q , iteratively using Eq. (4), that yields maximum effectiveness value η within the range 0–1 and produces robustness in the segmentation process. q generates a vector ϑ containing $q - 1$ threshold levels. The window is binarized using Eq. 5.

$$\vartheta = \begin{cases} \varrho \rightarrow \varrho + 1, & \text{if } \eta < 0.8, \text{ AND } \varrho \leq 20 \\ \text{stop, otherwise} \end{cases} \quad (4)$$

$$W_1(x, y) = \begin{cases} 0, & \text{if } W(x, y) > \min(\vartheta) \\ 1, & \text{else} \end{cases} \quad (5)$$

HEs with bigger sizes are prioritized because they are more disastrous for eyesight than small HEs. Therefore, two more oversized objects are retained, and the rest are removed. This maneuver reduces the risk of false detection because dark shades are often bigger and may mislead the segmentation stage. Furthermore, the window originates from a seed point that probably belongs to a HE. This estimation criterion is proposed because seed points are extracted using the characteristics of HEs. Therefore, the object with a minimum distance is a HE, and the other one is removed by computing the Euclidean distance as:

$$d_i = \min \sqrt{\{W_1(x_c) - I_i(x)\}^2 + \{W_1(y_c) - I_i(y)\}^2} \quad (6)$$

where $I_i(y)$ and $I_i(x)$ are the y and x spatial locations of the i_{th} object, and $i = \{1, 2\}$. The object's spatial locations assist in segmenting the complete HEs using SWAT. The size of the bounding box is increased to the particular directions according to the border pixels using the following relations:

$$V = \begin{cases} v_1 \rightarrow v_1 - 5, \text{if } q_1 = 1 \text{ AND } v_1 \cap S \\ v_2 \rightarrow v_2 - 5, \text{if } q_2 = 1 \text{ AND } v_2 \cap S \\ v_3 \rightarrow v_3 + 10, \text{if } q_3 = 1 \text{ AND } v_3 \cap S \\ v_4 \rightarrow v_4 + 10, \text{if } q_4 = 1 \text{ AND } v_4 \cap S \end{cases} \quad (7)$$

where S is the search region to achieve automation in the detection process. The search region S restricts SWAT from searching within the image domain. The region S is obtained by extending the retinal mask eighty pixels wide in each direction. It provides sufficient space, especially for those HEs that reside at the retinal periphery. The vector Q contains binary variables q_1, q_2, q_3 , and q_4 for corresponding left, top, right, and bottom border pixels. Equation 7 updates the vertices v_1, v_2, v_3 , and v_4 of vector V , accordingly. The window sample in Fig. 2 is used to train a classifier.

Training of deep network. Some seed points belong to other retinal structures like blood vessels, dark shades, and intensity variations. SWAT segments and other retinal structures originated from those seed points and are redundant in our experiment. Therefore, it demands a classifier to discriminate HEs in the detection process. We propose a novel HE network (HemNet) that is shallower than conventional deep models and efficiently classifies HEs from other retinal structures. Our deep model contains nineteen individual layers, including input and output layers. Table 1 elaborates on the architecture of the proposed HemNet and can be depicted visually in Fig. 4.

After segmenting the required objects using seed points, windows are extracted by the vector V from an enhanced colored image. This colored data is labeled and transformed into HSV²⁹ and CIE Lab³⁰ color spaces inspired by the HEs detection using conventional features³¹. For instance, the edges of HEs are sharper than the

Layer type	Layer name	Kernel size	Number of filters	Stride length	Output shape
Image input	–	–	–	–	(256, 256, 3)
Convolution	conv_1	11 × 11	16	3	(86, 86, 16)
ReLU	relu_1	–	–	–	–
Batch normalization	batchnorm_1	–	–	–	(86, 86, 16)
Average pooling	avgpool2d_1	2 × 2	–	–	(86, 86, 16)
Convolution	conv_2	5 × 5	16	2	(43, 43, 16)
ReLU	relu_2	–	–	–	–
Batch normalization	batchnorm_2	–	–	–	(43, 43, 16)
Average pooling	avgpool2d_2	2 × 2	–	–	(43, 43, 16)
Average pooling	avgpool2d_3	5 × 5	–	–	((43, 43, 16))
Batch normalization	batchnorm_3	–	–	–	(43, 43, 16)
Addition	Addition	–	–	–	(43, 43, 16)
Convolution	conv_4	11 × 11	16	2	(22, 22, 16)
ReLU	relu_4	–	–	–	(22, 22, 16)
Batch normalization	batchnorm_4	–	–	–	(22, 22, 16)
Average pooling	avgpool2d_4	2 × 2	–	–	(22, 22, 16)
Fully connected	Fc	–	–	–	(1, 1, 2)
SoftMax	SoftMax	–	–	–	(1, 1, 2)
Classification output	Classentropy	–	–	–	(1, 1, 2)

Table 1. The architecture of the proposed deep convolution network.

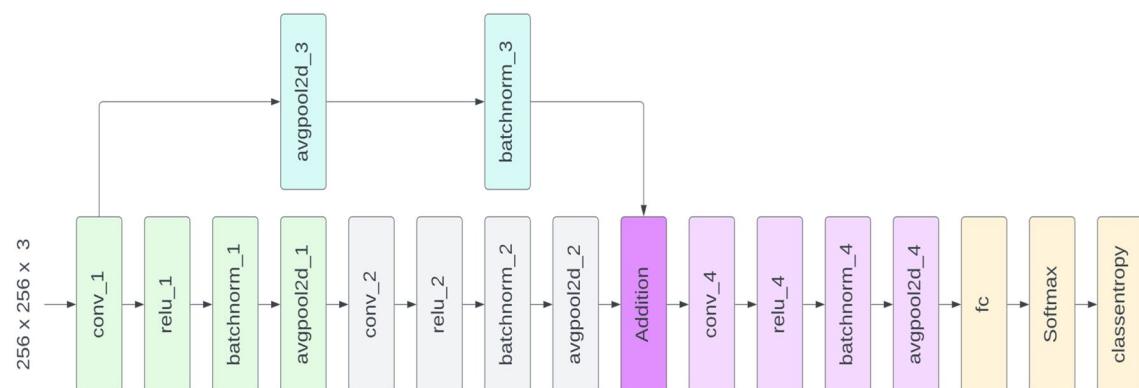


Figure 4. Overview of Proposed HemNet.

edges of the macula, so the first and second-order based gradient features distinguish HEs from the macula. The blood vessels are line-shaped retinal structures, and HEs are comparatively rounded-shaped elements. So, connected component-based shape features classify them. Color features discriminate the objects of interest from black shades in the green channel. Therefore, similar characteristics are transformed to train HemNet that models conventional features. Three channels are used to train the network: the green channel of RGB, the value channel of HSV, and the luminance channel of CIE Lab color spaces. In addition, surrounded regions of HEs are eliminated, and the segmented objects are used in the training so the deep network can effectively classify HEs from blood vessels and other retinal structures.

Experimental results

Experimental composition. The machine with a 2.4 GHz Core-i5 processor, 16 GB RAM, and a single 6 GB GPU is used for the implementation. The performance of our detection method is evaluated using the DIARETDB1 and the DIARETDB0 datasets. Images of the DIARETDB1 dataset that contain HEs are used for the training. Our proposed algorithm intends to detect HEs, and the interpretation of HEs from ground truths reveals that forty-five out of eighty-nine images contain hemorrhages. These forty-five images are separated into training and testing sets. The training set comprises twenty-five images, and the testing set includes twenty images. The training set is further divided into the training and validation sets. These examples are annotated using the ground truths. Twenty images of the DIARETDB1 dataset are tested randomly, and the results are compared. Twenty images of the DIARETDB0 dataset are randomly taken to independently benchmark the algorithm's performance. The classification results are determined using sensitivity (SE), specificity (SP), accuracy (AC), and precision (P). AC is an effective measurement for evaluating models' performance because it is a ratio of classified examples from the total number of samples³². These evaluation metrics are measured as follows:

$$SE = \frac{TP}{TP + FN} \quad (8)$$

$$SP = \frac{TN}{TN + FP} \quad (9)$$

$$P = \frac{TP}{TP + FP} \quad (10)$$

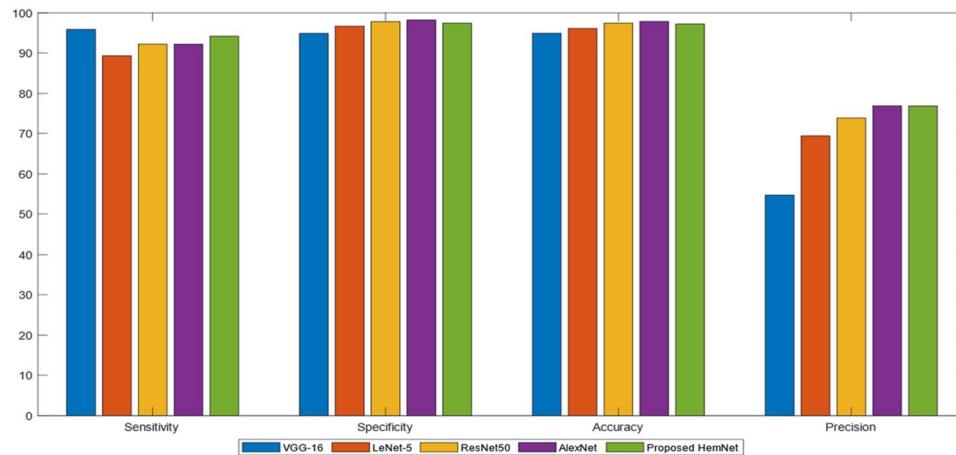
$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Time is a critical factor for screening and diagnosing diseases using medical images. Deep networks are efficient but often take a long time to train. Insufficient training of networks reduces classification accuracy. Therefore, the proposed network is analyzed using training time versus classification accuracy by synthetic experimentations using two datasets. Image of concrete crack³³ and Modified National Institute of Standards and Technology (MNIST)³⁴ digits datasets benchmark the HemNet. Thirty thousand images of the concrete crack dataset are included, where training, validation, and testing sets consist of twenty thousand, five thousand, and five thousand, respectively. Ten epochs are used with a learning rate of 0.009 in this trial. This experiment is assessed using training time, validation accuracy, and testing accuracy, and their statistics are provided in Table 4. Ten thousand images of the MNIST digits dataset are employed for the second synthetic experiment. One thousand images are equally distributed for validation and testing. While eight thousand images train the network using fifteen epochs and a 0.01 learning rate. The results of the digits classification are provided in Table 5.

Analysis of results. The performance of HemNet was compared with the state-of-the-art CNN models for HE detection and classification. Despite the fact that HemNet was a shallower network, it provided competitive results when compared with other deep networks like LeNet-5³⁵, AlexNet³⁶, ResNet50³⁷, and VGG-16³⁸ on the DIARETDB1 dataset. The SE of the HemNet was closer to the SE of VGG16, which was the highest, stating the true positive detection rate of HEs. While HemNet marked 94.21% SE, which was higher than LeNet-5, ResNet50, and AlexNet. It shows that HemNet identifies most of HEs with low false-negative rate. Additionally, the 94.76% SP of HemNet was slightly less than AlexNet but greater than VGG-16 and LeNet-5. The difference in SP was negligible when comparing ResNet50 and the proposed model. SP shows the misclassification of some segmented objects as HEs but they do not belong to this class. AC of ResNet50 and AlexNet was 97.46% and 97.88% and were the highest, but HemNet scored 97.19%. This difference was negligible and not critical because the SE of the proposed HemNet is higher than ResNet50 and AlexNet. It shows truly classified HEs detection of the HemNet. The AC of HemNet was greater than VGG-16 and LeNet-5. AlexNet and the proposed model marked high P of 76.90% and 76.87%, respectively. These statistics validated the excellent performance of HemNet in terms of misclassification rates; false-positives and false-negatives. The classification results of various deep learning models on DIARETDB1 are statistically compared in Table 2 and visualized in Fig. 5.

A SoftMax layer of a deep model normalizes each training data example into a probability distribution as a prediction score. These predicted probabilities can be distributed into regions by a threshold classification rule. The precision and recall values change with respect to threshold values. Therefore, it is critical to identify the optimum threshold. Using this evaluation criterion, the best classification model is considered to have a maximum area under the curve (AUC). The precision-recall (PR) curve plots the behavior of a deep model's precision

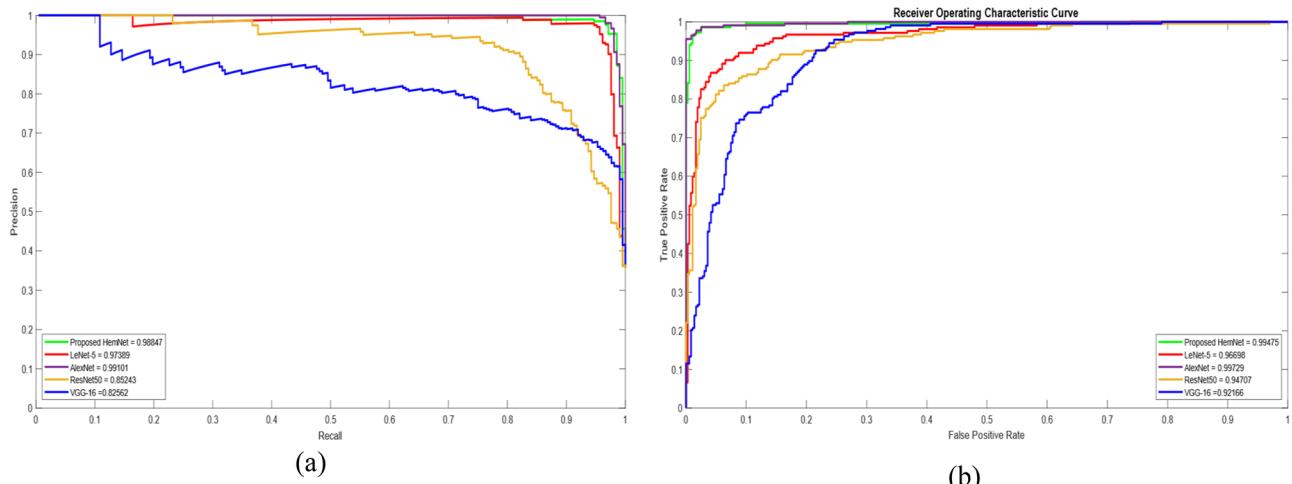
Methods	SE (%)	SP (%)	AC (%)	P (%)
VGG-16	95.88	94.87	94.93	54.69
LeNet-5	89.35	96.69	96.13	69.42
ResNet50	92.24	97.81	97.46	73.86
AlexNet	92.21	98.24	97.88	76.90
Proposed HemNet	94.21	97.46	97.19	76.87

Table 2. Classification of HEs using various models on the DIARETDB1 Dataset.**Figure 5.** Comparison of deep models with proposed HemNet Network on DIARETDB1 Dataset.

and recall over threshold values ranging from 0.0 to 1.0. Figure 6a shows the responses of LeNet-5, AlexNet, ResNet50, VGG-16, and proposed HemNet on the DIARETDB1 dataset. The AUC of the proposed HemNet was 0.988, closest to the optimum AUC of 0.991 provided by AlexNet. The plot shows the worst performance of VGG-16 with AUC of 0.826. Similarly, AlexNet and HemNet yielded the highest AUC in the receiver operating characteristic (ROC) curves of deep models can be depicted in Fig. 6b.

Discussion

The outcomes of the proposed HemNet model are also encouraging when independently benchmarked on the DIARETDB0 dataset. Effectively, the HemNet marked the highest SE of 90.98% among all the deep networks. However, its SP was less than AlexNet and ResNet50 but higher than VGG-16 and LeNet-5. AC of HemNet was 97.12%, which was also the highest and closer to AlexNet's 97.08%. Conversely, ResNet50 and AlexNet outclassed the proposed network by scoring the P of 87.25% and 89.58%, respectively. While HemNet marked 86.43% P,

**Figure 6.** Comparison of Deep Models on the DIARETDB1 Dataset (a) PR curves (b) ROC curves.

which was greater than LeNet-5 and VGG-16. The classification results of various deep learning models on DIARETDB0 are statistically compared in Table 3 and pictorially depicted in Fig. 7.

The assessment of training time with classification accuracy is problematic because it is laborious due to complex HEs detection applications and cannot be justified. Therefore, this criterion was analyzed using synthetic images by freezing parameters like learning rate and epochs in training. Table 4 explains the results of various deep networks on the concrete crack dataset. Ten epochs were fixed so that the convergence rate could also be assessed. It can be seen from the statistics, the propounded network's training time was the lowest, which is 11.19 min, but validation and testing accuracies were lower than AlexNet, ResNet50, and VGG-16. The convergence rate of VGG-16 was the highest as it scored 99.70% and 99.50% validation and testing accuracies within the ten epochs. VGG-16 outperformed all other deep models in terms of validation and testing accuracies. However, it took 16.56 min which was more than the 11.19 min of HemNet. Although, the differences in accuracies were not substantial because they could be increased by slightly increasing the epochs.

Table 5 elaborates the performances of deep models using the MNIST dataset. The training time of propounded HemNet was 2.33 min which was the minimum, and it scored 99.70 validation and testing accuracies. VGG-16 marked 100% validation accuracy and 99.90 testing accuracy, slightly higher than HemNet. However, its training time was 6.59 min, more than 2.33 of HemNet. The performance metrics of deep models can be observed in Table 5.

Methods	SE (%)	SP (%)	AC (%)	P (%)
VGG-16	89.10	96.95	96.01	80.07
LeNet-5	83.72	97.18	95.57	80.15
ResNet50	86.41	98.27	96.84	87.25
AlexNet	84.17	98.74	97.08	89.58
Proposed HemNet	90.98	97.98	97.12	86.43

Table 3. Classification of HEs using various models on the DIARETDB0 dataset.

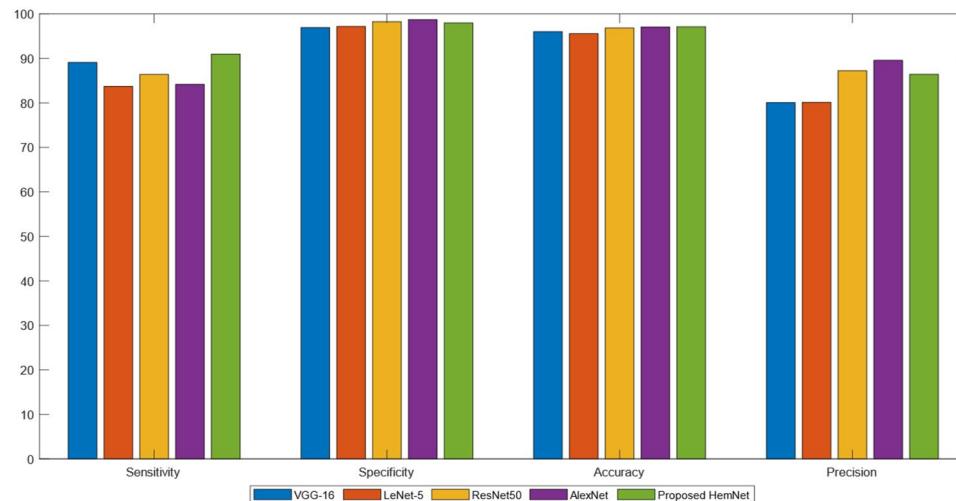


Figure 7. Comparison of Deep Models with Proposed HemNet Network on DIARETDB0 Dataset.

Method	Time (min)	Validation accuracy (%)	Testing accuracy (%)
LeNet-5	32.50	86.04	86.84
AlexNet	22.15	98.76	98.44
ResNet50	30.57	98.56	98.34
VGG-16	16.56	99.70	99.50
Proposed HemNet	11.19	96.28	95.92

Table 4. Analysis of deep models using time versus classification accuracy on concrete crack dataset.

Method	Time (min)	Validation accuracy (%)	Testing accuracy (%)
LeNet-5	2.53	92.80	91.90
AlexNet	3.11	99.10	98.60
ResNet50	9.52	99.70	99.60
VGG-16	6.59	100	99.90
Proposed HemNet	2.33	99.70	99.70

Table 5. Analysis of deep models using time versus classification accuracy on MNIST dataset.

Conclusion

The statistics of health organizations from various regions of the world indicate the proliferation of Diabetic retinopathy patients in the future. The insufficient medical resources and time-consuming treatment modalities would be unable to manage the outbreak. Furthermore, the ophthalmologist's involvement in the screening process and manual inference of the pathology causes adverse effects on the eye due to human error. It demands computer-based algorithms to expedite the screening process and to prevent the side effects of human interpretability. Additionally, these methods can assist in automating the diagnosis process, making it more efficient and less reliant on human interpretation.

This research has demonstrated automatic hemorrhage detection for screening Diabetic retinopathy using a novel hemorrhage network. The detection process is intelligent because it first estimates the prospective hemorrhage's locations which are imperative to infer the severity level of the ailment. The estimation process generates the data that suffices to deal with the limited data for training a deep model. The propounded network provides promising results while reducing training time significantly. A very deep network may not produce good results for some applications, as the experiment suggests that AlexNet and the proposed network are shallow but provide the highest results. Their overall performance is the best among all comparing networks. VGG-16 scored the best results for simple concrete crack and MNIST datasets. Its convergence rate is the highest. Conversely, its results are worse when applied to complex HEs classification problems. The reason might be the oscillation around the optimum solution due to the excess convergence rate. Therefore, it can be deduced that increasing the network's layer may not guarantee good results rather than increasing the training time. The arrangement of deep layers and the appropriate selection of parameters are crucial for enhancing the network's metrics.

The rigorous assessment of the propounded method reveals that the proposed detection scheme depends on seed point extraction. The constituent hemorrhage cannot be detected if a seed point is eliminated during the extraction phase. The parameters of the gaussian matched filter are empirically selected and should not be greater than the cross-section of the main artery. However, manual parameter selection may have adverse effects on the detection process. Therefore, the cross-section of the main artery could be automatically estimated for the robust selection of matched filter parameters and is proposed for future considerations. The feed-forward strategy in the architecture of a deep model might be effective for many applications, as represented by HemNet. Further experimentation needs to be conducted to evaluate the effects of feed-forward on deep networks. However, encouraging evaluation metrics of HemNet have been presented that show the efficacy of HemNet for hemorrhage detection. It is intended to conduct extensive experimentation for assessing this concept in the future.

Data availability

Publicly available datasets were analyzed in this study. The datasets can be found here: <https://www.it.lut.fi/project/imageret/diaretbdb1/index.html>, <https://www.it.lut.fi/project/imageret/diaretbdb0/index.html>.

Received: 8 November 2022; Accepted: 23 January 2023

Published online: 27 January 2023

References

1. Teo, Z. L. *et al.* Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology* **128**, 1580–1591 (2021).
2. Sharma, S., Hashmi, M. F., Kumar, A. Intracranial Hypertension. [Updated 2021 Aug 14]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing (2021) <https://www.ncbi.nlm.nih.gov/books/NBK507811/>.
3. Duh, E. J., Sun, J. K. & Stitt, A. W. Diabetic retinopathy: Current understanding, mechanisms, and treatment strategies. *JCI Insight*. **2**(14), e93751 (2017).
4. Gimbrone, M. A. Jr. & García-Cardeña, G. Endothelial cell dysfunction and the pathobiology of atherosclerosis. *Circ. Res.* **118**(4), 620–636 (2016).
5. Hattenbach, L. O. (ed.) *Management of Macular Hemorrhage* (Springer, 2018).
6. Herbort, C. P. *et al.* Advances and potential new developments in imaging techniques for posterior uveitis Part 2: Invasive imaging methods. *Eye* **35**(1), 52–73 (2021).
7. Alghadyan, A. A. Diabetic retinopathy—an update. *Saudi J. Ophthalmol. Off. J. Saudi Ophthalmol. Soc.* **25**(2), 99–111. <https://doi.org/10.1016/j.sjopt.2011.01.009> (2011).
8. Kamran, S. A., Hossain, K. F., Tavakkoli, A., Zuckerbrod, S., Baker, S. A., & Sanders, K. M. Fundus2Angio: A conditional GAN architecture for generating fluorescein angiography images from retinal fundus photography. In *International Symposium on Visual Computing* (pp. 125–138) (Springer, Cham, 2020)
9. Citirik, M., Tekin, K. & Teke, M. Y. Terson syndrome with persistent vitreous hemorrhage following traumatic brain injury. *Saudi J. Ophthalmol.* **33**(4), 392–397 (2019).
10. Li, T. *et al.* Applications of deep learning in fundus images: A review. *Med. Image Anal.* **69**, 101971 (2021).

11. Khojasteh, P., Aliahmad, B. & Kumar, D. K. Fundus images analysis using deep features for detection of exudates, hemorrhages and microaneurysms. *BMC Ophthalmol.* **18**(1), 1–13 (2018).
12. Lahmiri, S. & Shmuel, A. Variational mode decomposition based approach for accurate classification of color fundus images with hemorrhages. *Opt. Laser Technol.* **96**, 243–248 (2017).
13. Orlando, J. I., Prokofyeva, E., Del Fresno, M. & Blaschko, M. B. An ensemble deep learning based approach for red lesion detection in fundus images. *Comput. Methods Programs Biomed.* **153**, 115–127 (2018).
14. Son, J. *et al.* Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. *Ophthalmology* **127**(1), 85–94 (2020).
15. Gayathri, S., Gopi, V. P. & Palanisamy, P. A lightweight CNN for Diabetic Retinopathy classification from fundus images. *Biomed. Signal Process. Control* **62**, 102115 (2020).
16. Hacisofaoglu, R. E., Karakaya, M. & Sallam, A. B. Deep learning frameworks for diabetic retinopathy detection with smartphone-based retinal imaging systems. *Pattern Recognit. Lett.* **135**, 409–417 (2020).
17. Qureshi, I., Ma, J. & Abbas, Q. Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning. *Multimed. Tools Appl.* **80**(8), 11691–11721 (2021).
18. Hemanth, D. J., Deperlioglu, O. & Kose, U. An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network. *Neural Comput. Appl.* **32**(3), 707–721 (2020).
19. Asiri, N., Hussain, M., Adel, F. A., & Aboalsamh, H. (2021). A Deep Learning-Based Unified Framework for Red Lesions Detection on Retinal Fundus Images. [arXiv:2109.05021](https://arxiv.org/abs/2109.05021).
20. Van Grinsven, M. J., van Ginneken, B., Hoyng, C. B., Theelen, T. & Sánchez, C. I. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE Trans. Med. Imag.* **35**(5), 1273–1284 (2016).
21. Kauppi, T., Kalesnykiene, V., Kamarainen, J. K., Lensu, L., Sorri, I., Raninen, A., Voutilainen, R., Uusitalo, H., Kälviäinen, H., Pietilä, J. The DIARETDB1diabetic retinopathy database and evaluation protocol. In Proceedings of the British Machine Vision Conference, Warwick, UK, 10–13 (2007).
22. Kauppi, T. *et al.* DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms. *Mach. Vis. Pattern Recognit. Res. Group Lappeenranta Univ. Technol. Lappeenranta Finland* **73**, 1–17 (2006).
23. Alwazzan, M. J., Ismael, M. A. & Ahmed, A. N. A hybrid algorithm to enhance colour retinal fundus images using a wiener filter and CLAHE. *J. Digit. Imag.* **34**, 750–759 (2021).
24. Veluchamy, M. & Subramani, B. Image contrast and color enhancement using adaptive gamma correction and histogram equalization. *Optik* **183**, 329–337 (2019).
25. Huang, Z., Zhang, T., Li, Q. & Fang, H. Adaptive gamma correction based on cumulative histogram for enhancing near-infrared images. *Infrared Phys. Technol.* **79**, 205–215 (2016).
26. Gui, Z., Liu, Y., Gui, Z. & Liu, Y. An image sharpening algorithm based on fuzzy logic. *Optik* **122**, 697–702 (2011).
27. Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M. & Goldbaum, M. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Trans. Med. Imag.* **8**(3), 263–269 (1989).
28. Nie, F., Gao, C., Guo, Y. & Gan, M. Two-dimensional minimum local cross-entropy thresholding based on co-occurrence matrix. *Comput. Electr. Eng.* **37**(5), 757–767 (2011).
29. Li, Z. H., Han, D., Yang, C. J., Zhang, T. Y. & Yu, H. Q. Probing operational conditions of mixing and oxygen deficiency using HSV color space. *J. Environ. Manag.* **232**, 985–992 (2019).
30. Villaseñor-Aguilar, M. J. *et al.* Current status of optical systems for measuring lycopene content in fruits. *Appl. Sci.* **11**(19), 9332 (2021).
31. Aziz, T., Illesanmi, A. E. & Charoenlarpnopparut, C. Efficient and accurate hemorrhages detection in retinal fundus images using smart window features. *Appl. Sci.* **11**(14), 6391 (2021).
32. Wan, S., Liang, Y. & Zhang, Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. *Comput. Electr. Eng.* **72**, 274–282 (2018).
33. Liznerski, P., Ruff, L., Vandermeulen, R. A., Franks, B. J., Kloft, M., & Müller, K. R. (2020). Explainable deep one-class classification. [arXiv:2007.01760](https://arxiv.org/abs/2007.01760)
34. <http://yann.lecun.com/exdb/mnist>
35. Ebrahimi, A., Luo, S., Alzheimer's Disease Neuroimaging Initiative. Convolutional neural networks for Alzheimer's disease detection on MRI images. *J. Med. Imag.* **8**(2), 024503 (2021).
36. Lu, S., Lu, Z. & Zhang, Y. D. Pathological brain detection based on AlexNet and transfer learning. *J. Comput. Sci.* **30**, 41–47 (2019).
37. Theckedath, D. & Sedamkar, R. R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput. Sci.* **1**(2), 1–7 (2020).
38. Dung, C. V., Sekiya, H., Hirano, S., Okatani, T. & Miki, C. A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks. *Autom. Constr.* **102**, 217–229 (2019).

Author contributions

T.A.: investigation, methodology, software implementation, writing draft, conceptualization, formal analysis, and results verification. C.C.: funding acquisition, project administration, supervision, results verification, editing and finalizing the manuscript. S.M.: resources, visualization, writing original draft, results verification, and methodology.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Article

A Survey on Deep-Learning-Based Diabetic Retinopathy Classification

Anila Sebastian * , Omar Elharrouss , Somaya Al-Maadeed  and Noor Almaadeed

Department of Computer Science and Engineering, Qatar University, Doha P.O. Box 2713, Qatar
* Correspondence: as2000657@student.qu.edu.qa

Abstract: The number of people who suffer from diabetes in the world has been considerably increasing recently. It affects people of all ages. People who have had diabetes for a long time are affected by a condition called Diabetic Retinopathy (DR), which damages the eyes. Automatic detection using new technologies for early detection can help avoid complications such as the loss of vision. Currently, with the development of Artificial Intelligence (AI) techniques, especially Deep Learning (DL), DL-based methods are widely preferred for developing DR detection systems. For this purpose, this study surveyed the existing literature on diabetic retinopathy diagnoses from fundus images using deep learning and provides a brief description of the current DL techniques that are used by researchers in this field. After that, this study lists some of the commonly used datasets. This is followed by a performance comparison of these reviewed methods with respect to some commonly used metrics in computer vision tasks.

Keywords: diabetic retinopathy grading; diabetic retinopathy detection; deep learning; convolutional neural network; retinal fundus images

1. Introduction



Citation: Sebastian, A.; Elharrouss, O.; Al-Maadeed, S.; Almaadeed, N. A Survey on Deep-Learning-Based Diabetic Retinopathy Classification. *Diagnostics* **2023**, *13*, 345. <https://doi.org/10.3390/diagnostics13030345>

Academic Editor: Jae-Ho Han

Received: 17 November 2022

Revised: 21 December 2022

Accepted: 22 December 2022

Published: 18 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

During the past two decades, the number of people affected by diabetes has increased alarmingly. According to the IDF Diabetes Atlas [1], almost half a billion people of all ages have been diagnosed with it across the globe. This is expected to reach seven-hundred million by 2045. It is a global health concern. The IDF Diabetes Atlas also warns that, by 2040, one in three diabetes patients will develop Diabetic Retinopathy (DR). DR is a condition that can be identified by the presence of injured blood vessels behind the retina. This may result in serious complications such as the loss of vision when it goes undetected for a long time, hence the importance of addressing this issue. At present, doctors manually examine the fundus images of the eye to assess the severity of DR. This consumes much time, and there is a shortage of available medical professionals with respect to the actual number of patients. Due to these reasons, many patients do not receive medical care in a timely manner. Even though patients suffering from diabetes are advised by physicians to receive regular medical screenings of their fundus, many cases are left undetected until the disease becomes severe [2]. Hence, it is desirable to have an automated system to help in the detection of diabetic retinopathy.

Most studies in this field use fundus images, which provide visual records that document the present ophthalmic appearance of a person's retina. The presence of DR symptoms in these fundus images can be used to classify it using several steps such as retinal blood vessel segmentation, lesion segmentation, and DR detection [3]. The detection of DR and its current stage can be determined by investigating the presence/absence of several lesions. Some of the lesions are microaneurysms (MAs), superficial retinal hemorrhages (SRHs), exudates (Exs)—both soft exudates (SEs) and hard exudates (HEs)—intraretinal hemorrhages (IHEs), and cotton wool spots (CWSs). Figure 1 shows a comparison between a healthy retina and an unhealthy retina.

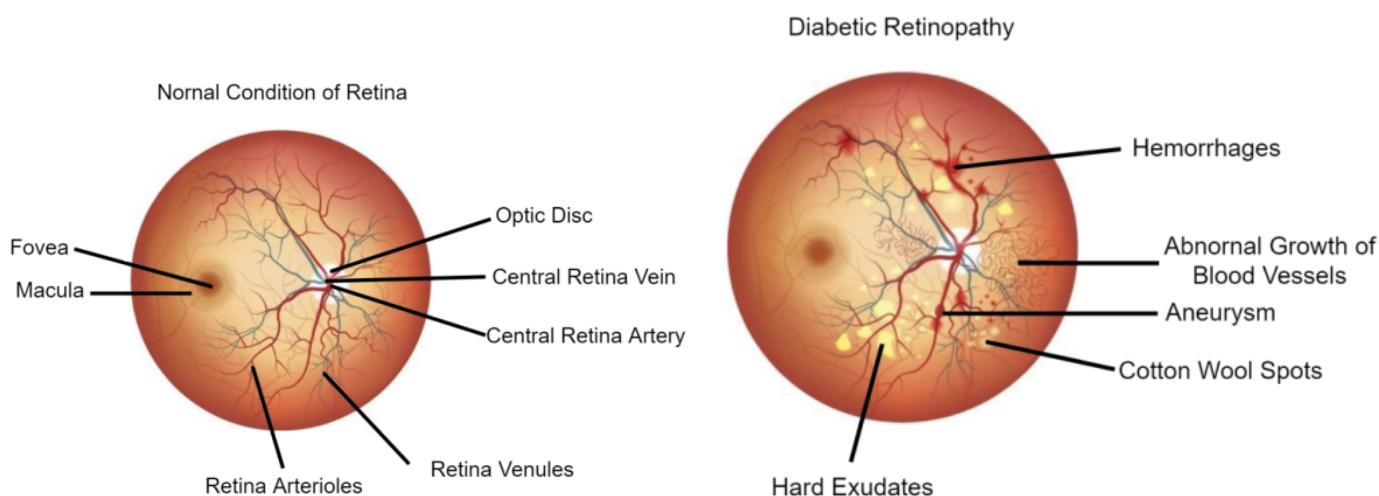


Figure 1. Visualization of a healthy retina and an unhealthy retina (<https://neoretina.com/blog/diabetic-retinopathy-can-it-be-reversed/>, accessed on 1 August 2022).

With the development of AI techniques, including machine learning and deep learning, high-performance detection and grading of the retina to detect and segment the infected parts of the retina become possible. Machine learning approaches are widely used for DR classification and grading. Nazir et al. [4] used a new way to represent fundus images called the “tetragonal local octa pattern (T-LOP) features”. Later, this classification was performed using extreme learning machine. Three ML classifiers—support vector machine (SVM), random forest, and J48—were used by the authors in [5]. The Gabor wavelet method followed by the AdaBoost classifier were used by the authors in [6] to grade DR. Recently, many deep learning techniques have been utilized by researchers to perform these tasks. This study provides a review of the present literature in this area with a focus on how DL is being used for DR detection and grading from fundus images. DL is a branch of AI that makes use of artificial neural networks with multiple processing layers to gradually extract the high-level features from the data. In this paper, we also summarize the DL architectures that have been used by the different reviewed studies.

However, significant research in this field using DL is also being carried out using optical coherence tomography (OCT) images, which have a higher resolution [7–9]. OCT images are more suitable than fundus images for developing systems that require micrometer resolution and a penetration depth of millimeters, which is why they are used by researchers for DR diagnosis, especially at the early stages [7].

The paper is organized as follows. The related works on DR detection and DR grading are presented in Section 2. Section 3 describes some of the preprocessing techniques that are used. Section 4 describes the datasets used. A comparison and discussion of the experiments are provided in Section 5. Some of future directions are provided in Section 6. The conclusion is presented in Section 7.

2. Literature Review

The diagnosis of diabetic retinopathy can be performed using two techniques: detection and grading. The detection of diabetic retinopathy is performed using binary classification (DR or normal retina), while diabetic retinopathy grading consists of detecting and annotating the infected parts, including the types of infection: mild, moderate, or severe. Figure 2 summarizes these two different types of DR studies. This section describes these studies by categorizing them into diabetic-retinopathy-detection-based studies and DR-grading-based studies. All these studies are summarized in Table 1.

2.1. DR-Detection-Based Studies

The diabetic retinopathy detection studies perform binary classification of the input images as healthy or DR. Here, we focus on deep-learning-based methods, which are the most effective approaches compared with other machine-learning-based or traditional techniques. For example, Kazakh-British et al. [10] proposed a simple convolutional neural network (CNN) to automatically classify DR. They used the original images and images filtered using an anisotropic diffusion filter in the experiments. From the obtained results, the authors found that the use of the anisotropic diffusion filter improved the performance. In the same context, the authors in [11–14] used CNN architectures to perform binary classification to identify the presence of diabetic retinopathy. After applying the Wiener filter to the fundus images and using OTSU for the segmentation, the authors of [15] proposed a deep CNN for multi-class classification of the fundus images into those having several vision-threatening diseases such as DR and the normal fundus images.

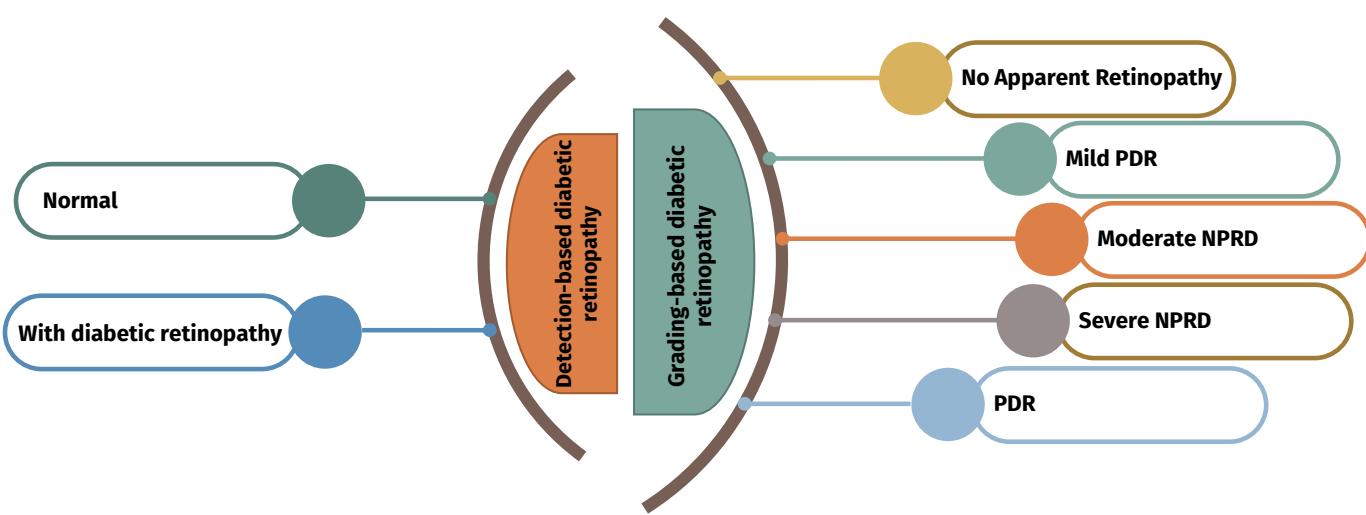


Figure 2. Types of diabetic retinopathy studies.

Instead of using simple convolutional neural networks, some authors have used pre-trained models (backbones) for transfer learning or for feature extraction to implement their methods. These are shown in Figure 3. For example, InceptionV3 was used by the authors in [16] to classify DR on RGB and textures features. Umapathy et al. [17] used a pre-trained InceptionV3 to perform DR classification. A binary CNN (BCNN) was proposed by the authors in [18] for DR classification to reduce memory consumption and improve runtime. Both binomial classification and multinomial classification of fundus images were performed by the authors in [19] using the MobileNetV2 architecture since this architecture requires less training time and can be used in mobile systems. Saranya et al. [20] used the DenseNet-121 model to detect DR from fundus images, while transfer learning using EfficientNet-B0, EfficientNet-B4, and EfficientNet-B7 were exploited to detect DR in [21]. The same backbones were used in [22] to classify DR into referable/vision-threatening DR. The EfficientNet-B3 backbone initialized with ImageNet weights and fully connected layers initialized with HE initialization were used for training by the author in [23]. From the experiments, the EfficientNet model gave good results compared to the ground-truth.

Another Backbone was used by Sudarmadji et al. [24] for diabetic retinopathy detection. The proposed method used the VGG network for feature extraction to implement the proposed CNN-based model. Boral and Thorat [25] used a transfer learning approach using InceptionV3 followed by SVM to perform DR classification. In another paper, five transfer learning models, Xception, InceptionResNetV2, MobileNetV2, DenseNet-121, and NASNetMobile, were used by the authors in [26] to perform binary classification of DR. DenseNet-121 was used as the transfer-learning-based method by the authors in [27] to

identify MAs, Exs, and hemorrhages from the input images to detect DR. Furthermore, transfer learning, VGG, AlexNet, Inception, GoogleNet, DenseNet, and ResNet were used by the authors in [28]. Another study [29] involved a comparison of three types of deep-learning-based architectures including Transformer-based networks, CNNs, and multi-layered perceptrons (MLPs) for DR classification. Different models included in the study were EfficientNet, ResNet, Swin-Transformer, Vision-Transformer (ViT), and MLP-Mixer. The models that are based on the transformer architecture were found to have the best accuracy among these. An ensemble model consisting of three CNN models was used by the authors in [30] for DR classification. It was based on stack generalization. ResNet-50 and VGG-16 were also used. Four vital features of using the CNN for DR classification, different architectures of the CNN, preprocessing techniques, class imbalance, and fine-tuning were evaluated by the authors in [31]. AlexNet, ResNet-50, and VGG-16 were employed for this purpose. The performances of twenty-eight deep hybrid architectures for binary classification of DR into referable DR and non-referable DR were empirically evaluated by the authors of [32]. This was compared with end-to-end deep learning (DL) architectures. A hybrid architecture using the SVM classifier and MobileNetV2 for feature extraction was found to be the best-performing among these. A three-class classification of fundus images into normal, glaucomatous, and diabetic retinopathy eyes was performed by the authors in [33]. Multiple CNN models—MobileNetV2, DenseNet-121, InceptionV3, InceptionResNetV2, ResNet-50, and VGG-16—were used for DR classification.

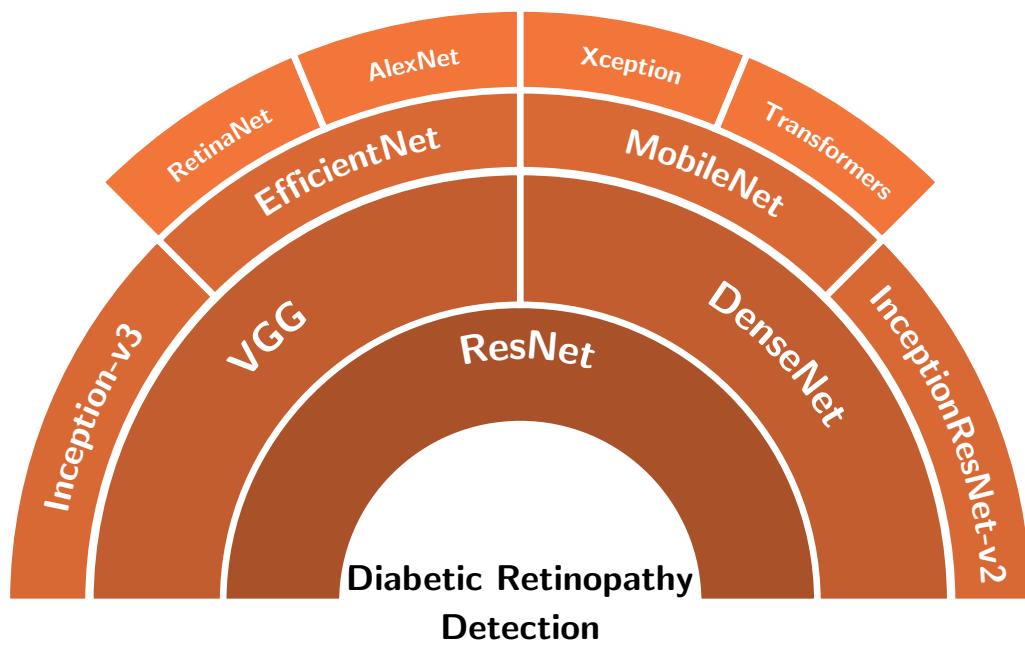


Figure 3. Backbones used for diabetic retinopathy detection studies.

A model based on ResNet with gradient-weighted class activation mapping (Grad-CAM) was used by the authors of [34] for lesion detection and DR classification. The lesions included MAs, HEs, hemorrhages, and CWSs. Quellec et al. [35] found that, when training for image-level classification was used with ConvNet, it became capable of performing lesion detection. The training was performed with a simplification of the back-propagation method. The images were classified into non-referable DR and referable DR. A new neural network called the lesion-guided network (LGN) was proposed by Tang et al. [36] to diagnose DR. For lesion detection, the backbone was RetinaNet with ResNet-50. A lesion-aware module (LAM) was also used to improve the rough lesion maps. Enhanced DR detection was performed by using the Harris hawks optimization (HHO) algorithm along with a DCNN by the authors of [37]. Gunasekaran et al. [38] used a deep RNN (DRNN) to perform early detection of DR. A CNN-based method was proposed in [39] to detect DR. A

very recent work [40] used seven different CNNs for DR diagnosis. Experiments in this study included single-modality and joint fusion strategies.

2.2. DR-Grading-Based Studies

As per the International Clinical Diabetic Retinopathy (ICDR) [41] scale, diabetic retinopathy can be graded into separate grades: no apparent retinopathy, mild non-proliferative diabetic retinopathy (NPDR), moderate NPDR, severe NPDR, and proliferative diabetic retinopathy (PDR). An example of each grade is presented in Figure 4. Many studies have been proposed for multi-class classification and grading of fundus images into the above-mentioned five stages.

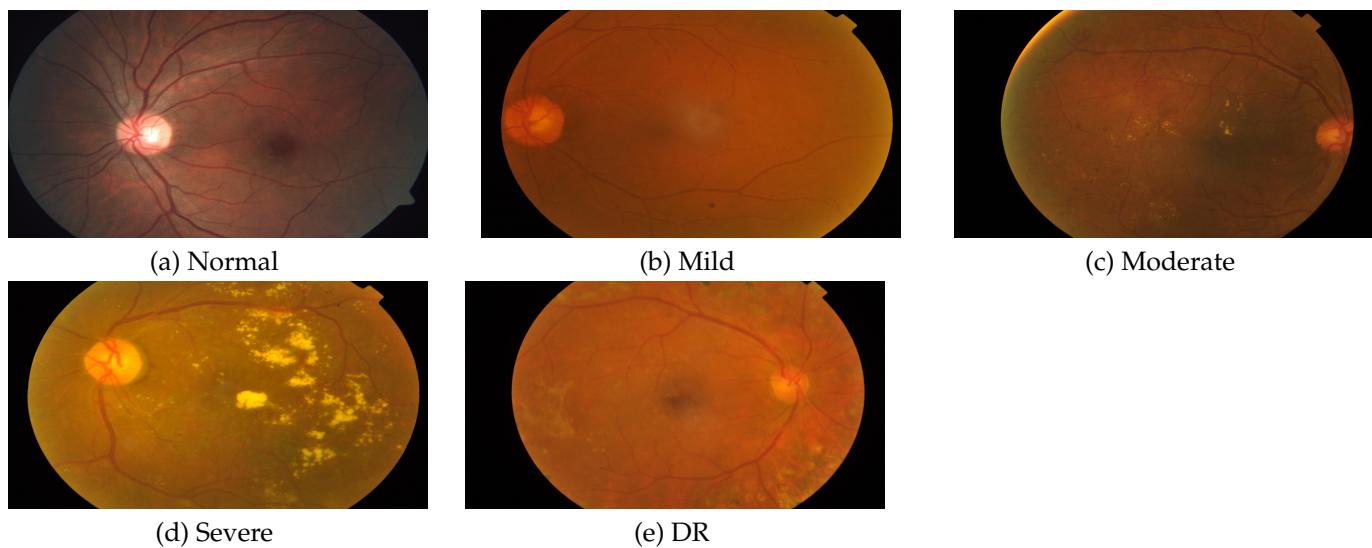


Figure 4. The five types of diabetic retinopathy.

A simple CNN model was used by the authors in [42] after applying a green channel filter to assess the stage of DR from fundus images. A CNN, which combined multi-view fundus images, was used along with attention mechanisms by the authors in [43]. It was called MVDRNet and used VGG-16 as the basic network. A locally collected dataset containing multi-view fundus images was employed for this. Another study that used a locally collected dataset from the University Hospital Saint Joan, Tarragona, Spain, is [44]. The CNN model used had batch normalization followed by the ReLU function. This was followed by a linear classifier and a softmax function. Two datasets—a balanced dataset with no augmentation and another one with augmentation—were used by the authors in [45]. A CNN was used to demonstrate the improvement in accuracy in DR grading due to the augmentation. Agustin and Sunyoto [46] performed a comparison of different regularization methods regarding how they reduce the overfitting of CNNs when used for DR severity grading. Dropout regularization was found to reduce overfitting and to increase accuracy.

Table 1. Retinopathy-grading-based studies during the period 2017–2020.

Method	Year	Method	Dataset(s)
Li et al. [47]	2017	CNN-based transfer learning, SVM	DR1 and MESSIDOR
Ardiyanto et al. [48]	2017	Deep-DR-Net	FINDeRS
Kwasigroch et al. [49]	2018	Transfer learning and VGG	Kaggle EyePACS
Wang et al. [50]	2018	AlexNet, VGG-16, and InceptionV3	Kaggle EyePACS
Zhou et al. [51]	2018	Inception-ResNet-v2, BaseNet	Kaggle EyePACS
Shrivastava and Joshi [52]	2018	InceptionV3, SVM	Kaggle EyePACS
Arora and Pandey [53]	2019	AlexNet, VGG-16, and InceptionV3	Kaggle EyePACS
Kassani et al. [54]	2019	InceptionV3, MobileNet, and ResNet-50	Kaggle APTOS
Hathwar and Srinivasa [55]	2019	Inception-ResNet-V2, and Xception	Kaggle EyePACS, IDRiD
Bellemo et al. [56]	2019	Ensemble of Adapted VGG and ResNet	Kitwe Central Hospital, Zambia
Kumar [57]	2019	Ensemble of GoogleNet, AlexNet, and ResNet-50	Kaggle EyePACS
Thota and Reddy [58]	2020	Pre-trained VGG-16	Kaggle EyePACS
Nguyen et al. [59]	2020	VGG-16 and VGG-19	Kaggle EyePACS
Lavanya et al. [60]	2020	ImageNet	Kaggle DR
Elzennary et al. [61]	2020	DenseNet-121	Kaggle APTOS
Barhate et al. [62]	2020	Autoencoder and VGG	Kaggle EyePACs
Wang et al. [63]	2020	Multichannel-based semisupervised GAN	MESSIDOR
Khaled et al. [64]	2020	VGG-16	Kaggle EyePACS
Islam et al. [65]	2020	Transfer Learning and VGG-16	Kaggle APTOS
Wang et al. [66]	2020	Hierarchical multi-task deep learning framework	Shenzhen, Guangdong, China
AbdelMaksoud et al. [67]	2020	E-DenseNet	Kaggle EyePACS, Kaggle APTOS
Yaqoob et al. [68]	2020	ResNet-50	MESSIDOR-2 and Kaggle EyePACS
Taufiqurrahman et al. [69]	2020	MobileNetV2, SVM	Kaggle APTOS
Vaishnavi et al. [70]	2020	AlexNet	Kaggle EyePACS
Shankar et al. [71]	2020	Synergic deep Learning (SDL) model	MESSIDOR
Karki and Kulkarni [72]	2021	EfficientNet	Kaggle APTOS
Qian et al. [73]	2021	Res2Net and DenseNet	Kaggle EyePACS
Shorfuzzaman et al. [74]	2021	CNN	Kaggle APTOS, MESSIDOR, IDRiD
Sugeno et al [75]	2021	EfficientNet-B3	Kaggle APTOS, DIARETDB1
Lee and Ke [76]	2021	VGG-16 and ResNet-50	IDRiD
Nazir et al. [77]	2021	DenseNet-100, CenterNet	Kaggle APTOS, IDRiD
Xiao et al. [78]	2021	SE-MIDNet	Kaggle EyePACS
Li et al. [79]	2021	SAGN, GCNN	Kaggle APTOS, Kaggle EyePACS
Martinez-Murcia et al. [80]	2021	ResNet-18 and ResNet-50	MESSIDOR
Rajkumar et al. [81]	2021	ResNet-50	Kaggle EyePACS
Swedhaasri et al. [82]	2021	SE-ResNet-50, EfficientNet	Kaggle APTOS
Reguant et al. [83]	2021	InceptionV3, ResNet50, and Xception	Kaggle EyePACS, DIARETDB1
Hari et al. [84]	2021	Xception, InceptionV3, and DenseNet-169	Kaggle EyePACS
Saeed et al. [85]	2021	VGG-19, ResNet, and DPN107	Kaggle EyePACS, MESSIDOR
Jabbar et al. [86]	2022	VGG	Kaggle EyePACS
Shaik and Cherukuri [87]	2022	HA-Net	Kaggle APTOS, IDRiD
Chandrasekaran and Loganathan [88]	2022	ResNet and AlexNet	Kaggle EyePACS
Oulhadj et al. [89]	2022	DenseNet, InceptionV3, and ResNet-50	Kaggle APTOS
Nair et al. [90]	2022	VGG-16, ResNet-50, and EfficientNet-B5	Kaggle APTOS
Deepa et al. [91]	2022	Xception, InceptionV3, and ResNet-50	Kaggle DR, DIARETDB, STARE
Farag et al. [92]	2022	DenseNet-169 with CBAM	Kaggle APTOS

Table 1. Cont.

Method	Year	Method	Dataset(s)
Canayaz [93]	2022	EfficientNet-B0, DenseNet-121	Kaggle APTOS
Bilal et al. [94]	2022	U-Net	Kaggle EyePACS, MESSIDOR-2
Murugappan et al. [95]	2022	DRNet	Kaggle APTOS
Chen and Chang [96]	2022	InceptionV3 and EfficientNet	Kaggle APTOS
Butt et al. [97]	2022	ResNet-18 and GoogleNet	Kaggle APTOS
Elwin et al. [98]	2022	DCNN, ShCNN	IDRiD, DDR
Deepa et al. [99]	2022	DCNN	Kaggle EyePACS, DIARETDB, STARE

A deep CNN model called DR|GRADUATE was presented by the authors in [100]. It was a new DL approach for DR grading, which could give a pathologically explainable description to support its judgment. It also provided an assessment of the ambiguity of its prediction. Feature extraction using a multipath CNN was used by the authors in [5]. After this, DR was graded using three different ML classifiers, SVM, random forest, and J48. Sugeno et al. [75] used the EfficientNet model to grade DR after using morphological operations and image processing for lesion detection. A multi-task model with EfficientNet-B5 was used by the authors of [101] for DR grading. Feature extraction performed with the EfficientNet backbone was fed to the dropout layer, which was followed by an ordinal regression section and a classification section. Shankar et al. [71] proposed a deep CNN model called the synergic deep learning (SDL) model to grade DR. Histogram-based segmentation was performed before this.

A pre-trained VGG-16 was used by the authors of [58] to train their proposed CNN to improve the accuracy of DR grading. VGG-16 and VGG-19 were used by the authors of [59] to grade DR. They mirrored and rotated the images to augment the dataset. The VGG-16 and ResNet-50 models were modified and used by the authors in [76] to grade DR with the help of the dropout concept. A cascaded model consisting of two VGG-16 models was used by the authors of [64]. The first model outputs “yes” or “no” to detect DR, and the second model classifies the fundus images into four different DR stages. Shaik and Cherukuri [87] used a model named “Hinge Attention Network (HA-Net)” which has multiple attention stages for DR severity grading. Initial spatial representations from the input images were extracted using a pre-trained VGG-16 base.

An automated DR detection system using a Raspberry Pi was developed by the authors of [60]. They used ImageNet for DR grading. Elzennary et al. [61] used the DenseNet-121 neural network architecture with the aid of transfer learning to determine the severity of DR. Both of these studies used the Python framework called Flask to create interfaces that can be used by doctors to detect DR. A custom CenterNet with DenseNet-100 support was used by the authors of [77] to detect eye diseases from retinal images. This study graded the severity of DR by separating the fundus images according to the lesions present.

Another classification network for DR-SE-MIDNet was introduced by the authors of [78]. It was built using an enhanced Inception module along with the squeeze-and-excitation (SE) module for grading. With the SE module, global information for the feature map on each channel was found. Feature extraction using InceptionV3 was performed using a hierarchical approach by the authors in [52]. The first layer was for binary classification into DR/no DR. The next one was to grade DR into the five DR stages. SVM with the radial basis function (RBF) kernel was utilized for classification. Wang et al. [63] used a multichannel-based semi-supervised GAN (SSGAN) for DR grading, which was capable of using labeled and unlabeled data as the training data. They used feature extraction to reduce the noise of the input images and for extracting the features of lesions. They also graded the lesions into three levels.

A new DL algorithm called Deep-DR-Net capable of being fit onto a small embedded board was introduced by the authors of [48] to grade DR. For this, they arranged a cascaded encoder–classifier network with a residual style to ensure that it was small in size. Li

et al. [79] proposed a semi-supervised auto-encoder graph network (SAGN) to diagnose DR. In this, an autoencoder was used for feature learning. After this, the RBF was used to calculate neighbor correlations. Finally, a graph CNN (GCNN) was used to grade DR. A graph neural network (GNN), which extracts lesion ROI sub-images to emphasize only lesions in fundus images, was proposed by Sakaguchi et al. [102]. A graph is constructed from these sub-images for DR classification.

Transfer learning and the VGG architecture were used by Kwasigroch et al. [49]. For this reason, the ImageNet dataset was used to pre-train the VGG architecture. Another DL model that used transfer learning—VGG-16—was used along with a new color version preprocessing method by Islam et al. [65] for DR grading. ResNet-18 and ResNet-50 were used along with residual transfer learning by Martinez-Murcia et al. [80] for the same. Another transfer learning approach—the ResNet-50 architecture trained on the ImageNet dataset—was used for DR classification and grading by the authors in [81]. Another study that used transfer learning by fine-tuning using a well-annotated ImageNet dataset to train Inception-ResNet-V2 and Xception models was given in [55]. The latter was found to have better performance. CNN-based transfer learning followed by SVM were used by the authors of [47]. AlexNet and VGG were pre-trained using the ImageNet dataset. Features extracted with the help of transfer learning were provided to SVM for DR grading. An ensemble model consisting of SE-ResNeXt50, EfficientNet-B4, and EfficientNet-B5 along with transfer learning was used by the authors of [82] for DR grading. The InceptionV3, ResNet-50, InceptionResNet50, and Xception models were used for DR grading by the authors in [83]. The parameters were initialized using transfer learning. They created visualization maps to investigate the clinical significance of the decisions made by the CNN models. Wang et al. [50] used AlexNet, VGG-16, and InceptionV3 along with transfer learning for DR grading. InceptionV3 was found to provide the best accuracy in their study. Jabbar et al. [86] used a transfer-learning-based VGG architecture for DR grading. Various data augmentation techniques were used to balance the classes in the training data.

Experiments using several deep neural networks (DNNs) were carried out to yield algorithms that grade DR conforming to the ICDR standards by the authors in [103]. The network was also trained to make several other binary classifications. Synchronized diagnosis of DR severity, DR features, and referable DR was conducted by the authors of [66]. A hierarchical multi-task DL framework with a skip connection was utilized for automatically merging the DR-related feature output with DR severity analysis. An ensemble of two CNN architectures—a modified VGG and RNN—was utilized for grading DR by the authors in [56]. Apart from the grading of DR as per the ICDR scale, the images were classified into referable DR/vision-threatening DR. Xception, InceptionV3, and DenseNet-169 were used by the authors of [84] for DR grading. They used the Kaggle DR dataset and created two versions of it: balanced and imbalanced. The Xception model, which was trained using the imbalanced version of the dataset, was found to have the best performance. VGG-19, ResNet-152, and DPN107 were used with two-stage transfer learning by the authors in [85] for grading DR. The initial layers of the pre-trained models were adjusted for the preceding layers to understand the lesions and also the normal areas. Zhou et al. [51] used a multi-cell architecture, which could increase the depth of the DNN, as well as the resolution of the input image. A three-layer architecture that used Inception-ResNet-v2 and BaseNet to grade DR was proposed. AlexNet, VGG-16, and InceptionV3 were used by the authors in [53] to determine DR stage classification. Image augmentation techniques were used before training. The DR grading performance of three models, a shallow CNN, ResNet with soft attention, and AlexNet for DR using a new hyper-analytic wavelet (HW) phase activation function, was compared by the authors in [88]. AlexNet for DR was found to show the maximum improvement in performance in their experiments. Oulhadj et al. [89] applied a deformable registration to the retina and graded DR using four CNN models, DenseNet-121, Xception, InceptionV3, and ResNet-50. Three pre-trained models, VGG-16, ResNet-50, and EfficientNet-B5, were used for DR grading by the authors in [90]. ResNet-50 was found to perform best among the three. The performance of three

pre-trained models, Xception, InceptionV3, and ResNet-50, in DR grading was compared by the authors of [91]. Their simulation result found the Xception model to perform better.

ResNet was used by the authors in [104] for feature extraction. After this, they used SVM, as well as a neural network (NN) pixelwise classifier to grade DR. AD2Net—a new CNN model having the qualities of Res2Net and DenseNet—was used by the authors of [73] for DR grading. An attention mechanism was used to make the network concentrate on understanding useful information from the images. A deep supervision of inception-residual network (DSIRNet) was used by [105], which was based on the network design ideas of GoogleNet and ResNet for feature extraction to grade DR. They also used a deep monitoring method to enhance the thermal classification effect of the training network. Yaqoob et al. [68] trained an optimized ResNet-50 having features from a canny edge detector and histogram of gradients to perform the grading of DR using two public datasets. An ensemble made of GoogleNet, AlexNet, and ResNet-50 was utilized by the authors of [57]. The images were preprocessed and fed to this ensemble model for DR grading. A CNN-based DL ensemble framework in which weights from distinct models were merged to make a solo model, which can extract prominent features from many lesions in the input images, was used by Shorfuzzaman et al. [74] to grade DR. Some CNN models that were pre-trained with the ImageNet dataset—the ResNet-50, DenseNet-121, Xception, and Inception models—were used for this. After preprocessing with CLAHE for segmentation, Vaishnavi et al. [70] used the AlexNet architecture for feature extraction. Finally, a softmax layer was utilized to grade the images according to DR severity.

An ensemble of five models from the EfficientNet family was used for DR grading by the authors in [72] by pre-training on ImageNet. These models were also used independently for the same, and EfficientNet-B3 performed better than the ensemble model and the other four models. A hybrid and effective model, MobileNetV2-SVM, was used by the authors of [69] to grade DR images. A stack of residual bottleneck layers, which consisted of a stack of bottleneck residual blocks, was used to construct the MobileNetV2 model. Jiang et al. [106] used three models—InceptionV3, ResNet-152, and Inception—ResNet-V2 to grade DR. An ensemble model consisting of these models, using the Adaboost algorithm, was also used. Another study used an embedded model consisting of five deep CNNs—ResNet-50, Xception, InceptionV3, DenseNet-121, and DenseNet-169 [107]. Stacked individual channels of the image were taken as the input. The forecast from separate models was averaged and used to fix the final target label. The green channel was found to give the best performance in grading DR. A novel hybrid DL model known as E-DenseNet was proposed by the authors of [67] to grade DR. It was a hybrid between a customized EyeNet and DenseNet based on DenseNet-121. The Xception deep feature extractor was used by the authors of [54] to advance the capability of the typical Xception architecture in classifying DR. They also used transfer learning along with hyper-parameter tuning.

A novel CNN model based on the DenseNet-169 architecture combined with a convolutional block attention module (CBAM) was used by the authors of [92] for DR severity classification. The ResNet-101 model was used for DR grading and to analyze the risk of macular edema by the authors in [108], and it was found to perform better than the ResNet-50 model. A heuristically constructed deep neural network was used by the authors of [109] to determine the severity levels of DR. An architecture consisting of an autoencoder along with a VGG network was used by the authors of [62] to reduce overfitting during DR detection. The network was pre-trained in a self-supervised manner.

The binary bat algorithm (BBA), equilibrium optimizer (EO), gravity search algorithm (GSA), and gray wolf optimizer (GWO) were used as the wrapper methods to select the best features that were obtained from the EfficientNet-B0 and DenseNet-121 models for DR grading by the authors in [93]. Transfer-learning-based InceptionV3 was used by the authors of [94] for DR grading. They used two separate U-Net models for OD and blood vessel segmentation. Five DL models—DenseNet-121, InceptionV3, ResNet-153, VGG-16, MobileNet, and InceptionResNet—were used with transfer learning for DR grading by the authors of [110]. Out of these, the VGG-16 model was found to provide the highest

accuracy in their experiments. Deepa et al. [99] used a pre-trained Xception model along with hierarchical clustering of image patches by the Siamese network to grade DR fundus images. A boosting-based ensemble learning method followed by a CNN was used by the authors of [111] for DR grading. A novel few-shot classification framework called DRNet was used by the authors of [95] for DR detection and grading. Episodic training was used to train the model on few-shot classification tasks. Both DR detection and DR grading were performed by the authors of [112] using a Bayesian neural network (BNN). Experiments using nine BNNs were performed to utilize their capability of uncertainty estimation in classifying DR. Chen and Chang [96] used the InceptionV3 and EfficientNet models to grade fundus images according to DR severity. A novel hybrid model called E-DenseNet was used by the authors of [113] for DR grading. It was a combination of the EyeNet and DenseNet models based on transfer learning. Another study by the authors of [97] used a similar hybrid model based on transfer learning for the detection and grading of DR. The model consisted of ResNet-18 and GoogleNet. Ar-HGSO, which is an autoregressive-Henry gas-sailfish-optimization-enabled deep learning model was used by the authors of [98]. The DCNN was used for DR detection, and the Shepard CNN (ShCNN) was used for severity classification. Rajavel et al. [114] introduced a cloud-enabled DR grading system that used an optimized deep belief network (O-DBN) classifier model. Dimensionality reduction and noise removal were performed by them using the stochastic neighbor embedding (SNE) feature extraction approach. LeNet-5 was used by the authors in [115] for DR grading. A spiking neural network (SNN) was used for DR grading by the authors in [116]. They used the chimp optimization algorithm with DenseNet (COA-DN) for feature extraction.

Table 1 summarize the studies that were presented in this section.

3. Preprocessing Techniques Used to Grade DR Fundus Images

Image enhancement is performed in most DR studies with the help of several preprocessing techniques. Preprocessing can consist of several steps such as image variation attenuation, intensity conversion, denoising, and contrast enhancement [117]. The attenuation of fundus images is required since there will be a wide variation in the color of the retina of different patients. Intensity conversion is used to make the features clearly visible in an image. Denoising of fundus images is required since much noise may be introduced into these images during the image acquisition process. Finally, contrast enhancement is essential since retinal images captured with the help of a fundus camera will have maximum contrast at the image center, which gradually reduces when moving away from the center. Other common preprocessing steps include image resizing and performing several image augmentations using techniques such as rotation, flipping, and zooming.

4. DR Datasets

The success of all these DL studies relies greatly upon the datasets that are used. The quality of the dataset used and the precision of the annotations will have a huge impact on the results that will be obtained by these methods. Hence, we created a list of some commonly used fundus image datasets for DR diagnosis. Table 2 presents this list.

A few of the commonly used publicly available datasets in these studies are STARE, IDRiD, MESSIDOR, DIARET DB1, the Kaggle APTOS dataset, and the Kaggle EyePACS dataset. Out of these, Kaggle's EyePACS and APTOS datasets are the most widely used datasets for DR detection/grading. However, these contain fundus images taken with different cameras and settings. The largest among these is the Kaggle EyePACS dataset with more than 88,000 fundus images, whereas some datasets, such as DIARETDB1, HRF, and DRiDB, have less than 100 fundus images.

Almost all of them are annotated for DR detection, while some datasets such as MESSIDOR and Kaggle EyePACS have been annotated also for DR grading. Most of the studies used different datasets/combinations of datasets for training and validation purposes since most of the datasets are small in size. However, some studies have used their own locally collected datasets for their experiments [43,44].

Table 2. Diabetic retinopathy datasets.

Dataset	No. of Images	Image Size
STARE	400	700 × 605
IDRID	516	4288 × 2848
MESSIDOR	1200	Different sizes
HRF	45	3504 × 2336
Kaggle EyePACS	88,702	Different sizes
Kaggle APTOS 2019	5590	Different sizes
MESSIDOR 2	1748	Different sizes
DDR	13,673	Different sizes

5. Discussion

In order to evaluate the diabetic retinopathy detection and grading methods on different datasets, a set of metrics is used, including model accuracy, sensitivity, specificity, and the AUC. These metrics are generally the most-used ones for detection and segmentation in computer vision tasks. In this section, we present the obtained results per dataset using the cited method for detection and grading methods. These results are reported in tables and figures in order to show the most-performed techniques using different architectures.

Tables 3 and 4 and Figures 5 and 6 show a comparison of the results obtained by some of the studies that have been reviewed. Studies that have used the same publicly available datasets have been grouped for comparison. Kaggle APTOS and Kaggle EyePACS are the largest datasets that have enabled these researchers to perform their experiments.

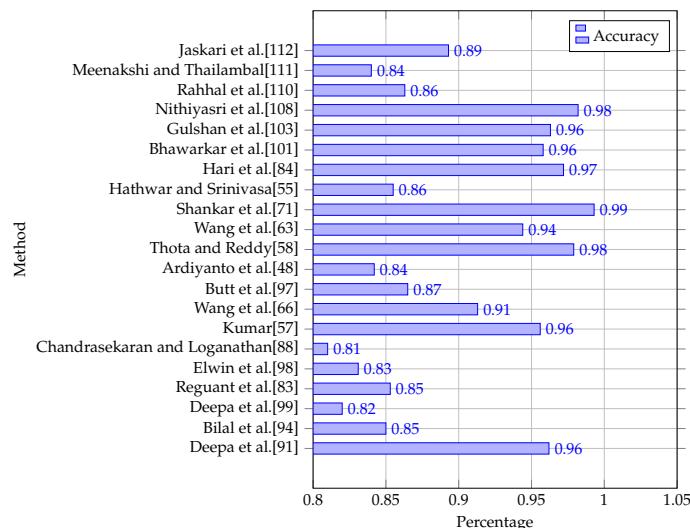


Figure 5. Performance comparison of diabetic-retinopathy-grading-based studies that used the Kaggle APTOS dataset.

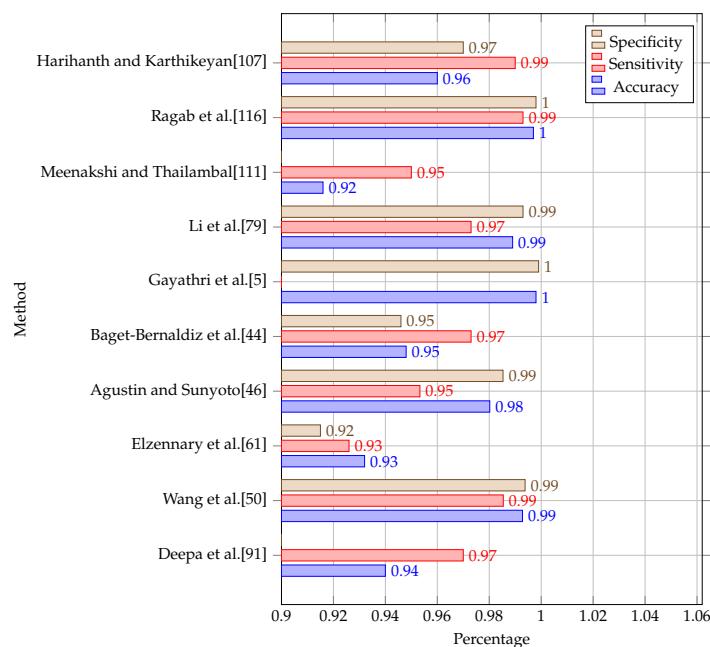


Figure 6. Performance comparison of diabetic-retinopathy-grading-based studies that used the MESSIDOR dataset.

5.1. Diabetic Retinopathy Detection

Diabetic retinopathy detection methods are performed on datasets of two classes that represent the images with diabetic retinopathy and the images without diabetic retinopathy. To show that, Table 3 compares some DR-detection-based studies. The most studies used Kaggle’s APTOS and EyePACS datasets, due to their size, which is large compared to the others. The binary classification to detect the fundus images that have DR lesions and, thus, detect the presence of DR is performed by the proposed methods. For that reason, we can see that all the methods can classify diabetic retinopathy with good performance in accuracy, while the sensitivity and specificity values were not mentioned in some of the studies. From the table of the obtained results using the proposed method on the Kaggle APTOS dataset, we can find that the authors in [11] achieved the best accuracy value of 94% with a difference of 4% better than the accuracy obtained using [40] and more than 8% for the other methods. Using the sensitivity and specificity metrics, the method in [22] achieved the best results. On the MESSIDOR and MESIDOR2 datasets, the methods used in [21,24] achieved the best accuracy, respectively. However, we can see that, for MESSIDOR2, the accuracies were lower than the obtained accuracies on MESSIDOR, due to the fact that the size of MESIDOR2 is larger than MESSIDOR, which can explain the difference between the accuracy on MESIDOR2 being 91% and 99% on MESSIDOR. The same observation is made for Kaggle EyePACS, which is a large-scale dataset; the accuracy performances were generally less than 91%, except for [18,24,25], which achieved an accuracy of up to 97%. For all the datasets including STARE, HRF, and IDRid, the performance of the proposed methods needs improvements due to the importance of the topic, as well as the impact of the error if these techniques are used in real-world diagnostics.

Table 3. Performance Comparison of diabetic-retinopathy-detection-based studies. The **bold** and underlined fonts, respectively, represent first and second place.

Dataset	Method	Accuracy	Sensitivity	Specificity
Kaggle APTOS	Anoop et al. [11]	0.946	0.860	<u>0.960</u>
	Pamadi et al. [19]	0.780	-	-
	Saranya et al. [20]	0.830	-	-
	Chetoui and Akhloufi [22]	-	0.991	0.972
	Sanjana et al. [26]	0.861	0.854	0.875
	Kumar and Karthikeyan [29]	0.864	-	-
	Lahmar and Idri [32]	0.890	-	-
MESSIDOR	El-Ateif and Idri [40]	0.907	0.928	0.893
	Rêgo et al. [16]	-	0.808	<u>0.973</u>
	Umapathy et al. [17]	0.944	-	-
	Sudarmadji et al. [24]	0.997	0.990	0.980
	Hossen et al. [27]	0.949	<u>0.926</u>	0.971
	Qomariah et al. [28]	0.958	-	-
	Mudaser et al. [21]	0.910	-	-
MESSIDOR2	Sanjana et al. [26]	<u>0.861</u>	0.854	<u>0.875</u>
	Lahmar and Idri [32]	0.841	-	-
	El-Ateif and Idri [40]	0.777	0.310	0.938
	Saranya et al. [20]	0.830	-	-
	Jiang et al. [14]	0.757	-	-
	Kaushik et al. [30]	0.979	-	-
	Boral and Thorat [25]	0.988	0.977	1.00
Kaggle EyePACS	Rêgo et al. [16]	-	0.808	0.973
	Kolla and Venugopal [18]	0.910	-	-
	Chetoui and Akhloufi [22]	-	0.981	<u>0.989</u>
	Sudarmadji et al. [24]	0.984	<u>0.980</u>	0.970
	Lian et al. [31]	0.790	-	-
	Lahmar and Idri [32]	0.840	-	-
	Quellec et al. [35]	0.954	-	-
STARE	Kazakh-British et al. [10]	<u>0.600</u>	-	-
	Umapathy et al. [17]	0.944	-	-
HRF	Chakrabarty [13]	1.00	1.00	-
	Umapathy et al. [17]	<u>0.944</u>	-	-
IDRid	Nasir et al. [12]	0.960	0.829	-

5.2. Diabetic Retinopathy Grading

Diabetic-retinopathy-grading-based studies comprise another classification category for diabetic retinopathy analysis. The proposed methods for diabetic retinopathy grading are based on deep learning using different CNN architectures. For that, transfer learning has been widely used in the reviewed studies. This is due to the effectiveness of the known backbones for the image classification tasks. This includes deep learning architectures/models such as encoder-decoder, VGG, DenseNet, Inception, Xception, EfficientNet, graph neural networks, etc. In addition, preprocessing techniques were also used in different studies to improve performance, as mentioned in Section 3. Grayscale conversion, resizing, CLAHE, and green channel extraction are some commonly preferred preprocessing techniques.

These techniques aid in improving the feature extraction process by removing unnecessary noise from the images.

In this section, we attempt to present the grading-based methods on popular DR datasets. The evaluation used a set of metrics including the accuracy, sensitivity, and specificity. Table 4 presents a comparison of the obtained results using the proposed method on studies that have used the Kaggle EyePACS, MESSIDOR2, DDR, and IDRid datasets. Figures 5 and 6 illustrate the experimental results using the proposed methods on the Kaggle APTOS and MESSIDOR datasets. From Table 4, we can find that the proposed methods succeeded in achieving high accuracies on MESSIDOR2, DDR, IDRid, reaching up to 97%. The same observation is made for the other metrics including the sensitivity and specificity. On Kaggle EyePACS, the proposed method in [5] achieved the best accuracy, as well as the best specificity metric value, while we can find that the majority of the methods achieved an accuracy of less than 90%. This is due to the complexity and size of the dataset. On Kaggle APTOS, from the obtained results represented in Figure 5, we can find that most methods that used accuracy as an evaluation metric achieved an accuracy of less than 97%, while only the method in [71] achieved an accuracy of 99%. For the MESSIDOR dataset, the proposed methods used the accuracy, sensitivity, and specificity metrics to evaluate their results. The obtained results are presented in Figure 6. It shows that many methods achieved an accuracy of up to 99% including [5,50,79,116], while the others achieved an accuracy of up to 92%.

From the presented results on different datasets, we can conclude that some of the methods such as [5] succeeded in classifying diabetic retinopathy with grading-based and detection-based methods with high accuracies, while some of the proposed methods were good for some datasets and less efficient for others. This makes diabetic retinopathy classification a challenging task even with the improvements achieved during the last ten years using different deep learning techniques.

Table 4. Performance comparison of diabetic-retinopathy-grading-based studies that used the Kaggle APTOS, MESSIDOR2, DDR, and IDRid datasets. The **bold** and underlined fonts, respectively, represent **first** and second place.

Dataset	Method	Accuracy	Sensitivity	Specificity
MESSIDOR 2	Yaqoob et al. [68]	0.970	-	-
	Bilal et al. [94]	<u>0.946</u>	0.948	0.944
DDR	Rahhal et al. [110]	1.00	-	-
	Elwin et al. [98]	<u>0.914</u>	0.925	0.905
IDRid	Shorfuzzaman et al. [74]	0.923	0.980	-
	Elsawah et al. [104]	0.866	-	-
	Sakaguchi et al. [102]	0.793	-	-
	Gayathri et al. [5]	0.990	-	0.997
	Lee and Ke [76]	0.972	0.702	0.921
	Nazir et al. [77]	<u>0.981</u>	-	-
	Shaik and Cherukuri [87]	0.664	-	-
	Nithiyasri et al. [108]	0.977	<u>0.978</u>	<u>0.989</u>
	AbdelMaksoud et al. [113]	0.930	0.967	0.720
	Elwin et al. [98]	0.914	0.925	0.905
	Sri et al. [115]	0.970	-	-
	Vaishnavi et al. [70]	0.958	0.920	0.978
	Thota and Reddy [58]	0.740	0.800	0.650
	Barhate et al. [62]	0.762	-	-
Kaggle EyePACS	Kwasigroch et al. [49]	0.508	-	-
	Wang et al. [50]	0.632	-	-
	Zhou et al. [51]	0.632	-	-
	Shrivastava and Joshi [52]	0.818	-	-
	Arora and Pandey [53]	0.744	-	-
	Kumar [57]	0.699	-	-
	Maistry et al. [45]	0.870	-	-
	Nguyen et al. [59]	0.820	0.800	0.820
	Khaled et al. [64]	0.631	-	-
	Harihanth and Karthikeyan [107]	0.819	-	-
	AbdelMaksoud et al. [113]	0.968	0.983	0.72
	Yaqoob et al. [68]	0.979	-	-
	Qian et al. [73]	0.832	-	-
	Gayathri et al. [5]	0.999	-	1.00
	Xiao et al. [78]	0.882	0.994	0.976
	Li et al. [79]	0.944	0.840	0.822
	Rajkumar et al. [81]	0.894	0.987	<u>0.999</u>
	Reguant et al. [83]	0.950	0.860	0.960
	Hari et al. [84]	0.830	-	-
	Saeed et al. [85]	<u>0.997</u>	0.960	0.998
	Jabbar et al. [86]	0.966	-	-
	Chandrasekaran and Loganathan [88]	0.980	<u>0.990</u>	-
	Bilal et al. [94]	0.979	0.969	0.969
	Deepa et al. [99]	0.960	-	-

6. Future Directions

Finally, we would like to provide some future research directions that were identified during this study. The latest trends such as using interpretable AI and cloud-enabled systems are also being used by some researchers in this field, as well as in medical imaging analysis [118–121]. Since interpretation will be preferred by doctors to diagnose DR, more studies on explainable AI may come up in the future such as those by Shorfuzzaman et al. [74] and Chetoui and Akhloufi [22]. Such DR-diagnosing systems will be able to help doctors rely on them with more confidence. The use of cloud-enabled systems for computer-aided DR detection systems such as the one by Rajavel et al. [114] will improve scalability. This will enable the development of large-scale systems for DR diagnosis.

Furthermore, developing low-cost standalone DR detection systems such as the one developed by the authors in [60] using a Raspberry Pi will be useful for deployment at health centers at a lower cost. Similar low-cost systems can also be created by developing DR diagnosis systems using smartphone-based retinal imaging systems such as the one by the authors in [122].

Another possible research direction is to develop more automated systems that are capable of determining more than one condition of the eyes, for example systems capable of diagnosing DR, as well as other conditions of the eyes such as glaucoma and diabetic macular edema, such as the one by the authors in [123].

7. Conclusions

In this work, we reviewed recent deep-learning-based approaches for diabetic retinopathy detection/diagnosis performed on fundus images. We classified the studies in this field into two categories including DR-detection-based studies and DR-severity-grading-based studies. Most studies graded fundus images into the severity levels suggested by the ICDR.

Almost all of the latest DL networks have been used efficiently by different studies for DR detection and grading. It was also noticed that there has been a considerable increase in the number of studies in this field recently. A list of the commonly used retinal fundus image datasets for DR detection and grading was also created in this study. Similar studies from each of the two categories of DR studies were compared according to their performance using the commonly used metrics of accuracy, sensitivity, and specificity. In future work, we will make a similar survey about the latest DR segmentation and lesion detection studies that have used DL.

Author Contributions: Conceptualization, A.S., O.E. and S.A.-M.; data curation, A.S.; formal analysis, A.S.; methodology, A.S., O.E. and S.A.-M.; project administration, S.A. and N.A.; supervision, S.A.-M. and N.A.; validation, A.S., O.E., S.A.-M. and N.A.; visualization, A.S. and O.E.; writing—original draft, A.S.; writing—review and editing, A.S., O.E., S.A.-M. and N.A. All authors have read and agreed to the published version of the manuscript.

Funding: Open Access funding provided by the Qatar National Library.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This publication was supported by Qatar University Internal Grant QUHI-CENG-22/23-548. The findings achieved herein are solely the responsibility of the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds ... are available from the authors.

Abbreviations

The following abbreviations are used in this manuscript:

DR	Diabetic retinopathy
DL	Deep learning
AI	Artificial intelligence
CNN	Convolutional neural network

References

1. IDF Diabetes Atlas 9th Edition. Available online: <https://diabetesatlas.org/atlas/ninth-edition/> (accessed on 1 August 2022).
2. Nijalingappa, P.; Sandeep, B. Machine learning approach for the identification of diabetes retinopathy and its stages. In Proceedings of the 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, Karnataka, India, 29–31 October 2015; pp. 653–658.
3. Raja, C.; Balaji, L. An automatic detection of blood vessel in retinal images using convolution neural network for diabetic retinopathy detection. *Pattern Recognit. Image Anal.* **2019**, *29*, 533–545. [CrossRef]
4. Nazir, T.; Irtaza, A.; Shabbir, Z.; Javed, A.; Akram, U.; Mahmood, M.T. Diabetic retinopathy detection through novel tetragonal local octa patterns and extreme learning machines. *Artif. Intell. Med.* **2019**, *99*, 101695. [CrossRef] [PubMed]
5. Gayathri, S.; Gopi, V.P.; Palanisamy, P. Diabetic retinopathy classification based on multipath CNN and machine learning classifiers. *Phys. Eng. Sci. Med.* **2021**, *44*, 639–653. [CrossRef] [PubMed]
6. Washburn, P.S. Investigation of severity level of diabetic retinopathy using adaboost classifier algorithm. *Mater. Today Proc.* **2020**, *33*, 3037–3042. [CrossRef]
7. Li, X.; Shen, L.; Shen, M.; Tan, F.; Qiu, C.S. Deep learning based early stage diabetic retinopathy detection using optical coherence tomography. *Neurocomputing* **2019**, *369*, 134–144. [CrossRef]
8. Islam, K.T.; Wijewickrema, S.; O’Leary, S. Identifying diabetic retinopathy from oct images using deep transfer learning with artificial neural networks. In Proceedings of the 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, 5–7 June 2019; pp. 281–286.
9. Wahid, M.F.; Hossain, A.A. Classification of Diabetic Retinopathy from OCT Images using Deep Convolutional Neural Network with BiLSTM and SVM. In Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 6–8 July 2021; pp. 1–5.
10. Kazakh-British, N.P.; Pak, A.; Abdullina, D. Automatic detection of blood vessels and classification in retinal images for diabetic retinopathy diagnosis with application of convolution neural network. In Proceedings of the 2018 International Conference on Sensors, Signal and Image Processing, Prague, Czech Republic, 12–14 October 2018; pp. 60–63.
11. Anoop, B. Binary Classification of DR-Diabetic Retinopathy using CNN with Fundus Colour Images. *Mater. Today Proc.* **2022**, *58*, 212–216.
12. Nasir, N.; Oswald, P.; Alshaltone, O.; Barneih, F.; Al Shabi, M.; Al-Shammaa, A. Deep DR: Detection of Diabetic Retinopathy using a Convolutional Neural Network. In Proceedings of the 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 21–24 February 2022; pp. 1–5.
13. Chakrabarty, N. A deep learning method for the detection of diabetic retinopathy. In Proceedings of the 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gorakhpur, India, 2–4 November 2018; pp. 1–5.
14. Jiang, Y.; Wu, H.; Dong, J. Automatic screening of diabetic retinopathy images with convolution neural network based on caffe framework. In Proceedings of the 1st International Conference on Medical and Health Informatics 2017, Taichung city, Taiwan, 20–22 May 2017; pp. 90–94.
15. Niranjana, R.; Narayanan, K.L.; Rani, E.F.I.; Agalya, A.; Chandraleka, C.; Indhumathi, K. Resourceful Retinal Vessel segmentation for Early Exposure of Vision Threatening Diseases. In Proceedings of the 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), Coimbatore, India, 4–5 March 2022; pp. 1–6.
16. Rêgo, S.; Dutra-Medeiros, M.; Soares, F.; Monteiro-Soares, M. Screening for Diabetic Retinopathy Using an Automated Diagnostic System Based on Deep Learning: Diagnostic Accuracy Assessment. *Ophthalmologica* **2021**, *244*, 250–257. [CrossRef]
17. Umapathy, A.; Sreenivasan, A.; Nairy, D.S.; Natarajan, S.; 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, Singapore, 7–9 January 2019; pp. 17–21.
18. Kolla, M.; Venugopal, T. Efficient Classification of Diabetic Retinopathy using Binary CNN. In Proceedings of the 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 17–18 March 2021; pp. 244–247.
19. Pamadi, A.M.; Ravishankar, A.; Nithya, P.A.; Jahnavi, G.; Kathavate, S. Diabetic Retinopathy Detection using MobileNetV2 Architecture. In Proceedings of the 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 25–26 March 2022; pp. 1–5.
20. Saranya, P.; Devi, S.K.; Bharanidharan, B. Detection of Diabetic Retinopathy in Retinal Fundus Images using DenseNet based Deep Learning Model. In Proceedings of the 2022 International Mobile and Embedded Technology Conference (MECON), Noida, India, 10–11 March 2022; pp. 268–272.

21. Mudaser, W.; Padungweang, P.; Mongkolnam, P.; Lavangnananda, P. Diabetic Retinopathy Classification with pre-trained Image Enhancement Model. In Proceedings of the 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 1–4 December 2021; pp. 629–632.
22. Chetoui, M.; Akhloufi, M.A. Explainable diabetic retinopathy using EfficientNET. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 1966–1969.
23. Zhang, Z. Deep-learning-based early detection of diabetic retinopathy on fundus photography using efficientnet. In Proceedings of the 2020 the 4th International Conference on Innovation in Artificial Intelligence, Shenzhen, China, 7–9 August 2020; pp. 70–74.
24. Sudarmadji, P.W.; Pakan, P.D.; Dillak, R.Y. Diabetic Retinopathy Stages Classification using Improved Deep Learning. In Proceedings of the 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 19–20 November 2020; pp. 104–109.
25. Boral, Y.S.; Thorat, S.S. Classification of Diabetic Retinopathy based on Hybrid Neural Network. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; pp. 1354–1358.
26. Sanjana, S.; Shadin, N.S.; Farzana, M. Automated Diabetic Retinopathy Detection Using Transfer Learning Models. In Proceedings of the 2021 5th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), Mirpur, Dhaka, 18–20 November 2021; pp. 1–6.
27. Hossen, M.S.; Reza, A.A.; Mishu, M.C. An automated model using deep convolutional neural network for retinal image classification to detect diabetic retinopathy. In Proceedings of the International Conference on Computing Advancements, Colombo, Sri Lanka, 10–11 December 2020; pp. 1–8.
28. Qomariah, D.U.N.; Tjandrasa, H.; Faticah, C. Classification of diabetic retinopathy and normal retinal images using CNN and SVM. In Proceedings of the 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 18 July 2019; pp. 152–157.
29. Kumar, N.S.; Karthikeyan, B.R. Diabetic Retinopathy Detection using CNN, Transformer and MLP based Architectures. In Proceedings of the 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Hualien City, Taiwan, 16–19 November 2021; pp. 1–2.
30. Kaushik, H.; Singh, D.; Kaur, M.; Alshazly, H.; Zagaria, A.; Hamam, H. Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models. *IEEE Access* **2021**, *9*, 108276–108292. [[CrossRef](#)]
31. Lian, C.; Liang, Y.; Kang, R.; Xiang, Y. Deep convolutional neural networks for diabetic retinopathy classification. In Proceedings of the 2nd International Conference on Advances in Image Processing, Chengdu, China, 6–18 June 2018; pp. 68–72.
32. Lahmar, C.; Idri, A. Deep hybrid architectures for diabetic retinopathy classification. In *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*; Taylor & Francis: Oxfordshire, UK, 2022; pp. 1–19.
33. Dwivedi, S.A.; Kalin, Y.A.; Roy, G.K.; Singla, K. Real-Time Detection for Normal, Glaucomatous and Diabetic Retinopathy Eyes for Ophthalmoscopy using Deep Learning. In Proceedings of the 2022 9th International Conference on Computing for Sustainable Global Development (INDIACOM), New Delhi, India, 23–25 March 2022; pp. 598–603.
34. Jiang, H.; Xu, J.; Shi, R.; Yang, K.; Zhang, D.; Gao, M.; Ma, H.; Qian, W. A multi-label deep learning model with interpretable Grad-CAM for diabetic retinopathy classification. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 1560–1563.
35. Quellec, G.; Charrière, K.; Boudi, Y.; Cochener, B.; Lamard, M. Deep image mining for diabetic retinopathy screening. *Med. Image Anal.* **2017**, *39*, 178–193. [[CrossRef](#)]
36. Tang, X.; Huang, Y.; Lin, L.; Li, M.; Yuan, J. Automated Diabetic Retinopathy Identification via Lesion Guided Network. In Proceedings of the The Fourth International Symposium on Image Computing and Digital Medicine, Shenyang, China, 5–8 December 2020; pp. 141–144.
37. Gundluru, N.; Rajput, D.S.; Lakshmann, K.; Kaluri, R.; Shoruzzaman, M.; Uddin, M.; Rahman Khan, M.A. Enhancement of Detection of Diabetic Retinopathy Using Harris Hawks Optimization with Deep Learning Model. *Comput. Intell. Neurosci.* **2022**, *2022*, 8512469. [[CrossRef](#)]
38. Gunasekaran, K.; Pitchai, R.; Chaitanya, G.K.; Selvaraj, D.; Annie Sheryl, S.; Almoallim, H.S.; Alharbi, S.A.; Raghavan, S.; Tesemma, B.G. A Deep Learning Framework for Earlier Prediction of Diabetic Retinopathy from Fundus Photographs. *BioMed Res. Int.* **2022**, *2022*, 3163496. [[CrossRef](#)] [[PubMed](#)]
39. Baba, S.M.; Bala, I. Detection of Diabetic Retinopathy with Retinal Images using CNN. In Proceedings of the 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 25–27 May 2022; pp. 1074–1080.
40. El-Ateif, S.; Idri, A. Single-modality and joint fusion deep learning for diabetic retinopathy diagnosis. *Sci. Afr.* **2022**, *17*, e01280. [[CrossRef](#)]
41. Wilkinson, C.; Ferris III, F.L.; Klein, R.E.; Lee, P.P.; Agardh, C.D.; Davis, M.; Dills, D.; Kampik, A.; Pararajasegaram, R.; Verdaguer, J.T.; et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* **2003**, *110*, 1677–1682. [[CrossRef](#)] [[PubMed](#)]
42. Harshitha, C.; Asha, A.; Pushkala, J.L.S.; Anogini, R.N.S.; Karthikeyan, C. Predicting the Stages of Diabetic Retinopathy using Deep Learning. In Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 20–22 January 2021; pp. 1–6.

43. Luo, X.; Pu, Z.; Xu, Y.; Wong, W.K.; Su, J.; Dou, X.; Ye, B.; Hu, J.; Mou, L. MVDRNet: Multi-view diabetic retinopathy detection by combining DCNNs and attention mechanisms. *Pattern Recognit.* **2021**, *120*, 108104. [CrossRef]
44. Baget-Bernaldiz, M.; Pedro, R.A.; Santos-Blanco, E.; Navarro-Gil, R.; Valls, A.; Moreno, A.; Rashwan, H.A.; Puig, D. Testing a Deep Learning Algorithm for Detection of Diabetic Retinopathy in a Spanish Diabetic Population and with MESSIDOR Database. *Diagnostics* **2021**, *11*, 1385. [CrossRef] [PubMed]
45. Maistry, A.; Pillay, A.; Jembere, E. Improving the accuracy of diabetes retinopathy image classification using augmentation. In Proceedings of the Conference of the South African Institute of Computer Scientists and Information Technologists 2020, Cape Town, South Africa, 14–16 September 2020; pp. 134–140.
46. Agustin, T.; Sunyoto, A. Optimization Convolutional Neural Network for Classification Diabetic Retinopathy Severity. In Proceedings of the 2020 3rd International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 November 2020; pp. 66–71.
47. Li, X.; Pang, T.; Xiong, B.; Liu, W.; Liang, P.; Wang, T. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In Proceedings of the 2017 10th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI), Shanghai, China, 14–16 October 2017; pp. 1–11.
48. Ardiyanto, I.; Nugroho, H.A.; Buana, R.L.B. Deep learning-based diabetic retinopathy assessment on embedded system. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, South Korea, 11–15 July 2017; pp. 1760–1763.
49. Kwasigroch, A.; Jarzembski, B.; Grochowski, M. Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujscie, Poland, 9–12 May 2018; pp. 111–116.
50. Wang, X.; Lu, Y.; Wang, Y.; Chen, W.B. Diabetic retinopathy stage classification using convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI), Salt Lake City, UT, USA, 7–9 July 2018; pp. 465–471.
51. Zhou, K.; Gu, Z.; Liu, W.; Luo, W.; Cheng, J.; Gao, S.; Liu, J. Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 2724–2727.
52. Shrivastava, U.; Joshi, M. Automated Multiclass Diagnosis of Diabetic Retinopathy using Hierarchical Learning. In Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing, Hyderabad, India, 18–22 December 2018; pp. 1–7.
53. Arora, M.; Pandey, M. Deep neural network for diabetic retinopathy detection. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; pp. 189–193.
54. Kassani, S.H.; Kassani, P.H.; Khazaeinezhad, R.; Wesolowski, M.J.; Schneider, K.A.; Deters, R. Diabetic retinopathy classification using a modified xception architecture. In Proceedings of the 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman, United Arab Emirates, 10–12 December 2019; pp. 1–6.
55. Hathwar, S.B.; Srinivasa, G. Automated grading of diabetic retinopathy in retinal fundus images using deep learning. In Proceedings of the 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 17–19 September 2019; pp. 73–77.
56. Bellemo, V.; Lim, Z.W.; Lim, G.; Nguyen, Q.D.; Xie, Y.; Yip, M.Y.; Hamzah, H.; Ho, J.; Lee, X.Q.; Hsu, W.; et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: A clinical validation study. *Lancet Digit. Health* **2019**, *1*, e35–e44. [CrossRef]
57. Kumar, S. Diabetic Retinopathy Diagnosis with Ensemble Deep-Learning. In Proceedings of the 3rd International Conference on Vision, Image and Signal Processing, Vancouver, BC, Canada, 26–28 August 2019; pp. 1–5.
58. Thota, N.B.; Reddy, D.U. Improving the accuracy of diabetic retinopathy severity classification with transfer learning. In Proceedings of the IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), Springfield, MA, USA, 9–12 August 2020; pp. 1003–1006.
59. Nguyen, Q.H.; Muthuraman, R.; Singh, L.; Sen, G.; Tran, A.C.; Nguyen, B.P.; Chua, M. Diabetic retinopathy detection using deep learning. In Proceedings of the 4th International Conference on Machine Learning and Soft Computing, Haiphong City, Vietnam, 17–19 January 2020; pp. 103–107.
60. Lavanya, R.V.; Sumesh, E.; Jayakumari, C.; Isaac, R. Detection and Classification of Diabetic Retinopathy using Raspberry PI. In Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 5–7 November 2020; pp. 1688–1691.
61. Elzennary, A.; Soliman, M.; Ibrahim, M. Early Deep Detection for Diabetic Retinopathy. In Proceedings of the 2020 International Symposium on Advanced Electrical and Communication Technologies (ISAECT), Kenitra, Morocco, 25–27 November 2020; pp. 1–5.
62. Barhate, N.; Bhave, S.; Bhise, R.; Sutar, R.G.; Karia, D.C. Reducing Overfitting in Diabetic Retinopathy Detection using Transfer Learning. In Proceedings of the 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 30–31 October 2020; pp. 298–301.

63. Wang, S.; Wang, X.; Hu, Y.; Shen, Y.; Yang, Z.; Gan, M.; Lei, B. Diabetic retinopathy diagnosis using multichannel generative adversarial network with semisupervision. *IEEE Trans. Autom. Sci. Eng.* **2020**, *18*, 574–585. [[CrossRef](#)]
64. Khaled, O.; El-Sahhar, M.; El-Dine, M.A.; Talaat, Y.; Hassan, Y.M.; Hamdy, A. Cascaded architecture for classifying the preliminary stages of diabetic retinopathy. In Proceedings of the 2020 9th International Conference on Software and Information Engineering (ICSIE), Cairo, Egypt, 11–13 November 2020; pp. 108–112.
65. Islam, M.R.; Hasan, M.A.M.; Sayeed, A. Transfer learning based diabetic retinopathy detection with a novel preprocessed layer. In Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 5–7 June 2020; pp. 888–891.
66. Wang, J.; Bai, Y.; Xia, B. Simultaneous diagnosis of severity and features of diabetic retinopathy in fundus photography using deep learning. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3397–3407. [[CrossRef](#)]
67. AbdelMaksoud, E.; Barakat, S.; Elmogy, M. Diabetic Retinopathy Grading Based on a Hybrid Deep Learning Model. In Proceedings of the 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, 26–27 October 2020; pp. 1–6.
68. Yaqoob, M.K.; Ali, S.F.; Kareem, I.; Fraz, M.M. Feature-based optimized deep residual network architecture for diabetic retinopathy detection. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 5–7 November 2020; pp. 1–6.
69. Taufiqurrahman, S.; Handayani, A.; Hermanto, B.R.; Mengko, T.L.E.R. Diabetic retinopathy classification using a hybrid and efficient MobileNetV2-SVM model. In Proceedings of the 2020 IEEE REGION 10 CONFERENCE (TENCON), Osaka, Japan, 16–19 November 2020; pp. 235–240.
70. Vaishnavi, J.; Ravi, S.; Anbarasi, A. An efficient adaptive histogram based segmentation and extraction model for the classification of severities on diabetic retinopathy. *Multimed. Tools Appl.* **2020**, *79*, 30439–30452. [[CrossRef](#)]
71. Shankar, K.; Sait, A.R.W.; Gupta, D.; Lakshmanaprabu, S.; Khanna, A.; Pandey, H.M. Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. *Pattern Recognit. Lett.* **2020**, *133*, 210–216. [[CrossRef](#)]
72. Karki, S.S.; Kulkarni, P. Diabetic Retinopathy Classification using a Combination of EfficientNets. In Proceedings of the 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 5–7 March 2021; pp. 68–72.
73. Qian, Z.; Wu, C.; Chen, H.; Chen, M. Diabetic retinopathy grading using attention based convolution neural network. In Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 March 2021; Volume 5, pp. 2652–2655.
74. Sharfuzzaman, M.; Hossain, M.S.; El Saddik, A. An Explainable Deep Learning Ensemble Model for Robust Diagnosis of Diabetic Retinopathy Grading. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2021**, *17*, 1–24. [[CrossRef](#)]
75. Sugeno, A.; Ishikawa, Y.; Ohshima, T.; Muramatsu, R. Simple methods for the lesion detection and severity grading of diabetic retinopathy by image processing and transfer learning. *Comput. Biol. Med.* **2021**, *137*, 104795. [[CrossRef](#)]
76. Lee, C.H.; Ke, Y.H. Fundus images classification for Diabetic Retinopathy using Deep Learning. In Proceedings of the 2021 The 13th International Conference on Computer Modeling and Simulation, Melbourne, Australia, 25–27 June 2021; pp. 264–270.
77. Nazir, T.; Nawaz, M.; Rashid, J.; Mahum, R.; Masood, M.; Mehmood, A.; Ali, F.; Kim, J.; Kwon, H.Y.; Hussain, A. Detection of diabetic eye disease from retinal images using a deep learning based CenterNet model. *Sensors* **2021**, *21*, 5283. [[CrossRef](#)] [[PubMed](#)]
78. Xiao, Z.; Zhang, Y.; Wu, J.; Zhang, X. SE-MIDNet Based on Deep Learning for Diabetic Retinopathy Classification. In Proceedings of the 2021 7th International Conference on Computing and Artificial Intelligence, Tianjin, China, 23–26 April 2021; pp. 92–98.
79. Li, Y.; Song, Z.; Kang, S.; Jung, S.; Kang, W. Semi-Supervised Auto-Encoder Graph Network for Diabetic Retinopathy Grading. *IEEE Access* **2021**, *9*, 140759–140767. [[CrossRef](#)]
80. Martinez-Murcia, F.J.; Ortiz, A.; Ramírez, J.; Górriz, J.M.; Cruz, R. Deep residual transfer learning for automatic diagnosis and grading of diabetic retinopathy. *Neurocomputing* **2021**, *452*, 424–434. [[CrossRef](#)]
81. Rajkumar, R.; Jagathishkumar, T.; Ragul, D.; Selvarani, A.G. Transfer learning approach for diabetic retinopathy detection using residual network. In Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 20–22 January 2021; pp. 1189–1193.
82. Swedhaasri, M.; Parekh, T.; Sharma, A. A Multi-Stage Deep Transfer Learning Method for Classification of Diabetic Retinopathy in Retinal images. In Proceedings of the 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 4–6 August 2021; pp. 1143–1149.
83. Reguant, R.; Brunak, S.; Saha, S. Understanding inherent image features in CNN-based assessment of diabetic retinopathy. *Sci. Rep.* **2021**, *11*, 1–12. [[CrossRef](#)]
84. Hari, K.N.; Karthikeyan, B.; Reddy, M.R.; Seethalakshmi, R. Diabetic Retinopathy Detection with Feature Enhancement and Deep Learning. In Proceedings of the 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 30–31 July 2021; pp. 1–5.
85. Saeed, F.; Hussain, M.; Aboalsamh, H.A. Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network. *IEEE Access* **2021**, *9*, 41344–41359. [[CrossRef](#)]
86. Jabbar, M.K.; Yan, J.; Xu, H.; Ur Rehman, Z.; Jabbar, A. Transfer Learning-Based Model for Diabetic Retinopathy Diagnosis Using Retinal Images. *Brain Sci.* **2022**, *12*, 535. [[CrossRef](#)]
87. Shaik, N.S.; Cherukuri, T.K. Hinge attention network: A joint model for diabetic retinopathy severity grading. *Appl. Intell.* **2022**, *52*, 15105–15121. [[CrossRef](#)]

88. Chandrasekaran, R.; Loganathan, B. Retinopathy grading with deep learning and wavelet hyper-analytic activations. *Vis. Comput.* **2022**, *1*–16. [CrossRef]
89. Oulhadj, M.; Riffi, J.; Chaimae, K.; Mahraz, A.M.; Ahmed, B.; Yahyaouy, A.; Fouad, C.; Meriem, A.; Idriss, B.A.; Tairi, H. Diabetic retinopathy prediction based on deep learning and deformable registration. *Multimed. Tools Appl.* **2022**, *81*, 28709–28727. [CrossRef]
90. Nair, A.T.; Anitha, M.; Kumar, A. Disease Grading of Diabetic Retinopathy using Deep Learning Techniques. In Proceedings of the 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 29–31 March 2022; pp. 1019–1024.
91. Deepa, V.; Kumar, C.S.; Cherian, T. Pre-Trained Convolutional Neural Network for Automated Grading of Diabetic Retinopathy. In Proceedings of the 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Tiruchirappalli, India, 16–18 February 2022; pp. 1–5.
92. Farag, M.M.; Fouad, M.; Abdel-Hamid, A.T. Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module. *IEEE Access* **2022**, *10*, 38299–38308. [CrossRef]
93. Canayaz, M. Classification of diabetic retinopathy with feature selection over deep features using nature-inspired wrapper methods. *Appl. Soft Comput.* **2022**, *128*, 109462. [CrossRef]
94. Bilal, A.; Zhu, L.; Deng, A.; Lu, H.; Wu, N. AI-Based Automatic Detection and Classification of Diabetic Retinopathy Using U-Net and Deep Learning. *Symmetry* **2022**, *14*, 1427. [CrossRef]
95. Murugappan, M.; Prakash, N.; Jeya, R.; Mohanarathinam, A.; Hemalakshmi, G.; Mahmud, M. A novel few-shot classification framework for diabetic retinopathy detection and grading. *Measurement* **2022**, *200*, 111485. [CrossRef]
96. Chen, C.Y.; Chang, M.C. Using Deep Neural Networks to Classify the Severity of Diabetic Retinopathy. In Proceedings of the 2022 IEEE International Conference on Consumer Electronics-Taiwan, Taipei, Taiwan, 6–8 July 2022; pp. 241–242.
97. Butt, M.M.; Iskandar, D.; Abdelhamid, S.E.; Latif, G.; Alghazo, R. Diabetic Retinopathy Detection from Fundus Images of the Eye Using Hybrid Deep Learning Features. *Diagnostics* **2022**, *12*, 1607. [CrossRef]
98. Elwin, J.G.R.; Mandala, J.; Maram, B.; Kumar, R.R. Ar-HGSO: Autoregressive-Henry Gas Sailfish Optimization enabled deep learning model for diabetic retinopathy detection and severity level classification. *Biomed. Signal Process. Control* **2022**, *77*, 103712. [CrossRef]
99. Deepa, V.; Sathish Kumar, C.; Cherian, T. Automated grading of diabetic retinopathy using CNN with hierarchical clustering of image patches by siamese network. *Phys. Eng. Sci. Med.* **2022**, *45*, 623–635. [CrossRef]
100. Araújo, T.; Aresta, G.; Mendonça, L.; Penas, S.; Maia, C.; Carneiro, Á.; Mendonça, A.M.; Campilho, A. DR | GRADUATE: Uncertainty-aware deep-learning-based diabetic retinopathy grading in eye fundus images. *Med. Image Anal.* **2020**, *63*, 101715. [CrossRef]
101. Bhawarkar, Y.; Bhure, K.; Chaudhary, V.; Alte, B. Diabetic Retinopathy Detection From Fundus Images Using Multi-Tasking Model With EfficientNet B5. In Proceedings of the ITM Web of Conferences; EDP Sciences: Navi Mumbai, India 2022; Volume 44, p. 03027.
102. Sakaguchi, A.; Wu, R.; Kamata, S.i. Fundus image classification for diabetic retinopathy using disease severity grading. In Proceedings of the 2019 9th International Conference on Biomedical Engineering and Technology, Tokyo, Japan, 28–30 March 2019; pp. 190–196.
103. Gulshan, V.; Rajan, R.P.; Widner, K.; Wu, D.; Wubbels, P.; Rhodes, T.; Whitehouse, K.; Coram, M.; Corrado, G.; Ramasamy, K.; et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol.* **2019**, *137*, 987–993. [CrossRef]
104. Elswah, D.K.; Elnakib, A.A.; Moustafa, H.E.d. Automated diabetic retinopathy grading using resnet. In Proceedings of the 2020 37th National Radio Science Conference (NRSC), Cairo, Egypt, 8–10 September 2020; pp. 248–254.
105. Dong, C.W.; Xia, D.; Jin, J.; Yang, Z. Classification of diabetic retinopathy based on DSIRNet. In Proceedings of the 2019 14th International Conference on Computer Science & Education (ICCSE), Toronto, ON, Canada, 19–21 August 2019; pp. 692–696.
106. Jiang, H.; Yang, K.; Gao, M.; Zhang, D.; Ma, H.; Qian, W. An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 2045–2048.
107. Harihanth, K.; Karthikeyan, B. Diabetic Retinopathy Detection using ensemble deep Learning and Individual Channel Training. In Proceedings of the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 3–5 December 2020; pp. 1042–1049.
108. Nithiyasri, M.; Ananthi, G.; Thiruvengadam, S. Improved Classification of Stages in Diabetic Retinopathy Disease using Deep Learning Algorithm. In Proceedings of the 2022 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, 24–26 March 2022; pp. 143–147.
109. Kassimi, A.E.B.; Madiafi, M.; Kammour, A.; Bouroumi, A. A Deep Neural Network for Detecting the Severity Level of Diabetic Retinopathy from Retinography Images. In Proceedings of the 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 3–4 March 2022; pp. 1–7.
110. Rahhal, D.; Alhamouri, R.; Albataineh, I.; Duwairi, R. Detection and Classification of Diabetic Retinopathy Using Artificial Intelligence Algorithms. In Proceedings of the 2022 13th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 21–23 June 2022; pp. 15–21.

111. Meenakshi, G.; Thailambal, G. Categorisation and Prognostication of Diabetic Retinopathy using Ensemble Learning and CNN. In Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 28–30 April 2022; pp. 1145–1152.
112. Jaskari, J.; Sahlsten, J.; Damoulas, T.; Knoblauch, J.; Särkkä, S.; Kärkkäinen, L.; Hietala, K.; Kaski, K.K. Uncertainty-aware deep learning methods for robust diabetic retinopathy classification. *IEEE Access* **2022**, *10*, 76669–76681. [[CrossRef](#)]
113. AbdelMaksoud, E.; Barakat, S.; Elmogy, M. A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. *Med. Biol. Eng. Comput.* **2022**, *60*, 2015–2038. [[CrossRef](#)]
114. Rajavel, R.; Sundaramoorthy, B.; GR, K.; Ravichandran, S.K.; Leelasankar, K. Cloud-enabled Diabetic Retinopathy Prediction System using optimized deep Belief Network Classifier. *J. Ambient. Intell. Humaniz. Comput.* **2022**, *1*–9. [[CrossRef](#)]
115. Sri, K.S.; Priya, G.K.; Kumar, B.P.; Sravya, S.D.; Priya, M.B. Diabetic Retinopathy Classification using Deep Learning Technique. In Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 28–30 April 2022; pp. 1492–1496.
116. Ragab, M.; Aljedaibi, W.H.; Nahhas, A.F.; Alzahrani, I.R. Computer aided diagnosis of diabetic retinopathy grading using spiking neural network. *Comput. Electr. Eng.* **2022**, *101*, 108014. [[CrossRef](#)]
117. Bhardwaj, C.; Jain, S.; Sood, M. Appraisal of preprocessing techniques for automated detection of diabetic retinopathy. In Proceedings of the 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan, India, 20–22 December 2018; pp. 734–739.
118. Al-Mohannadi, A.; Al-Maadeed, S.; Elharrouss, O.; Sadasivuni, K.K. Encoder-decoder architecture for ultrasound IMC segmentation and cIMT measurement. *Sensors* **2021**, *21*, 6839. [[CrossRef](#)]
119. Elharrouss, O.; Akbari, Y.; Almaadeed, N.; Al-Maadeed, S. Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv* **2022**, arXiv:2206.08016.
120. Riahi, A.; Elharrouss, O.; Al-Maadeed, S. BEMD-3DCNN-based method for COVID-19 detection. *Comput. Biol. Med.* **2022**, *142*, 105188. [[CrossRef](#)]
121. Elharrouss, O.; Subramanian, N.; Al-Maadeed, S. An encoder–decoder-based method for segmentation of COVID-19 lung infection in CT images. *SN Comput. Sci.* **2022**, *3*, 1–12. [[CrossRef](#)]
122. Hacisofotaoglu, R.E.; Karakaya, M.; Sallam, A.B. Deep learning frameworks for diabetic retinopathy detection with smartphone-based retinal imaging systems. *Pattern Recognit. Lett.* **2020**, *135*, 409–417. [[CrossRef](#)]
123. Keel, S.; Wu, J.; Lee, P.Y.; Scheetz, J.; He, M. Visualizing deep learning models for the detection of referable diabetic retinopathy and glaucoma. *JAMA Ophthalmol.* **2019**, *137*, 288–292. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

A Novel Proposal for Deep Learning-Based Diabetes Prediction: Converting Clinical Data to Image Data

Muhammet Fatih Aslan *  and Kadir Sabancı 

Department of Electrical and Electronics Engineering, Karamanoglu Mehmetbey University, Karaman 70100, Turkey
* Correspondence: mfatihaslan@kmu.edu.tr

Abstract: Diabetes, one of the most common diseases worldwide, has become an increasingly global threat to humans in recent years. However, early detection of diabetes greatly inhibits the progression of the disease. This study proposes a new method based on deep learning for the early detection of diabetes. Like many other medical data, the PIMA dataset used in the study contains only numerical values. In this sense, the application of popular convolutional neural network (CNN) models to such data are limited. This study converts numerical data into images based on the feature importance to use the robust representation of CNN models in early diabetes diagnosis. Three different classification strategies are then applied to the resulting diabetes image data. In the first, diabetes images are fed into the ResNet18 and ResNet50 CNN models. In the second, deep features of the ResNet models are fused and classified with support vector machines (SVM). In the last approach, the selected fusion features are classified by SVM. The results demonstrate the robustness of diabetes images in the early diagnosis of diabetes.

Keywords: convolutional neural network; diabetes prediction; numeric-to-image; PIMA dataset; support vector machines



Citation: Aslan, M.F.; Sabancı, K. A Novel Proposal for Deep Learning-Based Diabetes Prediction: Converting Clinical Data to Image Data. *Diagnostics* **2023**, *13*, 796. <https://doi.org/10.3390/diagnostics13040796>

Academic Editors: Sami Akbulut and Cemil Colak

Received: 24 January 2023

Revised: 14 February 2023

Accepted: 15 February 2023

Published: 20 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The most prevalent chronic non-communicable disease in the world is diabetes, also known as diabetes mellitus. Diabetes is fatal or drastically lowers quality of life and affects more women than men [1]. Diabetes is particularly risky for pregnant women, and unborn children are likely to be affected by this disease. Generally, if the glucose level in the blood rises above the normal value, the person is considered diabetic. This is due to the inability of the pancreas in the human body to fully perform its task. The person's blood sugar rises if the pancreas cannot utilize the insulin it produces or does not create enough of it. Diabetes can cause long-term damage to different organs such as the eyes, heart, kidneys and blood vessels [2]. There are three different types of diabetes: type 1, type 2 and gestational. In type 1 diabetes, the pancreas produces little or no insulin. Insulin therapy is needed. It is usually seen in young individuals (age < 30) or children. Type 2 is usually caused by insulin resistance and is more common in older (age > 65) and obese patients [3–5]. Gestational diabetes is hyperglycemia that occurs during pregnancy. In addition, after pregnancy, the risk of type 2 diabetes is higher in women, and in this case, babies are also at risk [6,7].

It is known that diabetes is a public health problem that affects 60% of the world's population [8]. Although the main cause of diabetes is unknown, scientists think it is related to genetic factors and environmental conditions. There are currently 425 million diabetics worldwide, according to the International Diabetes Federation, and 625 million will develop the disease in the next 23 years [9,10]. It is essential to identify the disease at an early stage in order to stop this rise. Only early detection can stop the growth of the disease because there is no cure for diabetes, which is a lifetime condition. With the right treatment, regular nutrition and drugs, the disease can be managed after early

diagnosis. [11,12]. However, a delayed diagnosis might result in heart conditions and serious harm to many organs. For the early diagnosis of diabetes, clinical (plasma glucose concentration, serum insulin, etc.) and physical data (for example, body mass index (BMI), age) are often used [13]. According to these data, a doctor carries out the diagnosis of the disease. However, making a medical diagnosis is a very difficult task for the doctor and can take a very long time. In addition, the decisions made by the doctor may be erroneous and biased. For this reason, the fields called data mining and machine learning are frequently used as a decision support mechanism for the rapid and accurate detection of diseases according to data [11,14,15].

Recent advances in computer technologies have led to the emergence of algorithms that allow human tasks to be performed faster and more automatically by computers. Tools such as data mining, machine learning and deep learning, which are generally referred to as artificial intelligence, have shown remarkable performance in interpreting existing data. Especially in the medical field, artificial-intelligence-based methods are used in the diagnosis or treatment of many different diseases as they provide fast and powerful results. Examples of these are diagnostic studies of cancer [16], diabetes [17], COVID-19 [18], heart diseases [19], brain tumors [20], Alzheimer's [21], etc. For more comprehensive information on the applications of artificial intelligence in the medical field, research studies by Kaur et al. [22] and Mirbabaie et al. [23] can be reviewed. Artificial intelligence is very useful for the medical field. Thanks to the superior success of artificial intelligence in medical studies so far, it has recently become common to record medical big data in hospitals. Considering that each patient is a real data point, much numerical data such as electrocardiograms (ECG), electromyograms (EMG), clinical data, blood values or a large number of image data such as X-ray, magnetic resonance imaging (MRI) or computed tomography (CT) can be produced after medical records. In this sense, such medical records constitute an important part of big data in the medical field [24].

Machine learning algorithms are generally used to interpret (regression, classification or clustering) big data based on artificial intelligence. Thanks to these algorithms, the relationship between them is learned based on samples and observations of the data. Machine learning methods that are frequently used in this sense are artificial neural networks (ANN), support vector machines (SVM), k-nearest neighbors (k-NN), decision trees (DT) and naïve Bayes (NB). These methods directly learn the correlation between input and target data. However, with the developments in artificial intelligence and computer processors in the last decade, ANN has been further deepened, and deep learning, which applies both feature extraction and classification together, has come to the fore. Especially in big data applications, deep learning has given a great advantage over traditional machine learning methods [25]. The most frequently used model in deep-learning-based medical diagnosis/detection applications is convolutional neural network (CNN). CNN models are very popular due to both their deep architecture and high-level feature representation. Since the architecture designed for CNN is end to end, raw data are given as input and classes are obtained as output. Therefore, the designed architecture is very important for the performance of the CNN model [26]. Recently, however, researchers have adopted transfer learning applications and used popular CNN architectures such as ResNet [27], GoogleNet [28], Inception [29], Xception [30], VGGNet [31], etc. In different data-driven studies [32], the direct use of pre-trained or pre-designed CNN architectures has provided advantages in terms of both performance and convenience.

1.1. Previous Artificial Intelligence Based Studies on Diabetes Prediction

This study performs deep-learning-based diabetes prediction using the PIMA dataset. In general, studies developed for diabetes prediction are based on machine learning or deep learning.

Some of the studies that applied diabetes prediction to the PIMA dataset using machine learning methods are as follows. Zolfaghari [33] performed diabetes detection based on an ensemble of SVM and feedforward neural network. For this, the results obtained from the

individual classifiers were combined using the majority voting technique. The ensemble approach provided a better result than the individual classifiers with 88.04% success. Sneha and Gangil [34] performed diabetes prediction using many machine learning methods such as naïve Bayes (NB), SVM and logistic regression. The best accuracy was obtained with SVM with 77.37%. In addition, the authors applied feature selection for the PIMA dataset. The features with low correlation were removed. Edeh et al. [35] compared four machine learning algorithms, Bayes, decision tree (DT), SVM and random forest (RF), on two different datasets for diabetes prediction. In the experimental results with PIMA, the highest accuracy was obtained with SVM at 83.1%. Chen et al. [36] reorganized the PIMA data with preprocessing and removed the misclassified data with the k-means algorithm (data reduction). They then classified the reduced data with DT. As a result of the study, diabetes was predicted with an accuracy of 90.04%. Dadgar and Kaardaan [37] proposed a hybrid technique for diabetes prediction. First, feature selection was performed with the UTA algorithm. Then, the selected features were given to the two-layer neural network (NN) whose weights were updated by genetic algorithm (GA). As a result, diabetes estimation was provided with an accuracy of 87.46%. Zou et al. [38] used DT, RF, and NN models for diabetes prediction. They also used principal component analysis (PCA) and minimum redundancy maximum relevance (mRMR) to reduce dimensionality. As a result, RF performed more successful predictions than the others, with 77.21% accuracy. For other proposed studies based on machine learning, studies by Choudhury and Gupta [39] and Rajeswari and Prabhu [40] can be examined.

The following are some studies that use the PIMA dataset with deep learning models: For diabetes prediction, Ashiquzzaman et al. [41] created a network with an input layer, fully connected layers, dropouts and an output layer architecture. It fed the PIMA dataset features directly into this designed MLP and achieved an accuracy of 88.41% at the end of the application. Massaro et al. [42] created artificial records and classified these data with long short-term memory (LSTM) (LSTM-AR). The LSTM-AR classification result, which was stated as 89%, was superior to both LSTM and the multi-layer perceptron (MLP) with cross validation previously performed. Kannadasan et al. [43] designed a deep neural network that extracts features with stacked autoencoders and performs diabetes classification with softmax. The designed deep architecture provided 86.26% accuracy. Rahman et al. [44] presented a model based on convolutional LSTM (Conv-LSTM). They also experimented with traditional LSTM and CNN to compare the results. They applied grid search algorithm for hyperparameter optimization in deep models. For all models, the input layer was one dimensional (1D). After training and test separation, Conv-LSTM for test data outperformed other models, with 91.38% accuracy. Alex et al. [45] designed a 1D CNN architecture for diabetes prediction. However, missing values were corrected by outlier detection. Then, they preprocessed the data with synthetic minority oversampling technique (SMOTE), and the imbalance in the data were removed. They then fed the processed data into the 1D CNN architecture and achieved 86.29% accuracy. For other applications based on deep learning for diabetes prediction, the studies presented by Zhu et al. [46] and Fregoso-Aparicio et al. [47] can be examined.

Previous studies show that the PIMA dataset is often used for machine learning, 1D-CNN and LSTM structures. The numerical nature of the PIMA dataset has limited the feature extraction and classification algorithms that researchers can use. In this study, this limitation is overcome by converting numerical data to images. Thus, the PIMA numerical dataset will be applicable with popular CNN models such as ResNet, VGGNet and GoogleNet.

1.2. The Structure, Purpose, Differences and Contribution of the Study

Examining the previous studies mentioned in Section 1.1. reveals that various machine learning and deep-learning-based applications predict diabetes quite successfully for the PIMA dataset containing clinical data records. Similar to the PIMA dataset, many clinical data in the medical field are composed of numerical values. Using numerical values directly

with conventional machine learning techniques is more typical because studies involving machine learning models such as SVM, NB, RF, DT, etc. feed raw data or data with small preprocessing directly to the model and give target (0 (negative)–1 (positive)) values to the output. Studies that design deep architecture using the same data feed the PIMA features either to the 1D convolution layer or to the fully connected layers. The study by Massaro, Maritati, Giannone, Convertini and Galiano [42] processed the PIMA dataset containing 1D data with a recurrent-neural-network (RNN)-based LSTM. Nevertheless, LSTM was designed for sequential data, whereas the PIMA dataset contains independent data.

Traditional machine learning techniques have been surpassed in many respects by deep learning, which has become more popular in recent years [48,49]. With the high-level capabilities they offer, particularly deep CNN models, they have shown greater performance, notably in computer vision applications. However, the PIMA dataset's inclusion of numeric values has thus far prompted researchers to create 1D CNN models. Popular CNN models are created for computer vision, and therefore the input layer only accepts 2D data. These models are employed in transfer learning applications. As a result, feature extraction using well-known CNN models and a diabetes prediction using these models have not yet been established from this PIMA dataset containing independent numerical data. Therefore, in order to provide more successful diagnoses, transformation can be applied to the raw data in accordance with popular CNN models.

This study converts each sample in the PIMA dataset to images (diabetes images) to overcome this limitation. Each diabetes image has cells representing features in the PIMA dataset. The ReliefF feature selection algorithm [50–52] was also used to make the feature with high correlation more dominant in the image. After each feature is placed on the image according to its importance, data augmentation is applied for these images. In fact, the easy application of data augmentation for diabetes data is one of the important contributions of this study because compared to numerical data, data augmentation for images is an easier and more common technique. The augmented image data are then fed to the ResNet18 and ResNet50 CNN models and diabetes prediction is performed. In order to improve these current results, the features of both models are then fused and classified with SVM (CNN-SVM). Finally, feature selection is made with the ReliefF algorithm, among many fusion features, and these selected features are classified by SVM. At the end of the study, all these results are compared. According to the results, the CNN-SVM structure with selected fusion features provides more successful diabetes prediction than others. In addition, the results of the proposed method are compared with those of previous studies, and the method is proven to be effective. The contributions of the proposed method can be stated as follows:

- An application with an end-to-end structure is suggested for diabetes prediction.
- PIMA dataset with numeric values is converted to images.
- It is provided to use numerical diabetes data together with popular CNN models.
- During the conversion to the image, the importance of the features is taken into account.
- The proposed method is superior to most previous studies.

2. PIMA Indians Diabetes Dataset

In this study, the PIMA Indians Diabetes dataset, which is taken from the Kaggle data repository and is frequently preferred for diabetes prediction, is used. The access link is <https://data.world/data-society/pima-indians-diabetes-database> (Access Date: 8 June 2022). The National Institute of Diabetes and Digestive and Kidney Diseases provided the source data for this dataset. The dataset's goal is to diagnose whether or not a patient has diabetes based on certain diagnostic metrics provided in the collection. All patients here, in particular, are PIMA Indian women over the age of 21.

The dataset includes the following measurements and ranges of clinical and physical characteristics. Pregnancies (number, [0–17]), glucose (value, [0–199]), blood pressure (mm Hg, [0–122]), skin thickness (mm, [0–99]), insulin (μU/mL, [0–846]), BMI (kg/m², [0–67.1]), diabetes pedigree function (PDF) (value, [0.078–2.42]), age (years, [21–81]), and

outcome (Boolean- 0, 1). The data are entirely numerical and comprise a total of 8 features and 768 samples. Table 1 shows a few samples from the dataset.

Table 1. Some examples of Pima Indians Diabetes dataset.

Pregnancy [0–17]	Glucose [0–199]	Blood Pressure [0–122]	Skin Thickness [0–99]	Serum Insulin [0–846]	BMI [0–67.1]	PDF [0.078–2.42]	Age [21–81]	Outcome (0–1)
1	89	66	23	94	28.1	0.167	21	0
2	197	70	45	543	30.5	0.158	53	1
1	189	60	23	846	30.1	0.398	59	1
1	103	30	38	83	43.3	0.183	33	0
9	171	110	24	240	45.4	0.721	54	1
5	88	66	21	23	24.4	0.342	30	0
2	141	58	34	128	25.4	0.699	24	0
2	100	66	20	90	32.9	0.867	28	1
7	83	78	26	71	29.3	0.767	36	0
7	160	54	32	175	30.5	0.588	39	1

3. Methodology

The methods used to determine the diabetes status of patients will be outlined in detail in this section. The steps of the proposed method are shown in Figure 1. The feature selection method initially selects the most useful features from the numerical data, as shown in Figure 1. The boundaries of all features are then adjusted for the numeric-to-image conversion stage once the numerical data has been normalized. The numerical to image conversion process is applied in such a way that the most effective features determined by the feature selection algorithm are dominant. The classification success of deep ResNet models is then increased by the use of data augmentation techniques. The three ResNet-based approaches suggested in this study are used to classify data in the final stage. Below, we go over each of these processes in more detail.

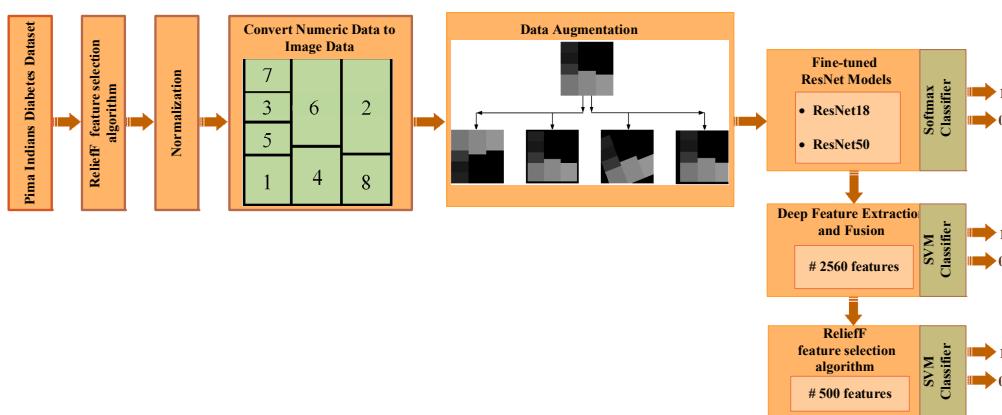


Figure 1. Application steps of proposed methods.

3.1. ReliefF Feature Selection Algorithm

To improve classification capability, a variety of feature reduction strategies have been explored in the literature [53]. In the literature, ReliefF is one of the distance-based feature selectors. ReliefF, developed by Kira and Rendell [52] in 1992, is one of the most successful feature filtering methods.

Dimension reduction strategies aid in the removal of superfluous attributes from a data set. These technologies aid in data compression, which saves storage space. It also shortens the time required for computational complexity and reduces the amount of time it takes to attain the same goal [54].

Kononenko [55] improved the algorithm for multi-class issues in 1994. With the help of this algorithm, feature selection can be performed successfully. The ReliefF algorithm is highly efficient and does not impose any restrictions on the data kinds' features. The ReliefF method assists in the solution of many classes of issues by selecting the nearest neighboring samples from each sample in each category [56].

ReliefF seeks to expose the connections and consistency found in the dataset's properties. Furthermore, by constructing a model that addresses the proximity to samples of the same class and distance to samples of different classes, it is feasible to discover the significant features in the dataset. Between samples of distinct qualities, the ReliefF model chooses neighboring attributes that are closest to each other [54]. The dataset is divided into two components in this model: training and test data. R_i random samples are chosen from the training set, and the difference function $diff$ is used to calculate the nearest neighbors of the same and different classes to identify the nearest neighbors to the selected R_i sample, as illustrated in Equation (1). When identifying nearest neighbors, the $diff$ function is also utilized to compute the distance between instances. The total distance is simply the sum of all attribute differences (i.e., Manhattan distance) [51].

Equation (1) is used to determine the difference between two separate I_1 and I_2 samples for the attribute A and to discover the closest distance between samples. The nearest neighbor H from the same class and the nearest neighbor M from a different class are chosen. The distance of adjacent sample A_f in the class and between the classes is compared based on the values of R_i , M , H , and the dataset's weighting vector. The WA_f weight is calculated as a result of the comparison by giving less weight to the distant attributes [57]. These processes are performed m times for each attribute, and the weight values are calculated for each attribute. The weights are updated using Equation (2) [55,58].

$$diff(A, I_1, I_2) = \frac{|value(A, I_1) - value(A, I_2)|}{\max(A) - \min(A)} \quad (1)$$

$$W_{new}(A_f) = W_{old}(A_f) + \frac{diff(A_f, R_i, M)}{m} - \frac{diff(A_f, R_i, H)}{m} \quad (2)$$

As a result of applying the ReliefF feature selection method described above to the PIMA dataset features, the importance weight of each feature is shown in Figure 2. The number of nearest neighbors was also determined as 10. As seen in Figure 2, the most effective features from the PIMA numerical data were determined by the ReliefF algorithm.

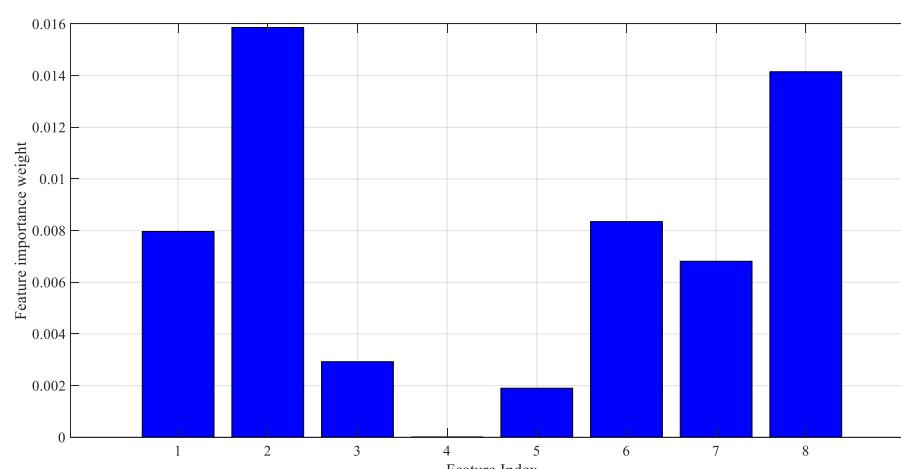


Figure 2. Importance weight of features in the PIMA dataset.

3.2. Normalization of Data

In artificial intelligence studies, normalizing data containing many features is a known process. Because different features have different limits. Setting features to the same or similar range, i.e., normalization, improves learning performance. The PIMA dataset also has different lower and upper bound values, as seen in Table 1. In this sense, normalization of these values is necessary. In addition, normalization is vital for the numeric-to-image conversion process in the proposed implementation because the value of each feature must be located on the image that represents that sample. According to the amplitude of the feature, the cell in the corresponding image has a brighter color. Therefore, the maximum and minimum values for all features must be the same.

The preferred method for normalization is feature scaling. With this method, feature values are rescaled to a certain range. The feature scaling method used in this study is the min–max normalization method. In this method, the new sample value (\hat{x}) is determined according to the maximum (x_{max}) and minimum (x_{min}) values of the features. As a result of normalization, all features are distributed between 0–1. In the application phase, normalization is applied for eight features in the PIMA dataset. Figure 3 shows that after this normalization, the glucose [0–199] and blood pressure [0–122] values range from 0–1. Equation (3) shows the formula for the min–max normalization method.

$$\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

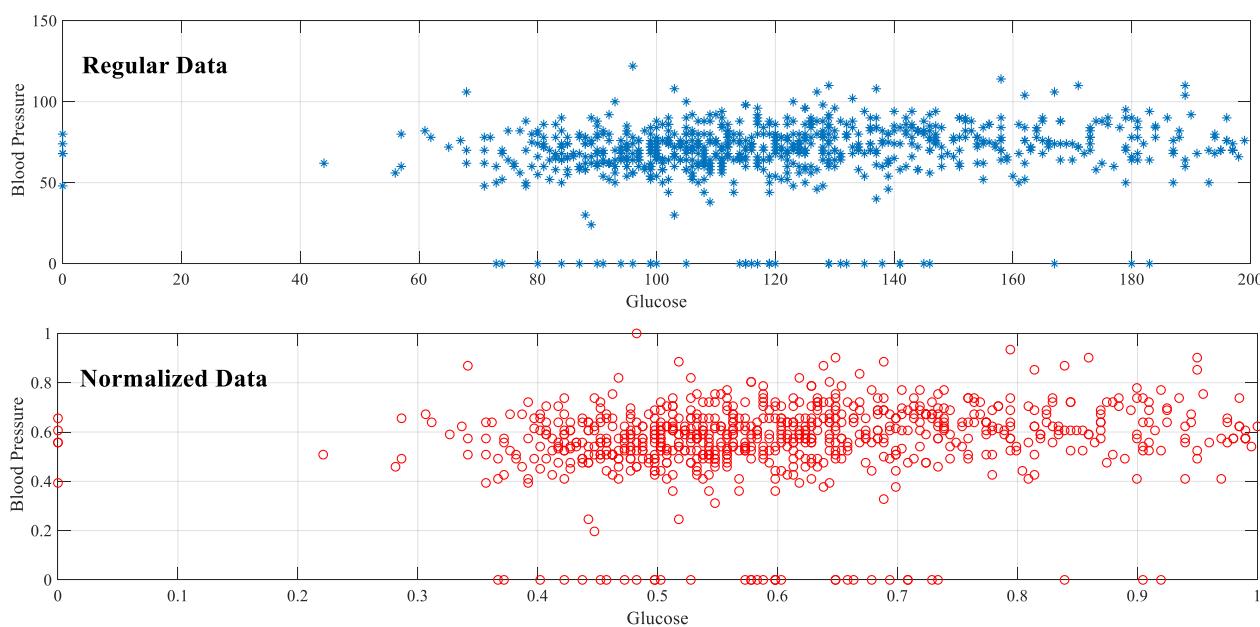


Figure 3. Min–max normalization of PIMA dataset.

3.3. Conversion of Numeric Data to Image Data

Although the number of image data in the medical field has increased considerably recently, there is still a large amount of numerical data available. Although numerical values are easily and cheaply obtained, the interpretation of these data is usually performed by machine learning methods. Recently proposed deep architecture studies prefer 1D CNN structures that take these numerical values as input because popular CNN models, which provide significant improvements in computer vision, cannot be used directly for such data. Because for these models, 2D data should be given as input to the input layer. CNN models such as ResNet, VGGNet, GoogleNet, etc., have an architecture designed for image data. Therefore, the inability to analyze datasets containing 1D samples with these powerful models is a major disadvantage in terms of both application diversity and

prediction performance. This section discusses the conversion of numeric data to images to overcome this limitation in the PIMA dataset, which is a numeric dataset.

In the process of converting PIMA data to images, the principle of determining the brightness of a specific region (cell) in the image according to the amplitude of each feature is adopted. In fact, each feature can be viewed as a piece of the sample image's puzzle. For each sample in the PIMA dataset, the 120×120 image structure shown in Figure 4 is used. The index on each cell corresponds to the feature index in the PIMA dataset. That is, Figure 4 shows feature locations in a sample image. In Figure 4, the location and size of features are determined not randomly but based on feature importance. As seen in Figure 2 as a result of the ReliefF algorithm, the order of importance of features is 2-8-6-1-7-3-5-4. Therefore, a larger cell is assigned for the more important feature. Each cell is colored according to the amplitude of the corresponding feature value. Because all data were previously normalized, each feature value ranges from 0 to 1. Each feature value is multiplied by 255, resulting in images with cells with brightness values between 0 and 255. Therefore, the resulting images are in gray spaces. Some sample diabetes images are shown in Figure 4.

Feature Positions in a Diabetes Image

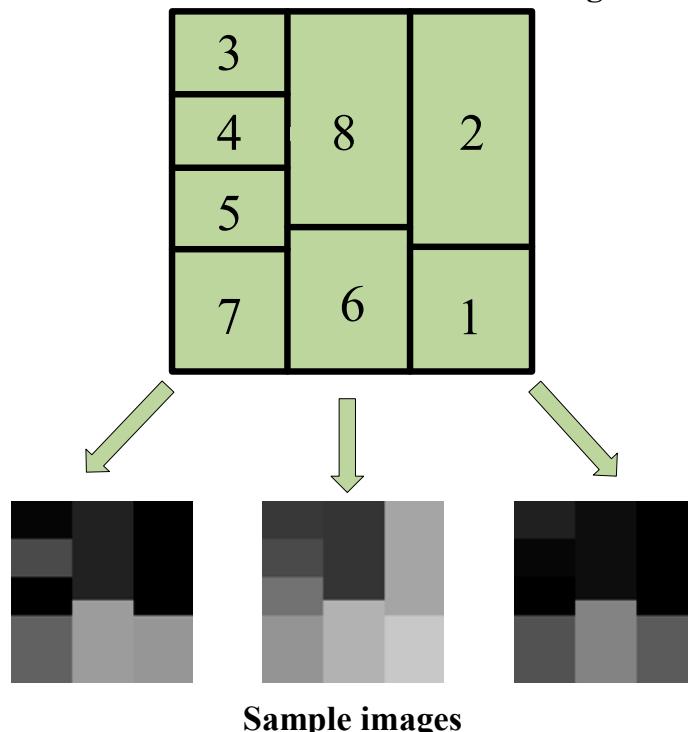


Figure 4. Conversion selected features to image (numeric to image).

As a result of applying the aforementioned image conversion method on the PIMA dataset, one image for each sample (that is, 768 images in total) is formed. These images, with all features included, can now be used in CNN models that require 2D data input. Furthermore, image data augmentation methods are easily applicable to these image data. For this purpose, the image structure in Figure 4 is designed asymmetrically because in the data augmentation stage, all images must be reproduced differently from each other.

3.4. Data Augmentation

The number of samples directly influences the success of deep learning approaches. However, accessing a significant volume of data is not always possible. As a result, researchers artificially increase the size of a training dataset by producing modified versions of the images in the dataset. These techniques, which are applied to raw images for this purpose, are known as data augmentation techniques.

In this study, the diabetic data contains 768 numerical samples in total, and hence 768 images are created during the conversion from numerical to image data. Data augmentation techniques are used because this amount is insufficient for a deep learning implementation. To ensure data diversity and robust training, four different data augmentation techniques (rotation, scale, reflection and translation) are applied to all images produced, as in Figure 4. Table 2 shows the lower and upper limit values for these data augmentation techniques. Additionally, Figure 5 shows the new diabetes images produced as a result of data augmentation techniques.

Table 2. Approaches to data augmentation's lower and upper limitations.

Parameter Name	Lower Limit	Upper Limit
Reflection	-	-
Rotation	-30°	30°
Scale	0.9	1.1
Translation	-10	+10

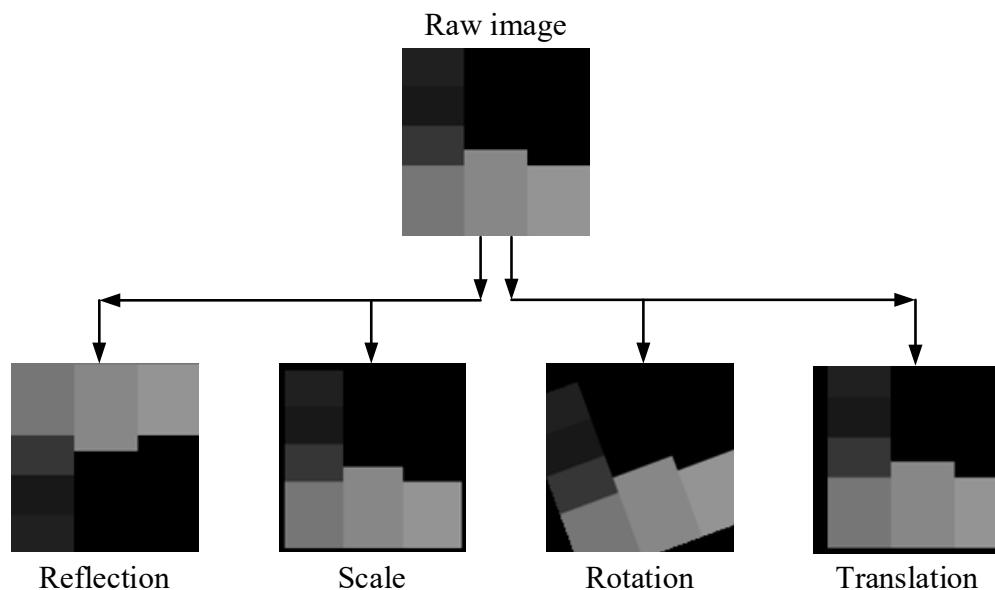


Figure 5. Data augmentation methodologies and sample augmented images.

After data augmentation, each original diabetes image is reproduced in four different ways. As a result, a total of five artificial samples is obtained from one sample. The sample numbers of the classes before and after the data augmentation stage are shown in Table 3. As a result of the data augmentation, the total number of images reached 3840.

Table 3. Examination of all data before and following data augmentation.

Class	0 (Negative)	1 (Positive)	Total
Before data augmentation	500	268	768
After data augmentation	2500	1340	3840

3.5. Diabetes Prediction via ResNet Models

After data augmentation, images separated into 80% training and 20% testing are fed to the CNN model. In this study, diabetes estimation is provided with the ResNet18 and ResNet50 models, which are frequently used for comparison purposes. Many studies apply ResNet models widely because of the advantages they provide [59]. What makes ResNet

preferable is that it transmits residual values to the next layers to avoid the vanishing gradients problem, for which it uses residual blocks. There are ResNet models with different depths. The depths of the ResNet18 and ResNet50 models used in this study are 18 and 50, respectively.

This study performs diabetes detection with existing models instead of designing a new CNN architecture. With only minor modifications (fine-tuning), existing ResNet models are adapted to our work. For both models, the last two layers of existing models are removed and replaced with two fully connected output layers and a classification (softmax) layer. In addition, while the diabetes images produced are 120×120 , the input size for ResNet models should be 224×224 . Therefore, all diabetes images are resized before and during training (see Figure 6). Information about the results obtained after the training and testing phases will be discussed in the results section.

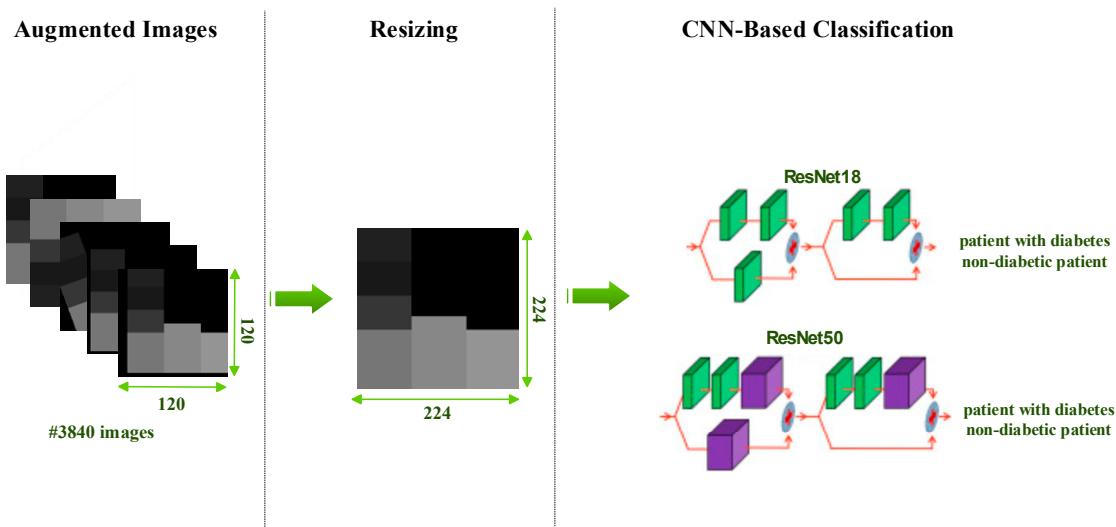


Figure 6. Classification of diabetes images as diabetic (1) and nondiabetic (0) with ResNet models.

3.6. Deep Feature Extraction, Feature Selection and Classification

While the previous section directly uses fine-tuning of ResNet models, this section describes the CNN-SVM structure. In other words, CNN is used for feature extraction and SVM is used for classification. This approach has been frequently preferred recently to increase the classification accuracy [60]. Two different experimental applications are presented at this stage. The features obtained with the CNN models in the previous stage are combined and fed to the SVM. In the previous step, 512 deep features were extracted from the ResNet18 model, and 2048 deep features were extracted from the ResNet50 model. Then, these features were combined to obtain a total of 2560 deep features, with 80% and 20% of these deep features being divided into two groups for training and testing. In the first experimental stage, these deep features are classified by the SVM machine learning algorithm. For this classification, linear, quadratic, cubic and Gaussian SVM kernel functions are used and results are obtained. In the second stage, the most effective 500 features from a total of 2560 features extracted from ResNet models are selected using the ReliefF feature selection algorithm. These 500 selected features are classified by the SVM machine learning algorithm. Similarly, at this stage, classification is made for SVM using linear, quadratic, cubic and Gaussian SVM kernel functions. All results are then compared. Figure 7 shows the proposed CNN-SVM structure.

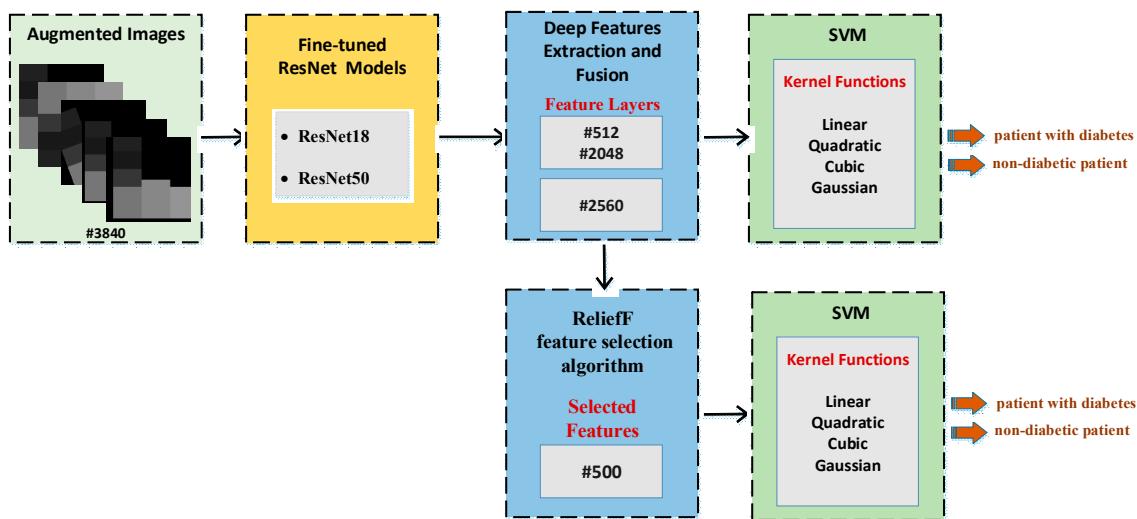


Figure 7. Implementation steps of the proposed CNN-SVM approach.

The results of the experimental studies are discussed in Section 4. Experimental application in the last step provided the most successful results. The flow graph containing the applications of this step is shown in Figure 8.

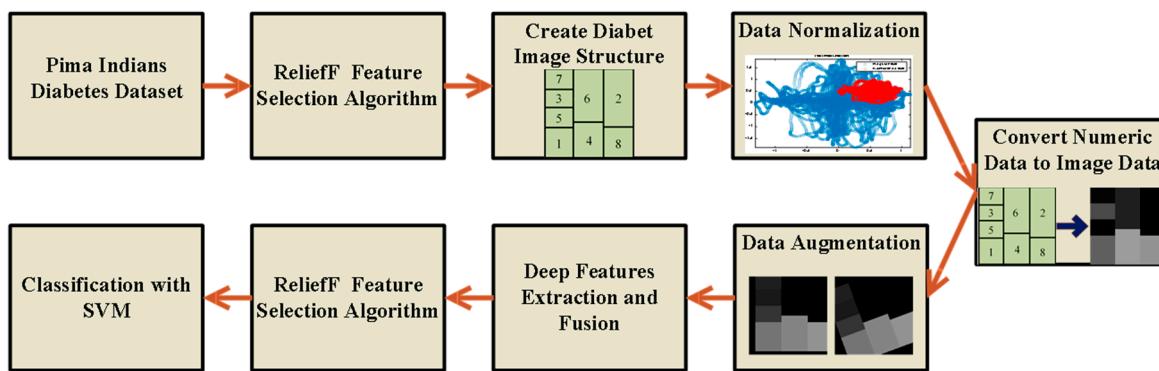


Figure 8. Application flow chart in the last step.

4. Results and Discussion

In this section, the results of the proposed approach are discussed. All deep learning applications for diabetes prediction were performed using a laptop with Intel Core i7-7700HG processor, NVIDIA GeForce GTX 1050 4 GB graphics card and 16 GB RAM. Applications are developed in Matlab environment. Figure 8 can be taken as reference for the software design or code implementation of the proposed approach. The algorithm of the method was created as in Figure 8. Toolbox and libraries used directly during coding prevented software complexity. Toolboxes used in this context are Machine Learning Toolbox, Deep Learning Toolbox and Image Processing Toolbox.

In order to demonstrate the superiority of the proposed method, results are produced with three different approaches. Methodological information about these three approaches has been shared in detail in the previous section. In the first approach, classification is performed with fine-tuned ResNet models to perform the diabetes prediction using diabetes images after data augmentation. For this, the ResNet18 and ResNet50 models are fine-tuned, and the output layer is changed according to the two classes. Then, 3840 diabetes images are divided into two groups as 80% and 20% training data and test data, respectively. While the models are trained using the training data, the performance of the network is obtained using the test data. In the second approach, deep features extracted from two fine-tuned ResNet models are combined and these fusion features (2560) are classified

by the SVM machine learning method. The performance of SVM differs according to the kernel function used. Therefore, in the second approach, classification accuracies are obtained by using linear, quadratic, cubic and Gaussian kernel functions and compared with each other. In the last approach, namely the proposed method, the most important 500 features from a total of 2560 fusion features extracted from fine-tuned ResNet models are selected with the ReliefF feature selection algorithm. In this way, we aimed to achieve similar success with fewer features. These features are classified with SVM as in the second approach. Classification results are obtained with linear, quadratic, cubic and Gaussian kernel functions, and the results are compared with other approaches.

ResNet models are trained once for all the approaches mentioned above. In other words, as a result of the three approaches, the features obtained with the ResNet models are the same. The parameters used for training ResNet models are: Mini Batch Size: 32; Max Epochs: 5; Learn Rate Drop factor: 0.1; Learn Rate Drop period: 20; Initial Learn Rate: 0.001. In addition, the optimizer used to update the weights during the training process is Stochastic Gradient Descent with Momentum.

After training is performed in the first approach, the features obtained in the first approach are used in other approaches. The accuracy and loss graph of the first approach, obtained during the training and testing phase of the ResNet18 and ResNet50 models, is shown in Figure 9. It is clear that overfitting does not occur during the training phase. In the second and third approaches, the CNN model is not trained, and 512 and 2048 features are extracted from ResNet18 and ResNet50, respectively, through fully connected layers used directly. The confusion matrixes obtained as a result of the classification of these features with SVM are shown in Figures 10 and 11. Figure 10 shows the application results obtained with the second approach using all the fusion features. Figure 11 shows the final application results that classify the selected features from the fusion features.

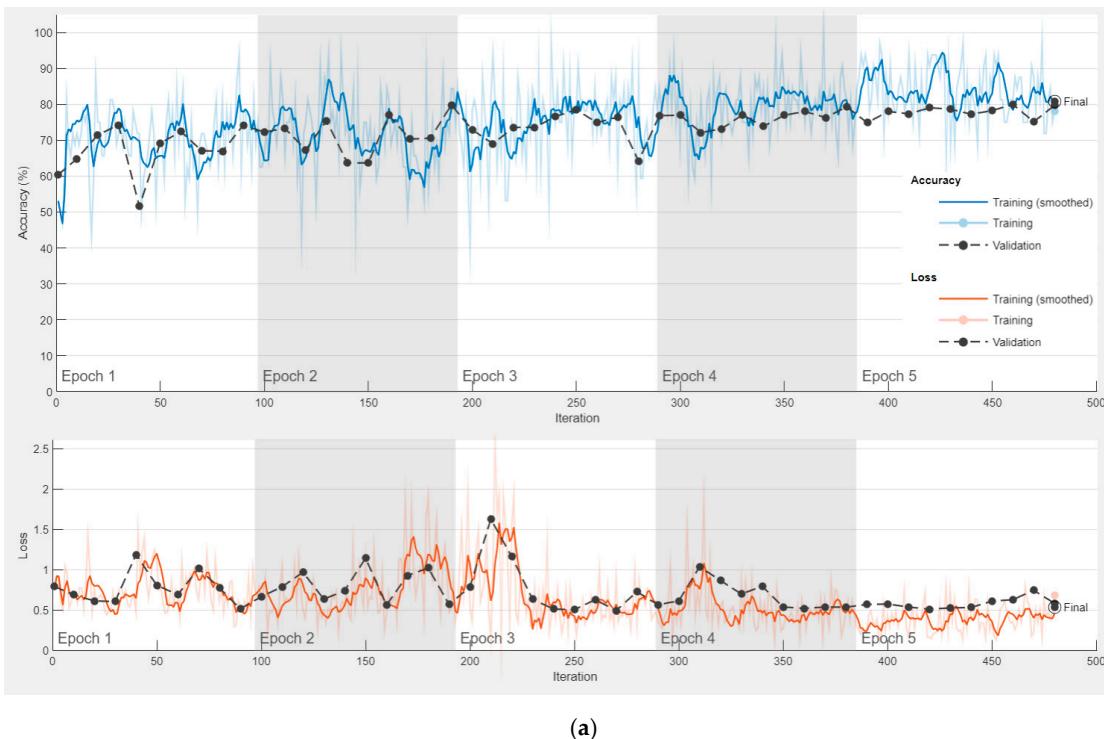
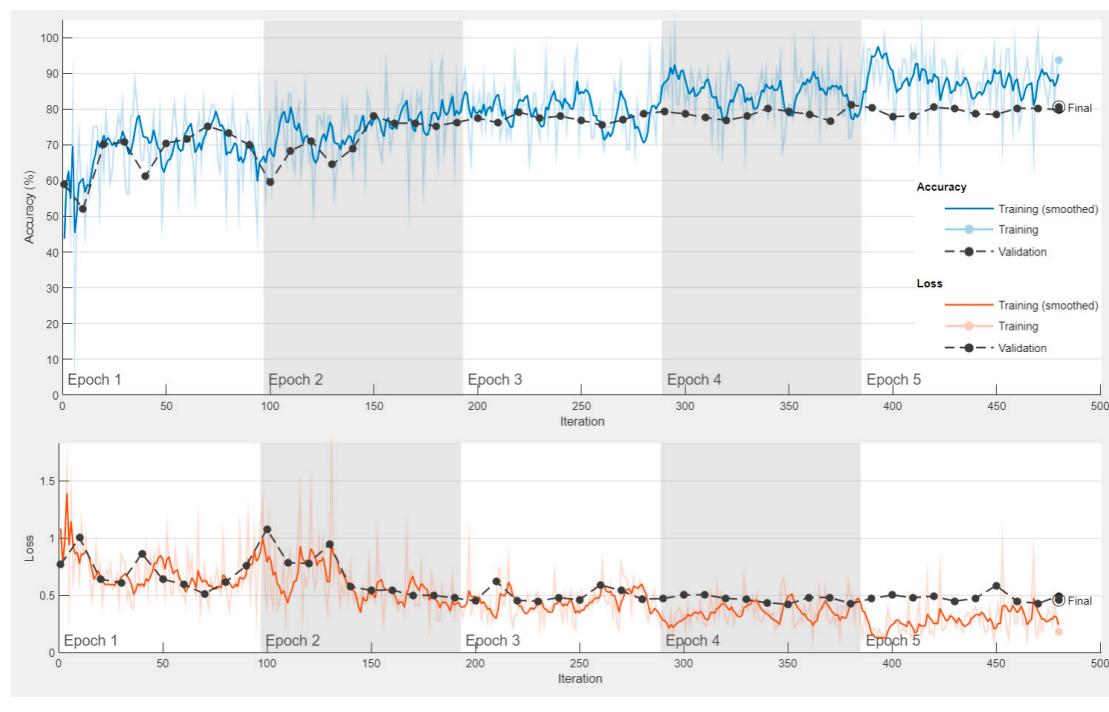
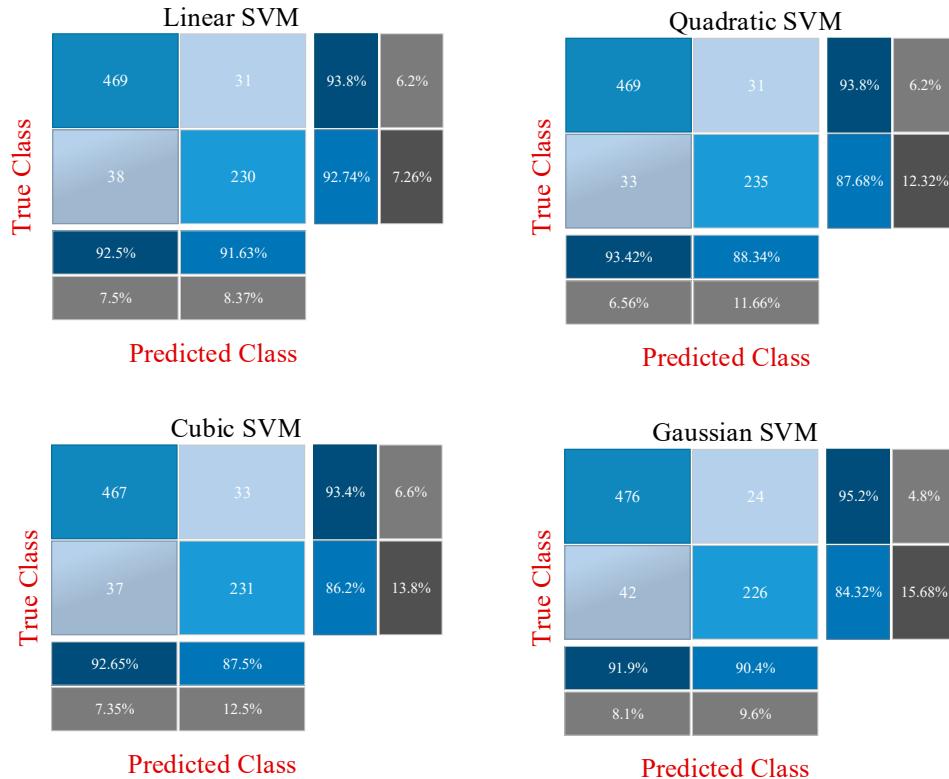


Figure 9. Cont.



(b)

Figure 9. Training and loss graphics of ResNet models. (a) ResNet18. (b) ResNet50.**Figure 10.** Confusion matrices obtained as a result of classification of all fused features with SVM.

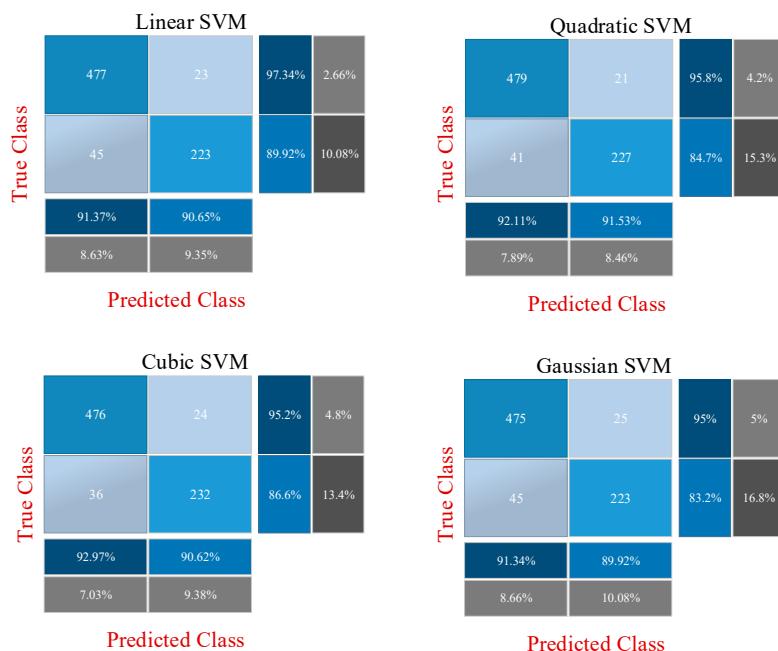


Figure 11. Confusion matrices obtained as a result of classification of selected features with SVM.

The confusion matrix structure that enables the calculation of these metrics is shown in Figure 12. The performance of the system is measured with the tp , tn , fp and fn values in this matrix. Using these values, accuracy, specificity, precision, sensitivity, F1-score and MCC performance metrics are calculated with the help of the formulas between Equations (4) and (9). Table 4 shows the performance metrics obtained as a result of the three approaches.

$$\text{Accuracy} = \frac{tp + tn}{tp + fp + tn + fn} \times 100 \quad (4)$$

$$\text{Specificity} = \frac{tn}{tn + fp} \quad (5)$$

$$\text{Precision} = \frac{tp}{tp + fp} \quad (6)$$

$$\text{Sensitivity} = \frac{tp}{tp + fn} \quad (7)$$

$$F1 - score = \frac{2tp}{2tp + fp + fn} \quad (8)$$

$$MCC = \frac{(tp \times tn) - (fn \times fp)}{\sqrt{(tp + fn) \times (tn + fp) \times (tp + fp) \times (tn + fn)}} \quad (9)$$

		PREDICTED	
		Positive	Negative
ACTUAL	Positive	TRUE POSITIVE tp	FALSE NEGATIVE fn
	Negative	FALSE POSITIVE fp	TRUE NEGATIVE tn

Figure 12. Structure of confusion matrices.

Table 4. Performance metrics for the three approaches.

Model	Kernel	Acc. (%)	Spec.	Prec.	Sens.	F1-Score	MCC
ResNet18	-	80.86	0.6689	0.8142	0.8947	0.8526	0.5868
ResNet50	-	80.47	0.5734	0.7826	0.9474	0.8571	0.5832
SVM with 2560 features	Linear	91.02	0.8582	0.9250	0.9380	0.9315	0.8012
	Quadratic	91.67	0.8769	0.9343	0.9380	0.9361	0.8163
	Cubic	90.89	0.8619	0.9266	0.9340	0.9340	0.7988
	Gaussian	91.41	0.8433	0.9189	0.9520	0.9352	0.8090
SVM with 500 selected features	Linear	91.15	0.8321	0.9138	0.9540	0.9335	0.8030
	Quadratic	91.93	0.8470	0.9212	0.9580	0.9392	0.8206
	Cubic	92.19	0.8657	0.9297	0.9520	0.9407	0.8268
	Gaussian	90.89	0.8321	0.9135	0.9500	0.9314	0.7972

According to Table 4, the highest accuracy in the first approach is obtained with the fine-tuned ResNet18 model. The accuracy rates obtained with ResNet18 and Resnet50 are 80.86% and 80.47%, respectively. In the second approach, in the classification made with SVM using 2560 features, the highest accuracy is calculated as 91.67% with the quadratic kernel function. In the last approach, in the classification made with the 500 most effective features selected by the feature selection algorithm, the highest accuracy is calculated with the SVM/cubic kernel function of 92.19%. The results of the first approach showed that converting diabetes data from numeric to image is an effective technique because these images were successfully classified with ResNet models. The second approach shows that fusing the features of different CNN models highly affects the success. In addition, SVM also showed a successful classification performance. The last approach showed that higher achievement can be achieved with fewer features. The results obtained with the last approach are compared with previous studies using the PIMA dataset, as shown in Table 5. As can be seen, the method proposed in our study outperformed many previous studies. Considering the methodological knowledge of previous studies, the numerical nature of the PIMA dataset has led researchers to use algorithms fed with numerical data such as traditional machine learning, 1D-CNN and LSTM. This study, unlike previous studies, transformed the PIMA dataset into image data and thus made the PIMA dataset suitable for popular CNN models.

Table 5. Comparative analysis with previous works.

Previous Work	Method	Accuracy (%)
Zolfaghari [33]	Ensemble of SVM and NN	88.04
Sneha and Gangil [34]	Feature Selection and SVM	77.37
Srivastava et al. [61]	ANN	92.00
Edeh, Khalaf, Tavera, Tayeb, Ghouali, Abdulsahib, Richard-Nnabu and Louni [35]	SVM	83.1
Massaro, Maritati, Giannone, Convertini and Galiano [42]	LSTM-AR	89
Dadgar and Kaardaan [37]	UTA-NN and GA	87.46
Zou, Qu, Luo, Yin, Ju and Tang [38]	mRMR-RF	77.21
Ashiquzzaman, Tushar, Islam, Shon, Im, Park, Lim and Kim [41]	Deep MLP	88.41
Kannadasan, Edla and Kuppili [43]	Stacked Autoencoders-DNN	86.26
Rahman, Islam, Mukti and Saha [44]	Conv-LSTM	91.38
Alex, Nayahi, Shine and Gopirekha [45]	DCNN/SMOTE/Outlier Detection	86.29
Kalagotla et al. [62]	Stacking of MLP, SVM, LR	78.2
Jakka and Vakula Rani [63]	LR	77.6
Proposed method	Diabetes images: ResNet18 and ResNet50-ReliefF	92.19

5. Conclusions, Discussion and Future Works

Diabetes is a chronic disease that limits people's daily activities, reduces their quality of life and increases the risk of death. In the past, machine learning and DNN solutions have been developed using clinical data and various diabetes prediction studies have been carried out. Despite the encouraging results of these studies, the numerical nature of clinical registry data has limited the use of popular CNN models. In this study, popular CNN models were used to determine the diagnosis of diabetes. Since these CNN models require two-dimensional data input, numerical clinical patient data (PIMA dataset) were first converted to images in this study. In this way, each feature was included in the sample image. This process was not performed randomly, and the most effective feature was made to stand out more in the image. During this process, the ReliefF feature selection method was used to determine the most effective features. After the number of generated images was increased by data augmentation and their size was adjusted for the ResNet model, diabetes prediction was carried out with three different approaches.

Diabetes images were successfully classified with the first approach using the fine-tuned ResNet18 and ResNet50 models. In the second approach, SVM was used to classify a total of 2560 deep features extracted from the fully connected layers of both ResNet models. In the last approach, the most effective 500 of these deep features were selected using the ReliefF feature selection algorithm, and the selected features were classified by SVM. The most successful prediction was obtained with the third approach. The accuracy of the classification using the SVM/cubic model with 500 selected features was 92.19%. All these classifications were performed on the image data. The conversion to image data removed the algorithm limitation that can be used for the PIMA dataset. In this way, the PIMA dataset or similar numerical data can be analyzed with different CNN models capable of extracting high-level and complex features. An application containing image data can be analyzed more diversely and comprehensively than an application containing numerical data because the different artificial intelligence combinations that can be applied to the image data are very rich. The results obtained with the ResNet18 and ResNet50 models in this study, therefore, outperform previous studies. For example, the number and variety of features can be increased with different CNN models. Based on all this, the experimental results have shown that converting clinical data into images is an effective technique.

The method proposed in this study can also be applied for different numerical data. Deep-learning-based studies have reduced the dependency on features and the designed architecture has come to the fore. However, the method proposed in this study is valuable in that deep and comprehensive architectures can also be used for numerical data. This application may involve more processing steps than studies using raw data directly. However, the generation of image data paves the way for further improvement of diabetes prediction performance because CNN models in many different architectures are now applicable to numerical data. Moreover, data augmentation can now be easily applied to diabetes images. In addition, the application results show that the fusion features used in the CNN-SVM architecture greatly increase the success. Additionally, using selected features, CNN-SVM is less costly and provides more accurate predictions. Based on these situations, the important trend of experimental simulation studies can be explained as follows: selected fusion features increase the performance of the system, although they are fewer in number. In addition, the CNN-SVM structure is quite effective. Different applications with fewer, more effective and more diverse features increase the classification accuracy of the system.

In future studies, it is planned to use of different CNN models and feature selection methods to improve diabetes prediction performance. A greater variety of features will be obtained by using more CNN models. In this case, it is expected that the classification accuracy will increase. In addition, future studies plan to apply the produced diabetes images with transformer-based networks.

Author Contributions: Conceptualization, M.F.A. and K.S.; methodology, M.F.A.; software, M.F.A. and K.S.; validation, M.F.A.; formal analysis, M.F.A. and K.S.; investigation, M.F.A.; writing—original draft preparation, M.F.A.; writing—review and editing, M.F.A. and K.S.; visualization, M.F.A. and K.S.; supervision, K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data utilized in this work can be found at <https://data.world/data-society/pima-indians-diabetes-database> (accessed on 8 June 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bhoi, S.K. Prediction of diabetes in females of pima Indian heritage: A complete supervised learning approach. *Turk. J. Comput. Math. Educ. (TURCOMAT)* **2021**, *12*, 3074–3084.
- Mellitus, D. Diagnosis and classification of diabetes mellitus. *Diabetes Care* **2005**, *28*, S5–S10.
- Madan, P.; Singh, V.; Chaudhari, V.; Albagory, Y.; Dumka, A.; Singh, R.; Gehlot, A.; Rashid, M.; Alshamrani, S.S.; AlGhamdi, A.S. An Optimization-Based Diabetes Prediction Model Using CNN and Bi-Directional LSTM in Real-Time Environment. *Appl. Sci.* **2022**, *12*, 3989. [[CrossRef](#)]
- Pippitt, K.; Li, M.; Gurgele, H.E. Diabetes mellitus: Screening and diagnosis. *Am. Fam. Physician* **2016**, *93*, 103–109. [[PubMed](#)]
- Xu, G.; Liu, B.; Sun, Y.; Du, Y.; Snetselaar, L.G.; Hu, F.B.; Bao, W. Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: Population based study. *BMJ* **2018**, *362*, k1497. [[CrossRef](#)] [[PubMed](#)]
- Chieffari, E.; Arcidiacono, B.; Foti, D.; Brunetti, A. Gestational diabetes mellitus: An updated overview. *J. Endocrinol. Investig.* **2017**, *40*, 899–909. [[CrossRef](#)] [[PubMed](#)]
- Mirghani Dirar, A.; Doupis, J. Gestational diabetes from A to Z. *World J. Diabetes* **2017**, *8*, 489–511. [[CrossRef](#)] [[PubMed](#)]
- Sarwar, A.; Ali, M.; Manhas, J.; Sharma, V. Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. *Int. J. Inf. Technol.* **2020**, *12*, 419–428. [[CrossRef](#)]
- Zhou, H.; Myrzashova, R.; Zheng, R. Diabetes prediction model based on an enhanced deep neural network. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 148. [[CrossRef](#)]
- Cho, N.; Shaw, J.; Karuranga, S.; Huang, Y.; da Rocha Fernandes, J.; Ohlrogge, A.; Malanda, B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res. Clin. Pract.* **2018**, *138*, 271–281. [[CrossRef](#)]
- Azrar, A.; Ali, Y.; Awais, M.; Zaheer, K. Data mining models comparison for diabetes prediction. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 320–323. [[CrossRef](#)]
- Larabi-Marie-Sainte, S.; Aburahmah, L.; Almohaini, R.; Saba, T. Current Techniques for Diabetes Prediction: Review and Case Study. *Appl. Sci.* **2019**, *9*, 4604. [[CrossRef](#)]
- Jahani, M.; Mahdavi, M. Comparison of predictive models for the early diagnosis of diabetes. *Healthc. Inform. Res.* **2016**, *22*, 95–100. [[CrossRef](#)] [[PubMed](#)]
- Ayon, S.I.; Islam, M. Diabetes Prediction: A Deep Learning Approach. *Int. J. Inf. Eng. Electron. Bus.* **2019**, *11*, 21.
- Naz, H.; Ahuja, S. Deep learning approach for diabetes prediction using PIMA Indian dataset. *J. Diabetes Metab. Disord.* **2020**, *19*, 391–403. [[CrossRef](#)]
- Aslan, M.F.; Celik, Y.; Sabanci, K.; Durdu, A. Breast Cancer Diagnosis by Different Machine Learning Methods Using Blood Analysis Data. *Int. J. Intell. Syst. Appl. Eng.* **2018**, *6*, 289–293. [[CrossRef](#)]
- Asiri, N.; Hussain, M.; Al Adel, F.; Alzaidi, N. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. *Artif. Intell. Med.* **2019**, *99*, 101701. [[CrossRef](#)] [[PubMed](#)]
- Aslan, M.F.; Sabanci, K.; Durdu, A.; Ulrsen, M.F. COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization. *Comput. Biol. Med.* **2022**, *142*, 105244. [[CrossRef](#)] [[PubMed](#)]
- Baashar, Y.; Alkawsi, G.; Alhussian, H.; Capretz, L.F.; Alwadain, A.; Alkahtani, A.A.; Almomani, M. Effectiveness of artificial intelligence models for cardiovascular disease prediction: Network meta-analysis. *Comput. Intell. Neurosci.* **2022**, *2022*, 5849995. [[CrossRef](#)]
- Saba, T.; Mohamed, A.S.; El-Affendi, M.; Amin, J.; Sharif, M. Brain tumor detection using fusion of hand crafted and deep learning features. *Cogn. Syst. Res.* **2020**, *59*, 221–230. [[CrossRef](#)]
- Abuhmed, T.; El-Sappagh, S.; Alonso, J.M. Robust hybrid deep learning models for Alzheimer’s progression detection. *Knowl.-Based Syst.* **2021**, *213*, 106688. [[CrossRef](#)]
- Kaur, S.; Singla, J.; Nkenyereye, L.; Jha, S.; Prashar, D.; Joshi, G.P.; El-Sappagh, S.; Islam, M.S.; Islam, S.M.R. Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives. *IEEE Access* **2020**, *8*, 228049–228069. [[CrossRef](#)]
- Mirbabaie, M.; Stieglitz, S.; Frick, N.R.J. Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction. *Health Technol.* **2021**, *11*, 693–731. [[CrossRef](#)]

24. Sun, H.; Liu, Z.; Wang, G.; Lian, W.; Ma, J. Intelligent Analysis of Medical Big Data Based on Deep Learning. *IEEE Access* **2019**, *7*, 142022–142037. [[CrossRef](#)]
25. Janiesch, C.; Zschech, P.; Heinrich, K. Machine learning and deep learning. *Electron. Mark.* **2021**, *31*, 685–695. [[CrossRef](#)]
26. Aslan, M.F.; Sabancı, K.; Durdu, A. A CNN-based novel solution for determining the survival status of heart failure patients with clinical record data: Numeric to image. *Biomed. Signal Process. Control* **2021**, *68*, 102716. [[CrossRef](#)]
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
28. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
29. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
30. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
31. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, preprint. arXiv:1409.1556.
32. Kirbaş, İ.; Çifci, A. An effective and fast solution for classification of wood species: A deep transfer learning approach. *Ecol. Inform.* **2022**, *69*, 101633. [[CrossRef](#)]
33. Zolfaghari, R. Diagnosis of diabetes in female population of pima indian heritage with ensemble of bp neural network and svm. *Int. J. Comput. Eng. Manag.* **2012**, *15*, 2230–27893.
34. Sneha, N.; Gangil, T. Analysis of diabetes mellitus for early prediction using optimal features selection. *J. Big Data* **2019**, *6*, 13. [[CrossRef](#)]
35. Edeh, M.O.; Khalaf, O.I.; Tavera, C.A.; Tayeb, S.; Ghouali, S.; Abdulsahib, G.M.; Richard-Nnabu, N.E.; Louni, A. A Classification Algorithm-Based Hybrid Diabetes Prediction Model. *Front. Public Health* **2022**, *10*, 829519. [[CrossRef](#)]
36. Chen, W.; Chen, S.; Zhang, H.; Wu, T. A hybrid prediction model for type 2 diabetes using K-means and decision tree. In Proceedings of the 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 24–26 November 2017; pp. 386–390.
37. Dadgar, S.M.H.; Kaardaan, M. A Hybrid Method of Feature Selection and Neural Network with Genetic Algorithm to Predict Diabetes. *Int. J. Mechatron. Electr. Comput. Technol. (IJMEC)* **2017**, *7*, 3397–3404.
38. Zou, Q.; Qu, K.; Luo, Y.; Yin, D.; Ju, Y.; Tang, H. Predicting Diabetes Mellitus With Machine Learning Techniques. *Front. Genet.* **2018**, *9*, 515. [[CrossRef](#)] [[PubMed](#)]
39. Choudhury, A.; Gupta, D. *A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques*; Springer: Singapore, 2019; pp. 67–78.
40. Rajeswari, M.; Prabhu, P. A review of diabetic prediction using machine learning techniques. *Int. J. Eng. Tech.* **2019**, *5*, 2395–1303.
41. Ashiquzzaman, A.; Tushar, A.K.; Islam, M.; Shon, D.; Im, K.; Park, J.-H.; Lim, D.-S.; Kim, J. Reduction of overfitting in diabetes prediction using deep learning neural network. In *IT Convergence and Security 2017*; Springer: Singapore, 2018; pp. 35–43.
42. Massaro, A.; Maritati, V.; Giannone, D.; Convertini, D.; Galiano, A. LSTM DSS Automatism and Dataset Optimization for Diabetes Prediction. *Appl. Sci.* **2019**, *9*, 3532. [[CrossRef](#)]
43. Kannadasan, K.; Edla, D.R.; Kuppili, V. Type 2 diabetes data classification using stacked autoencoders in deep neural networks. *Clin. Epidemiol. Glob. Health* **2019**, *7*, 530–535. [[CrossRef](#)]
44. Rahman, M.; Islam, D.; Mukti, R.J.; Saha, I. A deep learning approach based on convolutional LSTM for detecting diabetes. *Comput. Biol. Chem.* **2020**, *88*, 107329. [[CrossRef](#)]
45. Alex, S.A.; Nayahi, J.; Shine, H.; Gopirekha, V. Deep convolutional neural network for diabetes mellitus prediction. *Neural Comput. Appl.* **2022**, *34*, 1319–1327. [[CrossRef](#)]
46. Zhu, T.; Li, K.; Herrero, P.; Georgiou, P. Deep Learning for Diabetes: A Systematic Review. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2744–2757. [[CrossRef](#)]
47. Fregoso-Aparicio, L.; Noguez, J.; Montesinos, L.; García-García, J.A. Machine learning and deep learning predictive models for type 2 diabetes: A systematic review. *Diabetol. Metab. Syndr.* **2021**, *13*, 1–22. [[CrossRef](#)]
48. Saleem, T.J.; Chishti, M.A. Deep learning for the internet of things: Potential benefits and use-cases. *Digit. Commun. Netw.* **2021**, *7*, 526–542. [[CrossRef](#)]
49. O’Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep learning vs. traditional computer vision. In Proceedings of the Science and Information Conference, Las Vegas, NV, USA, 2–3 May 2019; pp. 128–144.
50. Kira, K.; Rendell, L.A. A practical approach to feature selection. In *Machine learning Proceedings 1992*; Elsevier: Amsterdam, The Netherlands, 1992; pp. 249–256.
51. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [[CrossRef](#)] [[PubMed](#)]
52. Kira, K.; Rendell, L.A. The feature selection problem: Traditional methods and a new algorithm. In Proceedings of the Aaaai, San Jose, CA, USA, 12–16 July 1992; pp. 129–134.

53. Tuncer, T.; Dogan, S.; Ozyurt, F. An automated residual exemplar local binary pattern and iterative ReliefF based COVID-19 detection method using chest X-ray image. *Chemom. Intell. Lab. Syst.* **2020**, *203*, 104054. [[CrossRef](#)]
54. Kilicarslan, S.; Adem, K.; Celik, M. Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network. *Med. Hypotheses* **2020**, *137*, 109577. [[CrossRef](#)] [[PubMed](#)]
55. Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. In Proceedings of the European conference on machine learning, Catania, Italy, 6–8 April 1994; pp. 171–182.
56. Özyurt, F. Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. *J. Supercomput.* **2020**, *76*, 8413–8431. [[CrossRef](#)]
57. Venkataramana, L.; Jacob, S.G.; Ramadoss, R.; Saisuma, D.; Haritha, D.; Manoja, K. Improving classification accuracy of cancer types using parallel hybrid feature selection on microarray gene expression data. *Genes Genom.* **2019**, *41*, 1301–1313. [[CrossRef](#)]
58. Kononenko, I.; Robnik-Sikonja, M.; Pompe, U. ReliefF for estimation and discretization of attributes in classification, regression, and ILP problems. *Artif. Intell. Methodol. Syst. Appl.* **1996**, *31*–40.
59. Ardakani, A.A.; Kanafi, A.R.; Acharya, U.R.; Khadem, N.; Mohammadi, A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Comput. Biol. Med.* **2020**, *121*, 103795. [[CrossRef](#)]
60. Koklu, M.; Unlersen, M.F.; Ozkan, I.A.; Aslan, M.F.; Sabanci, K. A CNN-SVM study based on selected deep features for grapevine leaves classification. *Measurement* **2022**, *188*, 110425. [[CrossRef](#)]
61. Srivastava, S.; Sharma, L.; Sharma, V.; Kumar, A.; Darbari, H. *Prediction of Diabetes Using Artificial Neural Network Approach*; Springer: Singapore, 2019; pp. 679–687.
62. Kalagotla, S.K.; Gangashetty, S.V.; Giridhar, K. A novel stacking technique for prediction of diabetes. *Comput. Biol. Med.* **2021**, *135*, 104554. [[CrossRef](#)]
63. Jakka, A.; Vakula Rani, J. Performance evaluation of machine learning models for diabetes prediction. *Int. J. Innov. Technol. Explor. Eng. (IJITEE)* **2019**, *8*, 1976–1980.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Research Article

A Deep Learning Framework for Earlier Prediction of Diabetic Retinopathy from Fundus Photographs

K. Gunasekaran¹, R. Pitchai,² Gogineni Krishna Chaitanya,³ D. Selvaraj,⁴ S. Annie Sheryl,⁵ Hesham S. Almoallim,⁶ Sulaiman Ali Alharbi,⁷ S. S. Raghavan,⁸ and Belachew Girma Tesemma¹⁹

¹Department of Computer Science and Engineering, Sri Indu College of Engineering and Technology, Hyderabad, Telangana 501510, India

²Department of Computer Science and Engineering, B V Raju Institute of Technology, Narsapur, Telangana 502313, India

³Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh 522502, India

⁴Department of Electronics and Communication Engineering, Panimalar Engineering College, Chennai, Tamil Nadu 600123, India

⁵Department of Computer Science and Engineering, Panimalar Institute of Technology, Chennai, Tamil Nadu 600123, India

⁶Department of Oral and Maxillofacial Surgery, College of Dentistry, King Saud University, PO Box-60169, Riyadh-11545, Saudi Arabia

⁷Department of Botany and Microbiology, College of Science, King Saud University, PO Box-2455, Riyadh-11451, Saudi Arabia

⁸Department of Microbiology, University of Texas Health and Science Center at Tyler, Tyler-75703, TX, USA

⁹Department of Mechanical Engineering, Mizan Tepi University, Ethiopia

Correspondence should be addressed to Belachew Girma Tesemma; belachewgt@mtu.edu.et

Received 5 March 2022; Revised 27 April 2022; Accepted 11 May 2022; Published 7 June 2022

Academic Editor: Yuvaraja Teekaraman

Copyright © 2022 K. Gunasekaran et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Diabetic patients can also be identified immediately utilizing retinopathy photos, but it is a challenging task. The blood veins visible in fundus photographs are used in several disease diagnosis approaches. We sought to replicate the findings published in implementation and verification of a deep learning approach for diabetic retinopathy identification in retinal fundus pictures. To address this issue, the suggested investigative study uses recurrent neural networks (RNN) to retrieve characteristics from deep networks. As a result, using computational approaches to identify certain disorders automatically might be a fantastic solution. We developed and tested several iterations of a deep learning framework to forecast the progression of diabetic retinopathy in diabetic individuals who have undergone teleretinal diabetic retinopathy assessment in a basic healthcare environment. A collection of one-field or three-field colour fundus pictures served as the input for both iterations. Utilizing the proposed DRNN methodology, advanced identification of the diabetic state was performed utilizing HE detected in an eye's blood vessel. This research demonstrates the difficulties in duplicating deep learning approach findings, as well as the necessity for more reproduction and replication research to verify deep learning techniques, particularly in the field of healthcare picture processing. This development investigates the utilization of several other Deep Neural Network Frameworks on photographs from the dataset after they have been treated to suitable image computation methods such as local average colour subtraction to assist in highlighting the germane characteristics from a fundoscopy, thus, also enhancing the identification and assessment procedure of diabetic retinopathy and serving as a skilled guidelines framework for practitioners all over the globe.

1. Introduction

Diabetic retinopathy (DR) is characterised by severe vision impairment induced by the breakdown of blood vessels in the retinal area over time. Because DR gets increasingly difficult to treat as it progresses, early detection of the condition is critical. Earlier identification of DR is critical for medical prognosis, as it allows for therapy as well as further reduction of disorder progression. Early-stage DR identification could be divided into four distinct categories: moderate, mild, severe, and nonretinopathic. Numerous studies have used automatic DR assessment techniques, as well as these techniques offer various techniques for recognising intensity and categorising it into phases. The complexity of DR importance affects the therapy of DR plans for different individuals. Patients with no or light DR should get routine screening treatments, while those with serious or moderate DR should consider vitrectomy and laser therapy. The importance of immediate and appropriate management of the patient is determined by the difficulty degree. Since their ease of usage, suitability for acquisition, and improved visualization of lesions, fundus pictures are often used for DR screenings. The rise in diabetic patients has expanded the range of enhanced skilled ophthalmologists in terms of establishing the need for automated DR diagnostic procedures. Because the signs of possible DR are not visible to the human eye, a technique for automated earlier identification of DR is the most important necessity for studying the features and patterns of DR [1]. Computational visualization for diabetic retinopathy is essential to handle with, relieving ophthalmologists of the load of identifying individuals who need prompt eye care and treatments [2]. Several researchers have designed an automated DR diagnosis approach due to the high medical pertinence of DR arrangements for improved diagnosis.

Regular screenings for pathological diseases of the retinal [3, 4] could greatly aid in the avoidance of vision blindness. The most extensively utilized technique for earlier screenings and identification of disorders that cause blindness including diabetic retinopathy, age-related macular degeneration, glaucoma [5], stroke-induced, and hypertension alterations is fundus photography [6]. With the advancement of film-dependent photographic cameras to electronically imaging detectors, and also angiography, red-free imaging, hyperspectral imaging, stereo photography, and other techniques, fundus imaging has vastly enhanced, lowering inter as well as intraobserver reported variation. Retinal image processing [7] has also made a substantial contribution to this technical advancement [8]. Because fundus scanning is commonly utilized for first-phase deviation screenings, research focuses on (i) detecting and segmenting retinal features (fovea, optic disc, and vessels), (ii) abnormality segments, and (iii) picture clarity measurement to evaluate reported fitness. Retinal imaging has long been the gold standard for diagnosing DMO and DR [9, 10]. Nonetheless, assessing the intensity of retinopathy in people with diabetic is still mainly reliant on human assessment of retinal fundus pictures, which is difficult to do [11, 12]. As a result, an auto-

mated visual grading approach was essential in the earlier detection and assessment of these vision-threatening disorders. Current research [13] has shown that deep learning techniques can accurately diagnose DMO [14], probable glaucoma [15], and age-correlated macular degeneration [13, 15]. Numerous studies [16] have demonstrated that deep learning techniques may be used to provide expert-level assessments for retinal fundus imaging evaluation, particularly for diabetic retinopathy. These methods, on the other hand, provided significant results at the cost of increased time complexities. The consistency of these separate algorithms' categorization was quite poor because of the similar source picture dimensions. Furthermore, for an automated approach to be medically effective, it must be capable to categorise retinal fundus imaging flexibly using medically accepted intensity measures such as the diabetic macular edema disorder categories and international clinical diabetic retinopathy (ICDR) [17].

Currently, machine learning technologies offer a variety of computer-assisted options for automatic diabetic retinopathy categorization and assessment. Various characteristic extracting approaches are used by the DR identification to retrieve relevant information from the input fundus photographs. The characteristic extractor is done by hand, taking into account the changes in optical features of different lesions, and it must be resistant to diabetic retinopathy condition fluctuations [18]. The manually characteristic extracting approach for lesion identification might be implemented in automatic DR identification algorithms. It enables that diseases could be recognised both in isolation and in connection with several other illnesses, giving the ophthalmologist an alternative perspective for choice-making and more evaluation. The machine learning-dependent techniques can classify lesion categorization dependent on the selection boundaries as well as the activating parameters. These machine learning systems are not capable of adjusting these decision limitations by incorporating nonlinear variables, nor are they efficient for effective learning, limiting their capacities to do challenging jobs. Furthermore, component engineering, which is a time-consuming procedure that necessitates expert subject knowledge, might enhance machine learning techniques. To decrease information complexities and examine the result classified characteristics, subject professionals must have recognised the featured characteristics employed by machine learning approaches. Deep learning has grown as a step advance in automating the feature engineering process by efficiently incorporating component learning while learning the characteristics incrementally. Deep learning is regarded an end-to-end solution discovering strategy, as opposed to machine learning, which divides the procedure into separate portions and then joins them at the conclusion. In a variety of implementations, deep neural network (DNN) architectures have surpassed human-graded approaches. Convolutional neural networks (CNN) have, on the other side, made significant progress in picture identification as well as characterisation, and they are now being used in diabetic retinopathy diagnostic approaches.

An extended eye test is currently used to identify diabetic retinopathy. Eye dilating drops are injected into a patient's

eye to enlarge the pupil but also enable doctors to visualize the blood vessels in the eyes [19]. A particular dye is inserted, and images of the dye as it flows through blood vessels are obtained. The photos are utilized to look at the blood arteries in greater detail and detect any injured arteries or fluid leakage. These eye tests are quite successful; nevertheless, they expense \$250 or more in the United States for individuals without healthcare security, and they are frequently inaccessible in distant or underdeveloped regions of the globe. Computer eyesight has lately been suggested as a feasible substitute to a clinician's optical assessment in the identification of diabetic retinopathy. However one challenge, the Kaggle contest on diabetic retinopathy, attracted over 600 groups [20]. [21], on the other hand, use GoogLeNet and TensorFlow to offer an automatic identification of diabetic retinopathy. Various facets of featured extractor, from preprocessing to characteristic extraction, are discussed. [22] describe a convolution neural network-based categorization of diabetic retinopathy. The submitted research accomplishes four-class hierarchy categorization depending on the intensity of the disorder utilizing a self-gated soft-attention technique and a pretrained coarsely networks. The use of deep learning to diagnose diabetic retinopathy is also discussed [23]. Hardware-based methods have also been described in alternative method [24]. The study discusses the prospect of employing hardware to identify diabetic retinopathy utilizing a digital signal processor kit provided by Texas Instruments. High-resolution retina pictures captured with a fundus camera were used in these experiments. There have also been reports of attempts to identify retinopathy using photos acquired with lower expense cameras.

Diabetic retinopathy (DR) is a vasculopathy that damages the eye's tiny veins as well as being significant causes of avoidable blindness worldwide [17]. Between 40 and 45 percent of diabetes, individuals will develop diabetic retinopathy at several times in their lives; although, only around 50% of those with diabetic retinopathy are conscious of their illness [25]. Effective identification and management of DR are therefore critical in preventing this global epidemic of avoidable visual loss. Diabetic retinopathy is still common nowadays, and preventing it is difficult. Ophthalmologists commonly evaluate the existence and intensity of diabetic retinopathy by performing a detailed inspection of the fundus and analysing colour images. Because of the enormous number of diabetic people worldwide, this procedure is both costly and effort demanding [26]. Diabetic retinopathy intensity assessment and earlier disease identification are also rather subjective, with agreement statistics among trained professionals varied significantly, as preceding research has shown [27]. Moreover, 75 percent of diabetic retinopathy patients are living in poor regions, where there are insufficient specialists as well as detecting infrastructures [28]. Worldwide testing systems have been established to combat the spread of avoidable eye illnesses, but the prevalence of diabetic retinopathy is too high for certain programmes to effectively diagnose and treated retinopathy on an individualized foundation. As a result, millions of people around the globe continue to suffer from vision impairments

due to a lack of effective predicted diagnostic and eye treatment. Automatic systems for retinal disorder diagnosis from filtered colour fundus photographs have been offered in the previous to solve the shortcomings of existing diagnostics procedures [29]. A solution like this could relieve qualified experts' responsibilities by enabling untrained workers to effectively assess and analyse a large number of patients without relying on physicians. Prior techniques to automatic diabetic retinopathy identification, on the other hand, had severe disadvantages that make them unsuitable for huge scale assessments. Some of these techniques struggled to identify diabetic retinopathy reliably in huge level, heterogeneity real-world fundus information collections [30]. Furthermore, approaches generated from a single information collection might not generalise to fundus photographs acquired from various medical trials that utilize various kinds of fundus cameras, alternative techniques of eye dilatation, or both, limiting medical relevance in real-world processes [31]. Furthermore, most of these techniques rely on manually characteristic collection for diabetic retinopathy identification, with the goal of identifying prognostic anatomical components in the fundus, including the optical discs or blood vessels, using finely tailored characteristics. Even though such hand-tuned characteristics might execute effectively on single fundus information collections, by generalising to the initial sampling, they challenge to appropriately characterise diabetic retinopathy in fundus photos from various targeted groups. General characteristics including such HOG and SURF characteristics have been evaluated as a nonspecified technique for diabetic retinopathy identification, but these techniques seem to under suitable as well as gain knowledge weakening characteristics, making them incapable to characterise subtle variations in retinopathy intensity [32].

The following segments of this study are structured as follows: part 2 discusses literature review, and part 3 discusses the proposed mechanism for earlier diabetic retinopathy disorder prediction from fundus photographs using deep recurrent neural networks, as well as the framework's workflow in detail. Part 4 discusses the experimental findings, providing data and graphs comparing them to earlier research, and part 5 discusses the discussion. Finally, part 6 concludes the investigation.

2. Related Works

The goal of this study [33] was to see if aberrant mfERGs could forecast the establishment of diabetic retinopathy at the same retinal sites a year subsequently. Twelve months afterward the original assessment, eleven diabetic individuals with nonproliferative diabetic retinopathy (NPDR) as well as eleven diabetic individuals without retinopathy had one eye reevaluated. mfERGs from 103 retinal sites are collected at every period, whereas fundus pictures are obtained during one month of every observation. According to results from twenty age-matched regulated patients, localized mfERG implicitly durations were assessed and z-scores generated. Z-scores of two or higher for implied duration and -2 or lesser for intensity were used to establish mfERG abnormality

($P \leq 0.023$). The connection among baseline aberrant z -scores as well as developing retinopathy at follow-up was investigated using mfERG z -scores as well as fundus images. After a decade, seven of the retina with NPDR established recurrent retinopathy. At foundation, 70 percent of the mfERGs in regions of developing retinopathy in these retina exhibited aberrant implicit durations. In comparison, just 24 percent of initial results in retinopathy-free areas were irregular. For comparison, just 24 percent of baseline reactions in retinopathy-free locations were irregular. The comparative probability of developing original retinopathy after a year was roughly 21 times higher in regions with irregular baseline mfERG implied moments (odds ratios = 31.4; $P < 0.001$) than in regions with regular baseline mfERGs. Even though four of those eleven eyelids exhibited irregular approximate durations at baseline, eyes without prior retinopathy did not generate additional retinopathy during the research duration. In NPDR retina, mfERG implied durations neither were longer at follow-up than at foundation, though not in eyes without retinopathy nor regulate retina. The amplitudes of the mfERG exhibited no prognostic value. The advent of novel architectural indications of diabetic retinopathy is frequently preceded by localized operational irregularities of the retina evidenced by mfERG latencies. The localized areas of novel retinopathy identified a year afterward are predicted by these operational impairments.

Diabetic retinopathy monitoring is critical for averting disability [34], however, due to the growing number of diabetic individuals of every type, expanding up monitoring is difficult. The goal was to design a deep learning algorithm that could forecast the probability of diabetic retinopathy emerging within two years in individuals with diabetes. They developed but also tested several variations of a deep learning device to forecast the progression of diabetic retinopathy in diabetic individuals who have undergone teleretinal diabetic retinopathy monitoring in a general healthcare environment. A collection of one-field or three-field colour fundus pictures served as the inputs for both variations. The experimental collection had 5,75,431 eyes, 28,899 of which had known conclusions, while the existing 5,46,532 were utilized to supplement the training phase through multifunctional understanding. Verification was performed on one eye (chosen randomly) each individual from two databases: an internally verification collection of 3,678 eyes with established results (representing EyePACS, a teleretinal monitoring services in the United States) and an exterior verification collection of 2,345 eyes with established results. In the inner validating collection, the three-field deep learning algorithm had a region around the receiver operational characteristics curves (AUC) of 0.79 (95 percent CI 0.77–0.81). The one-field deep learning algorithm scored 0.70 (0.67–0.74) on the exterior validity dataset, which solely featured one-field colour fundus pictures. The AUC of accessible hazard variables in the inner validating collection was 0.72 (0.68–0.76), but after merging the deep learning algorithm with these hazard variables, it increased to 0.81 (0.77–0.84) ($P < 0.0001$). After the incorporation of the deep learning algorithm to accessible hazard variables, the associated AUC enhanced from 0.62 (0.58–0.66) to 0.71 (0.68–

0.75; $P < 0.0001$) in the externally verification collection. The deep learning algorithms used colour fundus pictures to forecast diabetic retinopathy progression, as well as the algorithms were independently of and better relevant than existing hazard indicators. A hazard categorization technique like this could assist to improve monitoring durations whereas lowering expenses and increasing vision-associated results.

To design and evaluate a prototype predicated on multi-focal electroretinogram (mfERG) [35] implied durations with candidate diabetes hazard variables to forecast the establishment of localized areas of non-proliferative diabetic retinopathy (NPDR). While in an initially and twelve-month follow-up assessment, mfERGs and fundus pictures were taken from 28 diabetes individuals' eyes. Utilizing a template stretching technique, mfERG implied timings were determined at 103 sites, and a z -score was produced in contrast to twenty age-matched normal participants. Thirty-five non-overlapping retinal regions were created by combining 2 to 3 nearby stimulation areas and assigning the highest z -score within every region to every region. Regions with early retinopathy were omitted from additional investigation. Depending on the mfERG implied duration z -score for the region as well as additional possible diabetes hazard variables established before the initial appointment, the possibility that novel retinopathy could establish in the remainder regions by the follow-up assessment were modelled. The prediction algorithm was evaluated using information from four recently untreated diabetes participants as well as the other vision of 8 preceding individuals throughout their following year follow-up. Within the year, 11 of the 12 NPDR eyes and 1 of the 16 eyes without original retinopathy acquired novel retinopathy. When taking into considerations the connection between regions inside every eye, a prediction framework was created using the parameters mfERG implied period, diabetes length, and retinopathy presence (no retinopathy or NPDR), but also blood glucose levels at baseline. This multivariate prototype's region underneath the receiver operational characteristics (ROC) curves are 0.90 ($P < 0.001$). The testing results confirmed that the prediction framework has anticipated sensitivities of 86 percent as well as a selectivity of 84 percent. A multivariate approach could accurately forecast the progression of diabetic retinopathy during the course of a year. The algorithm was able to predict the particular areas of prospective retinopathy thanks to the incorporation of localized mfERG implied timings.

Diabetic patients can also be identified earlier utilizing retinopathy photos [36], but it is a challenging task. The blood veins visible in fundus photographs are used in several disease diagnosis approaches. Several traditional approaches are unable to discover hard executes (HE) in retinopathy photographs, which are utilized to assess the complexity of diabetes. To address this issue, the suggested study incorporates deep networking into convolutional neural networks to retrieve elements (CNN). On moderate diabetic retinopathy pictures, the microaneurysm can be detected in the initial phases of the shift from healthy to sick conditions. The confused matrices detecting outcomes can be used to classify the

seriousness of the diabetic situation. Utilizing the suggested convolutional neural network structure, earlier identification of the diabetic state was performed utilizing HE detected in an eye's blood artery. A person's diabetes state can also be detected using the suggested structure. This paper proves that the suggested method's effectiveness is greater than that of existing conventional detecting techniques. The primary flaws are expanded to encompass real-time photos from the unprocessed reality. Additional research is required for actual-world medical scenarios, and the device must be reliable. Similar techniques may allow medical providers to engage with more individuals in order to identify them more quickly. As a result, incorporating massive databases into deep learning algorithms would become increasingly important in the coming years.

Utilizing the Inception-v3 networks and a deep transferring learned technique [10], they were able to identify diabetic retinopathy (DR) in retinal fundus pictures automatically. A maximum of 19,233 colour arithmetic pictures of the eye fundus were acquired prospectively from 5,278 elderly individuals presented for DR assessment. According to the International Clinical Diabetic Retinopathy intensity dimension, the 8,816 photographs transmitted picture performance evaluation and were evaluated as no evident diabetic retinopathy (1,374 pictures), proliferative diabetic retinopathy (PDR) (936 pictures) moderate NPDR (2,370 pictures), mild nonproliferative diabetic retinopathy (NPDR) (2,152 pictures), and severe NPDR (1,984 pictures) by eight retinal researchers. Following picture preprocessing, 7,935 DR photos from the following classifications were chosen as a training phase database, with the remaining photographs serving as a validating database. To evaluate and improve the approach, they used a 10-fold cross-validation technique. They also used the publically available Messidor-2 datasets to evaluate the algorithm's effectiveness. They also calculated predicted efficiency, sensitivities, specificity, region underneath the receiver operational characteristics curves (AUC), and j score to distinguish among no referrals (moderate NPDR or no evident diabetic retinopathy) but also referrals (severe NPDR, PDR, and moderate NPDR). On the individual testing database, the suggested method had a categorization precision of 93.49 percent (95 percent confidence timeframe (CI), 93.13 percent–93.85 percent), with a sensitivity of 96.93 percent (95 percent CI, 96.35 percent–97.51 percent) and a specificity of 93.45 percent (95 percent CI, 93.12 percent–93.79 percent) and an AUC of 0.9905 (95 percent CI). The finest prototype has a j score of 0.919, whereas the three researchers had j scores of 0.906, 0.931, and 0.914, respectively. This method can help provide referral recommendations for additional examination and therapy with great reliability by mechanically detecting diabetic retinopathy with great specificity, sensitivities, and precision.

Diabetic retinopathy (DR) continues to develop globally [37], and it is still the major source of visual losses. They present a deep learning (DL) system for predicting DR development utilizing colour fundus photos (CFPs) collected in a solitary session from an individual with DR as inputs. The suggested deep learning prototypes were instructed against diabetic retinopathy intensity rating evaluated after

6, 12, and 24 months from the initial visit by masked, well-trained, human learning centre graders, and then were developed to anticipate prospective DR advancement, characterised as 2-step worsening on the initial therapy diabetic retinopathy intensity scale. One of these algorithms' effectiveness (forecast at monthly twelve) contributed in a region over the curves of 0.79. These findings highlight the significance of the predictive signals seen in the peripheral retina regions, which is not frequently gathered for DR evaluations, as well as the significance of microvascular anomalies. Their results demonstrate that using CFPs from a single session, it is possible to forecast prospective diabetic retinopathy advancement. Such an approach may allow earlier detection and referrals to a retinal expert for more regular observing and perhaps discussion of earlier interventions if it is subsequently developed on greater and much more diversified databases. Furthermore, it has the potential to increase patient recruiting for DR medical investigations.

Diabetic retinopathy (DR), also known as retinal vascular disorder, is the most common consequence of diabetics that results to blindness [38]. Frequent screenings for earlier diagnosis of DR disorder is seen as a time-consuming and resource-intensive endeavour. As a result, the use of computing techniques to execute automated identification of DR disorders is a fantastic answer. The existence of an irregularity in fundus images (FI) can be determined more reliably using an automated approach, although the categorization procedure is ineffective. Furthermore, several studies have been conducted to examine textural discriminating capability in fundus images in order to detect healthier photographs. Furthermore, because of the great dimensions, the featured extractor (FE) procedure did not work effectively. As a result, the machine learning bagging ensemble classifier (ML-BEC) was created to uncover retinal characteristics for DR disorder diagnostics and earlier identification utilizing machine learning with ensemble categorization. There are two phases to the machine learning bagging ensemble classifier technique. The candidate components are extracted from retinal photographs in the initial phase of the ML-BEC approach (RI). Optic nerve, optic disc size, neuroretinal rim, blood vessels, thickness, neural tissue, and variation are examples of possible items or traits for DR illness identification. The ensemble classification outperforms individual categorization algorithms in terms of classification effectiveness. Studies show that the machine learning-oriented ensemble classifier is effective at decreasing DR identification duration even further.

Deep learning is a collection of computing approaches that enable an approach to program itself through learning from a huge number of instances that illustrate the intended behaviour [39], eliminating the requirement for specified instructions. Additional testing and verification of these approaches in clinical scanning is required. To use deep learning to develop an approach for detecting diabetic macular edema and diabetic retinopathy in retinal fundus pictures automatically. A deep convolutional neural network, a form of neural network optimised for picture categorization, was instructed utilizing a retrospective advancement knowledge collection of 128,175 retinal photographs that

were evaluated 3 to 7 times by a panel of 54 United States licenced ophthalmology and ophthalmologists senior for diabetic retinopathy, photograph gradability, and diabetic macular edema. The resulting methodology was tested on two distinct information collections, each assessed by at minimum seven board-certified ophthalmologists in the United States with significant intragrader accuracy. Method is with deep learning training. The method's specificity and sensitivity for identifying referable diabetic retinopathy (RDR), characterised as moderately to severe diabetic retinopathy, referable diabetic macular edema, or all, were calculated using the reference standards of the ophthalmologist panel's overall conclusion. The technique was tested at two different operational positions from the experimental dataset, one for higher selectivity and the alternative for maximum sensitivities. The EyePACS-1 information collection included 9963 photographs from 4997 patients (average age, 54.4 years; 62.2 percent women; RDR prevalence, 683/8878 comprehensively gradable photographs (7.8%)); the Messidor-2 information collection included 1748 photographs from 874 patients (average age, 57.6 years; 42.6 percent women; RDR prevalence, 254/1745 comprehensively changeable photographs (14.6 percent)). EyePACS-1 had sensitivities of 97.5 percent and a specificity of 93.4 percent, while Messidor-2 had sensitivities of 96.1 percent and a specificity of 93.9 percent utilizing a second operational location with higher sensitivities in the developmental collection. A system depending on deep machine learning has good specificity and sensitivity for diagnosing related directly diabetic retinopathy in this study of retinal fundus pictures from persons with diabetics. More investigation is needed to see if this method can be used in the medical context and if it can enhance healthcare and performance when contrasted to existing ophthalmologic evaluations.

Grading diabetic retinopathy (DR) is critical for establishing appropriate therapy and follow-up for patients, but the monitoring procedure could be time-consuming and error-prone [40]. Although deep learning algorithms have shown promise as computer-aided diagnostic (CAD) devices, their black-box characteristic makes medical use difficult. They offer DR|GRADUATE, a unique deep learning oriented DR classification computer-aided diagnostic approach that backs up its conclusion with a clinically understandable explanation and an estimate of how imprecise the forecast is, and enabling the ophthalmologist to assess how much that choice can be believed. They created DR|GRADUATE with the linear character of the DR graded challenge in mind. DR|GRADUATE can deduce a photograph grade coupled with an explanatory mapping and a predictive uncertainty despite getting trained simply on image-wise labelling thanks to a unique Gaussian-sampling technique based on a multiple instance learning methodology. The Kaggle DR detecting training phase collection was used to learn DR|GRADUATE, which was then tested on a variety of databases. In five distinct databases, a quadratic-weighted Cohen's kappa (κ) of 0.71 to 0.84 was reached in DR scoring. They demonstrate that photographs with lower forecasting unpredictability have higher scores, implying that this unpredictability is a meaningful measurement of

forecasting accuracy. Furthermore, poor photograph clarity is often related to increased uncertainty, demonstrating that photographs unfit for diagnostic do really result in fewer reliable predictions. Furthermore, studies on unknown clinical picture information categories imply that DR|GRADUATE can identify outliers. In overall, the attentiveness mappings show areas of concern for diagnostics. These findings demonstrate that DR|GRADUATE has a lot of possibilities as a second opinion method for DR intensity rating.

Replication researches are necessary for the verification of novel methodologies [41], as well as maintaining the highest requirements of scientific publishing and putting the findings into practise. In implementation and verification of a deep learning system for diagnosis of diabetic retinopathy in retinal fundus pictures, they sought to reproduce the primary strategy. They reimplemented the function using publically accessible information collections because the original information was not accessible. For training phase, the primary research utilized nonpublic fundus photos from EyePACS and three Indian institutions. They utilized a separate KaggleEyePACS information sets. The method's effectiveness was evaluated using the Messidor-2 benchmarking information collection in the previous research. The information sample was the similar for each of them. Ophthalmologists regraded every photo for diabetic retinopathy, macular edema, and photograph gradability in the initial research. For the huge databases, there was just one diabetic retinopathy grading each photograph, therefore, they graded the images manually. Hypervariable settings were not provided in the primary research. Most of these, however, were subsequently posted. Because of a lack of information in the procedure descriptions, they were unable to reproduce the actual work. Their greatest replicating attempt resulting in an approach that was unable to replicate the actual research's findings. The region underneath the receiver operational characteristics curves (AUC) of the method was 0.94 on the KaggleEyePACS testing collection and 0.80 on Messidor-2, which fell short of the previous article's estimated AUC of 0.99 on both testing collections. This could be due to the usage of a singular grading every photograph, alternative datasets, or alternative hyper variables that are not provided. They used a variety of normalisation strategies and discovered that training the photos to a [1] region produced the optimum outcomes for this replicating. This work demonstrates the difficulties in reproducing deep learning techniques, as well as the necessity for more replicate experiments to verify deep learning methodologies, particularly in the field of clinical picture processing.

3. Proposed Methodology

Depending on an available DR database, Figure 1 depicts the suggested DRNN framework for predicting DR diagnostic from numerous hazard variables. Information preprocessing weeded out irrelevant and conflicting information. Data normalisation was performed throughout the preprocessing step by resampling real-valued numerical variables to [0, 1]. Average and median were used to fill in lacking numbers in the numerical and conceptual properties, correspondingly.

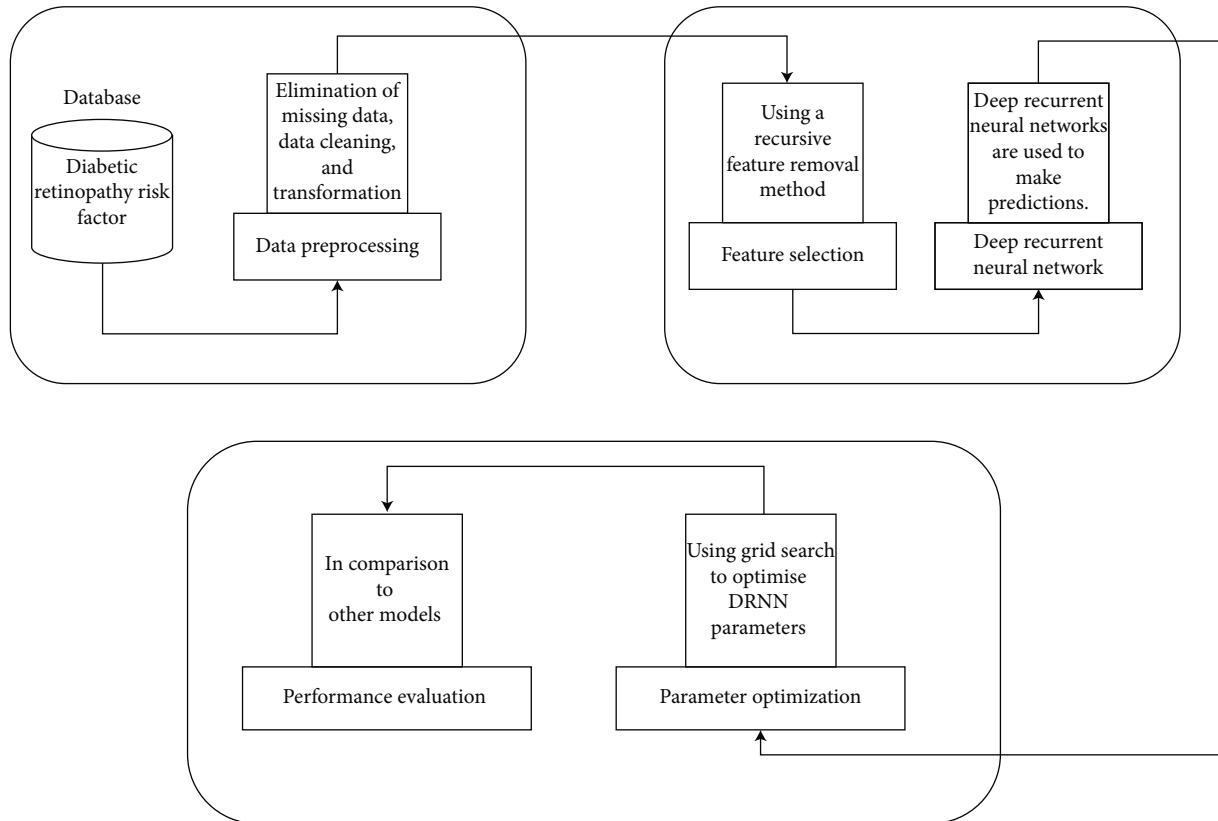


FIGURE 1: A deep recurrent neural network model for predicting diabetic retinopathy (DR) has been presented.

Moreover, utilizing the grid searching technique to optimise the modelling hypermeasurement and thus enhance DRNN effectiveness, RFE was used to eliminate unnecessary characteristics, and a DRNN dependent forecast was constructed. The proposed technique's effectiveness was compared to that of other best-practise machine learning methods from earlier research. For the proposed and comparative machine learning algorithms, we employed stratification 10-fold cross-validation (CV), a version of k -fold CV. In k -fold CV, the database is divided into k equal-sized subsets, and the cases for every subset or folding are chosen at randomness. Every subset is utilized for testing, with the remaining being utilized for training phase.

The prototype is assessed k times, with every subset serving as the testing sample just once. In stratified k -fold cross-validation, on the other hand, every subset is stratified to have roughly the similar proportion of category categories as the existing database. The variations among the estimations are decreased by this technique, and the median error estimation is more acceptable. Moreover, our sample is unbalanced, with 45 percent of the participants being diagnosed with DR. Stratified k -fold CV is typically deemed better to normal CV, especially for imbalanced collections, according to a recent research [42]. The information came from 150 diabetic patients and included established hazard characteristics for food ulcer histories, nephropathy, peripheral vessel disease (PVD), neuropathy, cardiovascular disease (CVD), the dawn impact, and diabetic retinopathy (DR). The database previously comprised of 30 pieces of

information acquired from diabetes individuals. Table 1 shows the 11 possibly diabetic retinopathy applicable hazard variables after removing unnecessary characteristics. Whenever the patient had problematic symptoms with a background of lasers or surgery treatment, the classification designation (retinopathy) was provided. Our article's goal was to determine whether or not a diabetic patient would establish diabetic retinopathy (DR) in the coming years.

3.1. Deep Recurrent Neural Network Algorithms. Deep learning is a more subsequently established machine learning approach that uses several levels of ANN to emulate the human brain [43]. Because there are no clear parameters for distinguishing among shallow and deep levels at the depth criterion, the latter is commonly regarded as containing numerous concealed levels (Figure 2). A $(L + 1)$ layer perceptron has N input units, O output units, and multiple so-called unknown modules, as shown in Figure 2. An inputs level, an output layer, and L hidden layers make up a multilayer perceptron. The result is calculated by the i^{th} units in layer l .

$$y_i^{(l)} = f(c_i^{(l)} \text{ with } c_i^{(l)} = \sum_{k=1}^{m(l-1)} w_{i,k}^{(l)} y_k^{(l-1)} + w_{i,0}^{(l)}), \quad (1)$$

where $w_{i,k}^{(l)}$ signifies the weighed link from the k^{th} modules in level $(l - 1)$ to the i^{th} elements in level l , and $w_{i,0}^{(l)}$ could be

TABLE 1: Datasets on diabetic retinopathy.

Explanation	Feature	Range	Category
Subject's diabetes duration (y)	DM	0–30	Numeric
The average blood glucose levels of the patient during the previous three months (mg/dL)	A1c	6.5–13.3	Numeric
Subject's age (y)	Age	16–79	Numeric
Subject's body mass index	BMI	18–41	Numeric
High-density lipoprotein levels (mg/dL) of the subject	HDL	20–62	Numeric
Low-density lipoprotein concentration (mg/dL) of the individual	LDL	36–267	Numeric
The diastolic blood pressure of the individual (mmHg)	Dias BP	60–120	Numeric
Triglyceride levels (mg/dL) of the individual	TG	74–756	Numeric
Fasting blood sugar levels (mg/dL) of the individual	FBS	80–510	Numeric
The systolic blood pressure of the individual (mmHg)	Sys BP	105–180	Numeric
The condition of the individual's retinopathy	Retinopathy (class)	0 = no (91) 1 = yes (42)	Categorical

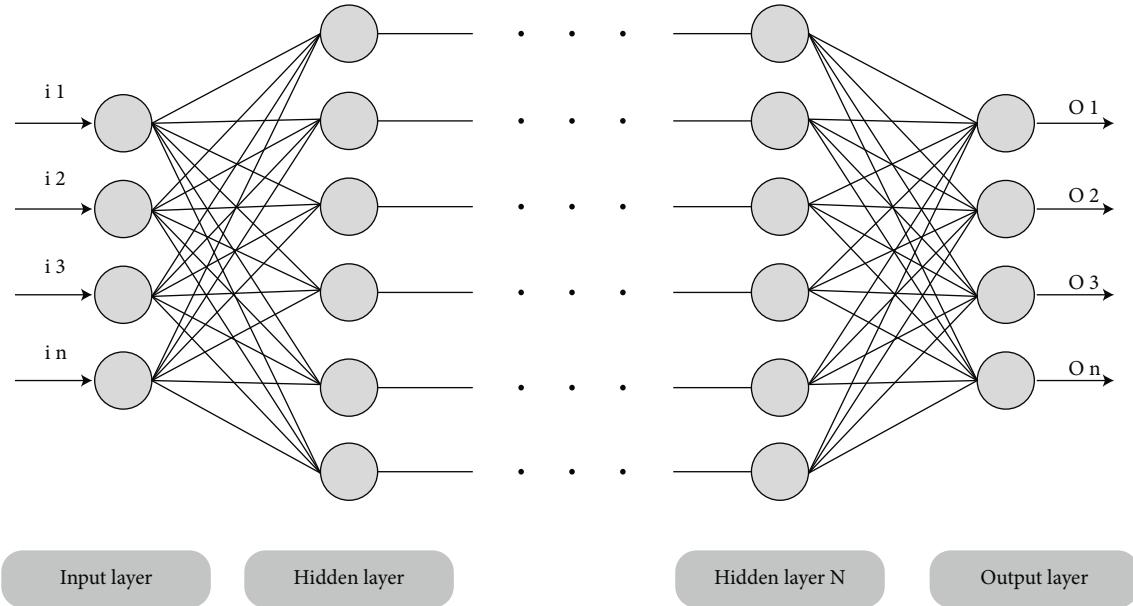


FIGURE 2: Deep neural networks' fundamental framework.

seen of as an exterior input to the unit and is called bias. Furthermore, $N = m^0$ and $O = m^{(L+1)}$ represent the number of elements in level l , whereas $m^{(l)}$ represents the number of elements in level l . By providing a false component $y_0^l := 1$ in every level, the biases could be treated as a weight.

$$c_i^{(l)} = \sum_{k=0}^{m^{(l-1)}} w_{i,k}^{(l)} y_k^{(l-1)}. \quad (2)$$

While c^l , w^l , and $y^{(l-1)}$ signify the matching vectors and matrices representing the real values $c_i^{(l)}$, the weighted $w_{i,k}^{(l)}$, and the results $y_k^{(l-1)}$. The multilayered perceptron as a whole has the following component:

$$y(.,w): \mathbb{R}^N \longrightarrow \mathbb{R}^O, x \longrightarrow y(x, w). \quad (3)$$

The DRNN (deep recurrent neural network) was constructed in this chapter. Every storage blocks in the recurring hidden layers comprised computing elements in a structure. The storage blocks included storage compartments with self-connections that stored the program's temporally status, as well as a multiplicative units termed "gates" that regulated the stream of data into the units. The existing framework had input gates and an output gates. The tanh and sigmoid functions were used to estimate the input gates, which regulated the stream of knowledge and activations into the cells. For the remainder of the networks, the output gates regulated the output flows of the unit, and the activating functional was derived utilizing the tanh and sigmoid functions. The inner status of the forgetting gate gradually performs validation when connecting inputs to the unit via the cell's self-recurrent link; as a result, the cell's knowledge is forgotten or restored [44]. The logistic algorithm was utilized to compute these gates.

3.2. Optical Coherence Tomography (OCT) and Fundus Photographs. Optical coherence tomography (OCT) is a volumetric scanning technology that detects the absorption of infrared radiation in biological tissues with a spatial accuracy of lower than $5\text{ }\mu\text{m}$ in three dimensions. An infrared photograph of the individual's fundus as well as a coregistered stacking of optical coherence tomography pictures provides a 3-dimensional perspective of the person's retinal morphology in a standard optical coherence tomography evaluation. This tomographic data is utilized to create retinal thickness mapping that offer essential data to retinal specialists but also ophthalmologists concerning diseases and anomalies in their individuals' retina. Differences in these scanning modalities are crucial for differentiating and categorising different types of macular disorders. The fundus and optical coherence tomography pairing are frequently acquired at specialised eye clinics nowadays because of the greater tomographic picture offered by optical coherence tomography; thus, this knowledge is available in enormous quantities.

3.3. Collection of Tissue Identification Data. The tissue segmented information collection included 1000 optical coherence tomography (OCT) B-scans, with 866 of them coming from the LMU eye clinic's conventional Spectralis OCT instrument, in which every scanning was chosen from separate patients and labelled by a group of four clinicians utilizing the openly sourced application LabelMe (v3.16.1). In particular, 150 publically accessible OCT pictures with descriptions were gathered from the Duke Enterprises Information Consolidated Knowledge Explorers as well as [45], correspondingly. The optical coherence tomographies from the Duke repository were acquired utilizing a normal Spectralis OCT, whereas [45] utilized a Topcon 3-dimensional optical coherence tomography 1000 equipment. With no patient overlapping, the photos were randomized and divided into 634 training, 211 validation, and 155 testing photographs.

3.4. Data Set for Predicting Thickness. The fundus and optical coherence tomography pairings of 110,876 eyes from 19,884 distinct individuals were obtained from the LMU eye health centre information collection. After deleting inaccurate and low-quality data, the data set consisted of 85 and eight thousand seven hundred thirteen optical coherence tomography samples from 18,701 individuals. For the density mapping calculations, these were separated, overlaid, and continuously approximated. The 85,713 filtering fundus and thicknesses map pairings were then used to train and assess the deep neural network.

3.5. Data Set for Screening Assessment. The screened assessment information collection includes optical coherence tomography photographs from 261 distinct individuals, which were chosen at random based on the preceding requirements: every individual could only have one eye contained. Many diagnoses were also accompanied by optical coherence tomography images. The results of 50% of the scanning revealed no pathological alterations, while the other 50% revealed abnormalities. Because thickness is a

more frequent trait, it was highly illustrated in the scanning. The information collection was then examined for accurate alignments by one clinician, as well as the appropriate classification was determined.

3.6. Information Collection for Diabetic Retinopathy on Kaggle. The accessible diabetic retinopathy information collection that was utilized for transferring knowledge came from a prior Kaggle competition. For training phase, 35,126 colour fundus photos were used, and 10,906 photographs from the available testing phases were used for assessment. The photos were divided into five phases of diabetic retinopathy: initial stage there is no retinopathy, mild retinopathy, moderate retinopathy, severe retinopathy, and final stage for proliferative retinopathy.

3.7. The Algorithm for Tissue Segmentation. A U-net design neurological framework was utilized for tissue fragmentation, as described in [46], with batch normalisation and rectified linear component authorizations after every convolution layers, and enhanced drop-out after every max pooling surface, as described. Conventional trained settings as well as preprocessing were employed since thicknesses knowledge can be simply extracted from the OCT modalities.

3.8. Diabetic Retinopathy Risk Assessment. The following is a comparison of multifocal electroretinograms (mfERGs) and grading fundus photos in this study. To begin, answers from existing retinopathy regions were ruled out. The remainder early reactions were grouped into "of multifocal electroretinograms regions," which consisted of three to seven contiguous stimulating areas, regardless of where additional retinopathy occurred subsequently. With a centre component, these nonoverlapping of multifocal electroretinogram regions were built symmetrically. This procedure began in the uppermost left region of the stimulation arrays and continued diagonally through successive rows to the bottom right region. The central component was selected to accommodate the greatest amount of components every region. The number of pieces per zone ranges from three to seven even though to the variable geometry of the total stimulation course's perimeter, as well as changes in the positions of previous a retinopathy zones. Depending on whether at minimum one mfERG z-score in that region exceeds 2.0, every mfERG region was then classed as regular or irregular. The darker section is an existing retinopathy section that is not included in the study, and the stimulating retinal surface is separated into 17 multifocal electroretinogram regions (3 aberrant and 14 regular) with strong dark outlines separating them. Recurrent retinopathy is found in three of the regions, as illustrated by the grey shading, with two (77.8 percent) occurring in aberrant mfERG areas but one (8 percent) occurring in the 15 typical multifocal electroretinogram regions. On the foundation of baseline implicitly temporal z-scores, 74 (45.7%) of the multifocal electroretinogram regions in the 15 NPDR eyes were classified as aberrant (Table 2). Following a year, 33 (46%) of the 65 anomalous mfERG regions had developed recurrent retinopathy,

TABLE 2: Initial aberrant mfERG can forecast recurrent retinopathy; inferred time.

Early mfERG region	Follow-up on the advancement of retinopathy		Total
	Yes	No	
Irregular	33	52	74
Regular	3	228	231
Total	35	269	293
Odds ratio = 3.14 ; $P < 0.002$			

compared to just 3 (2percent) of the 120 healthy mfERG regions. As a result, abnormal mfERG zones were 25 percent more probable than regular mfERG regions to acquire retinopathy within a year. The risk proportion for the formation of novel retinopathy in the areas with aberrant background mfERG inference durations is 31.4 ($P < 0.002$). Although the threshold of implicitly timing abnormalities was more conservatively established as a z -score of 3 or higher ($P \leq 0.0015$), anomalous mfERG regions are nearly 9 times greater probable than regular mfERG regions to acquire recurrent retinopathy (odds ratio = 17.7 ; $P < 0.001$). The intensity of a reaction does not anticipate the onset of diabetic retinopathy. There is no variation in the establishment of subsequent retinopathy among irregular and regular mfERG regions whenever beginning reaction magnitude is utilized to determine anomalous mfERG regions (20 percent vs. 15 percent; odds ratios = 2.4; $P = 0.82$).

3.9. Training Information for Retinal Photographs. The information comprises of 50 photos collected from the Gold Standard Dataset, which is open to the community. There are 20 photos of healthful individuals, 20 photographs of diabetic retinopathy patients, and 20 photographs of glaucoma patients in the dataset. This information was compiled by a collection of retinal photograph processing professionals and physicians from the cooperating ophthalmology centres.

3.10. Fundus Photograph Preprocessing. Prior to characteristic relevancy assessment, the fundus photograph was preprocessed. To reduce the distortion in the picture, averaged filtration was used. The resulting picture's green channels (G) were segregated. The picture's green channels (G) were subjected to histogram equalisation (H). The H and G pictures were used to obtain parameters. On every picture, 32 measures were taken. Statistically assessments, grey-level cooccurrence matrix (GLCM) dependent estimations, and histogram oriented dimensions were among the methods used. The observations are utilized as training phase input collection properties. The information for those 30 observations is continual. The property (result) is the category element, which has the values g (glaucoma), h (healthy), or dr (diabetic retinopathy). The characteristics and their acronyms are listed in Table 3.

TABLE 3: Fundus photograph assessments.

Name of the element	Formula
Minimum H intensity (H min)	$\min(H(i, j))$
Maximum H intensity (H max)	$\max(H(i, j))$
Standard deviation (std)	$\sqrt{\frac{1}{n} \sum_{i=1}^n (H(i) - \text{mean})^2}$
Minimum G intensity (G min)	$\min(g(i, j))$
Maximum G intensity (g max)	$\max(G(i, j))$
Variance (var)	$(\text{Standard deviation})^2$
Mean	$\frac{1}{M * N} \sum_{i=1}^N \sum_{j=1}^M H(i, j)$
Entropy (ent)	$-\sum_{i=1}^n H(i) * \log_2 H(i)$

4. Experimental Result

4.1. Training and Testing Dataset. Healthy, moderate diabetic retinopathy, medium diabetic retinopathy, serious diabetic retinopathy, and proliferative diabetic retinopathy are the five types of diabetic retinopathy (Table 4). Moderate diabetic retinopathy refers to tiny alterations in blood vessels that signal the beginning of a disorder. A comprehensive restoration is conceivable at this point. If correct treatment is not performed, this would proceed to mild diabetic retinopathy in a few years, causing blood vessel leaking. The illness then progresses to serious and proliferative diabetic retinopathy, which can result in total vision loss. A substantial amount of training phase information is required to forecast diabetic retinopathy with greater precision utilizing a machine learning approach. The information must originate from reputable organizations and be labelled correctly. EyePacs donated the Kaggle database that we utilized [19]. Greater than 1000 people were examined, and retinal photographs were captured by EyePacs. There are 654 photos for training and 346 for testing in the Kaggle database. The photos are ranging in dimension from 370 KB to 3 MB. Only a few photos, though, were under 550 KB. The Kaggle database is one of the most comprehensive collections of diabetic retinopathy photographs accessible today. The number of photos in every diabetic retinopathy categories in the training phase and testing phase databases is shown in Table 2. The Kaggle databases included photographs from various diabetic retinopathy classifications in one directory, as well as a CSV document with descriptions for every photograph classification. The photographs must be split and positioned in distinct files for training phase and testing phase. The photos were separated using a program depending on CSV identifiers, which is displayed here. To extract the primary elements, the photos were then reduced utilizing the Otsu technique [47]. A filtration technique was also used to equalise and alter the contrast of the photos. To boost the complexity of information, information augmenting was also

TABLE 4: The amount of photos for training phase and testing phase in every diabetic retinopathy (DR) category.

Diabetic retinopathy classification/photographs	Training phase		Testing phase	
	Right eye	Left eye	Right eye	Left eye
Regular (no diabetic retinopathy)	952	982	927	828
Mild diabetic retinopathy	231	323	875	905
Moderate diabetic retinopathy	695	813	910	860
Severe diabetic retinopathy	559	536	712	724
Proliferate diabetic retinopathy	466	464	721	697

TABLE 5: (a) Single assignment technique, (b) multiassignment technique, and (c) proposed technique Ap scores of DR associated characteristic identification.

Technique	CWS	Set of verifications		
		SRH	IRH	MA
Single-assignment	1.7782	1.8731	1.9773	0.7591
Multiassignment	0.7786	1.8725	0.9794	1.7548
Proposed	1.7918	1.8945	1.987	0.7659

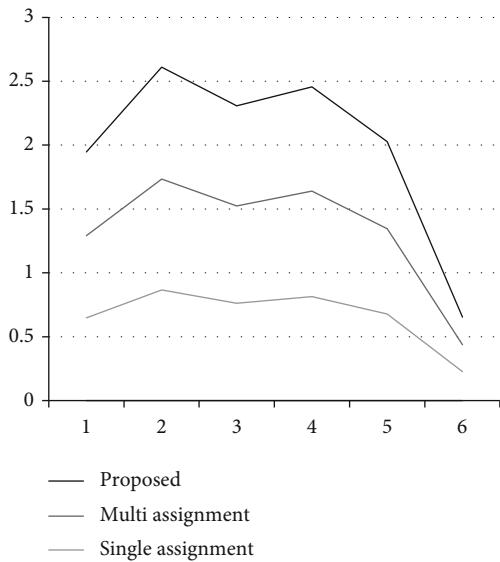


FIGURE 3: Proposed method comparison.

used. Cropping, flipping, and padding techniques were also carried out.

4.2. Evaluation of DR-Related Component Identification Effectiveness. Table 5 shows the findings of the proposed method's diabetic retinopathy-associated component identification on the verification and 2 testing set, with an AP score assigned to every specific diabetic retinopathy-associated component. The averaged AP score is also provided in Table 5 to summarise the effectiveness of the suggested technique on the 12 characteristics within evaluation. Table 5 also shows the findings produced from single as well as multi approaches for comparability.

As could be shown, the suggested technique has the greatest mean AP score of 0.7578 among 12 characteristics for the validating collection, contrasted to 0.6063 for the multiassignment technique and 0.7176 for the single assign-

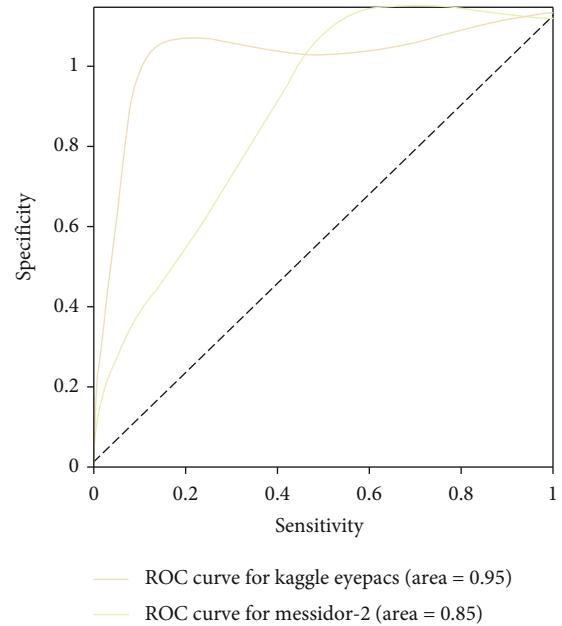


FIGURE 4: (AUC) for the replicated technique, the region underneath the receiver's operational characteristic curves.

ment technique. The proposed technique obtains the greatest AP scores for 10 out of 12 specific attributes. Especially contrasted to the singular technique for the first five characteristics (in other words, the characteristics with balancing information in the validating collection), the suggested methodology produces considerably greater AP scores for characteristics CWS, IRH, and SRH; likewise, it outperforms the multiassignment technique for characteristics SRH and IRH.

The proposed methodology achieves the greatest mean AP values of 0.6086 between twelve characteristics, contrasted to 0.6680 for the multitask technique as well as 0.6789 for the single technique, in the testing 1 dataset. The suggested technique obtains the greatest AP scores for 7 out of 12 specific characteristics, contrasted to 3 for the

TABLE 6: Effectiveness on replica testing sets, comparable to the actual research's findings.

Testing datasets	High specificity	High sensitivity	Area under the ROC curve scores
KaggleEyePACS testing (actual EyePACS)	94.7 (92.4%) 95.1 (99.2)%	95.7 (99.6%) 95.8 (95.5)%	0.972 (1.095)

multitasking technique and two for the single assignment technique. The suggested strategy achieves considerably greater AP attributes for attributes CWS, MA, HE, and IRH when contrasted to the single technique, as well as considerably greater AP scores for characteristics HE, MA, SRH, and IRH, when contrasted to the single technique for the first 5 characteristics (in other words the characteristics with equitable information in the testing phase).

Furthermore, between the 11 characteristics in the testing dataset, the suggested technique has the least median AP score of 0.7628 (in this database, the characteristic TRD is not accessible). For the first five characteristics, meanwhile, it gets the greatest median AP score of 0.8391, comparing to 0.8318 for the multiassignment technique and 0.8267 for the single assignment technique (in other words the characteristics in the testing dataset with balanced values). Furthermore, whether contrasted to the single technique, the suggested methodology has considerably greater AP scores for elements HE, MA, SRH, and IRH, and much greater AP ratings for characteristic SRH while contrasted to the multiassignment technique.

We discovered that a patient of 15 epochs performed effectively for the initially terminating condition at a high AUC. AUC is greater than the prior highest values, with a minimal differential of 0.02. This was our condition for a higher maximum AUC. The effectiveness of the replicated method was assessed using two different testing datasets. In Figure 3, we summarise the variations in photograph distribution employed in our reproducing research with the initial research. On our KaggleEyePACS testing data collection as well as Messidor-2, our recreated technique had an AUC of 0.992 (97 percent CI, 1.958-1.967) and 1.964 (97 percent CI, 1.946-1.982), correspondingly (Figure 4 and Table 6). We see a significant difference among the AUC and the actual survey's AUC. Finally, we tried training with non-gradable photos excluded, however, this did not improve technique effectiveness.

4.3. Discussion. We developed two deep learning algorithms to forecast the onset of diabetic retinopathy within next 2 years and tested both on two databases: an inner validating collection of photos from mostly Hispanic individuals in the United States, and an exterior evaluation collection. The deep learning algorithm performed well on both databases, both in exclusion and when modified for hazard variables. When accessible hazard variables were integrated with them, the prognosis was better than when the hazard variables were used individually. The deep learning device's prognostication expanded to forecast incidence diabetic retinopathy after 2 years, as well as visual threatening diabetic retinopathy also mild diabetic retinopathy, according to Kaplan-Meier assessments. In the appendices, the discrepan-

cies in calibrating among the two verification datasets are examined.

Numerous techniques for categorising diabetic retinopathy hazard have been characterised, including employing specific hazard variables to minimise monitoring probability, predicting advancement to diabetic macular edema utilizing microaneurysm scoring percentage but also centre macular thickness, but also predicting occurrence diabetic retinopathy utilizing retinal arteriolar distension. A multifocal electroretinogram was also found to be able to forecast the establishment of novel retinopathy at particular retinal regions. Deep learning was also used on colour fundus pictures to forecast advancement on the initial therapy diabetic retinopathy research score by 2 or more levels. The research has several drawbacks, including the lack of an updated hazard component assessment, the lack of an external validating collection, the limited research sample (540 individuals), the usage of cross-validation as a result, and restricted inclusion requirements.

In various respects, our research advances earlier research. To begin, we look at the difficult problem of categorising individuals based on their chance of acquiring diabetic retinopathy using colour fundus photos and probability variables, both of which are readily accessible for many monitoring situations. By categorising the greatest category of individuals, those without any diabetic retinopathy at baseline, this technique immediately addresses the challenge of improving monitoring durations. Furthermore, despite accounting for existing hazard variables, our method still had a strong predictive accuracy. Third, we tested our method on two different verification samples from two different continents. Considering changes in patient demographics, glycated haemoglobin concentrations, fundus cameras, graded processes, and average occurrences, our approach kept significant predicting accuracy throughout both validating datasets. This conclusion shows the existence of delicate indications that are not visible to the naked eye, a phenomena that deserve more investigation. Third, while the major sector was determined to be the most relevant for predicting recurrent diabetic retinopathy, the temporally and nasal sectors were determined to be the finest combination. These findings could be reconciled by remembering that the major sector is made up of the temporal and nasal sectors, and that these two domains when merged offer a greater wide picture of the retinal. Furthermore, whenever the superior and inferior parts of the fundamental domain were eliminated, they had the lowest impact on deep learning algorithm predictive capability, implying that they were the lowest essential. In contrast, when the macular area is excised, it had the greatest impact and was also the more prognoses in solitude. As a result, the deep learning algorithm gave the macular and the retinal's peripheral equal weight in predictions.

The optimization of screening frequencies could be one use of this deep learning technique. The earliest retinopathy is discovered, ever more efficient therapies, including such intravitreal doses and infrared photocoagulation of antivascular endothelial development factors. While conventional diabetic retinopathy testing depended on indirectly or directly slit lamp biomicroscopy or ophthalmoscopy, fundus photography's simplicity of usage, expense efficiency, and precision has resulted to its inclusion in a number of diabetic retinopathy monitoring recommendations. Medical hazard elements, resource accessibility (for example, scanning technology, materials, and people), and other economic considerations all figure into the spectrum of testing periods in these recommendations (for example, 12 to 24 months for patients with no obvious diabetic retinopathy). Patients at higher hazard may also be worked up more regularly to guarantee earlier identification, while individuals at minimal hazard may be following up less regularly to decrease the monitoring burden experienced by individuals, physicians, and the medical systems. Individuals at the greatest hazard acquired diabetic retinopathy at a frequency of more than 90 percent in our research, while patients at the least hazard had a probability of getting diabetic retinopathy of fewer than 10 percent. Furthermore, additional research would require focusing on the particular cut-offs for determining higher and lower hazard category, as well as the corresponding treatments, which would most likely be adjusted to regional resources accessibility and practise structures.

5. Conclusion

A DRNN-deep recurrent neural network system for evaluating colour fundus photographs for diabetic retinopathy identification is presented in this research. On the DRIVE datasets, the approach is tested effectively. With 95.6 percent precision, the suggested algorithm categorised healthy and harmful photos. With drastically minimised cross-entropy losses functional of 0.4356 the suggested model achieves the best precision efficiency. When compared current approaches to the traditional inception deep RNN approach, there is a 14.69 percent boost in efficiency. This research focuses on attributes relevancy and categorization strategies for effectively categorising diseases related to the retina using characteristics derived from retinal pictures via photograph processor approaches. The long-term objective of this study is to build lower expense technology capable of on-site real-time retina picture categorization. Because of the complexity of the proposed applications, the sensitivity (true negatives) must be near to 100 percent to minimise the dangers of a false categorization outcome. In particular, the device's sensitivities must be increased for a practical use in order to eliminate unwanted testing and correlated expenses.

Data Availability

The data used to support the findings of this study are included within the article. Further data or information is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors appreciate the supports from MizanTepi University, Ethiopia for the research and preparation of the manuscript. The authors thank Sri Indu College of Engineering and Technology, Panimalar Engineering College, and Koneru Lakshmaiah Education Foundation for providing assistance to complete this work. This project was supported by Researchers Supporting Project number (RSP-2021/283) King Saud University, Riyadh, Saudi Arabia.

References

- [1] D. S. Fong, L. Aiello, T. W. Gardner et al., "Retinopathy in diabetes," *Diabetes Care*, vol. 27, suppl_1, pp. s84-s87, 2004.
- [2] G. G. Yen and W.-F. Leong, "A sorting system for hierarchical grading of diabetic fundus images: a preliminary study," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 118–130, 2008.
- [3] A. Tuulonen, P. J. Airaksinen, A. Montagna, and H. Nieminen, "Screening for glaucoma with a non-mydriatic fundus camera," *Acta Ophthalmologica*, vol. 68, no. 4, pp. 445–449, 1990.
- [4] E. Stefánsson, T. Bek, M. Porta, N. Larsen, J. K. Kristinsson, and E. Agardh, "Screening and prevention of diabetic blindness," *Acta Ophthalmologica Scandinavica*, vol. 78, no. 4, pp. 374–385, 2000.
- [5] V. Elizabeth Jesi, S. Mohamed Aslam, G. Ramkumar, A. Sabarivani, A. K. Gnanasekar, and P. Thomas, "Energetic glaucoma segmentation and classification strategies using depth optimized machine learning strategies," *Contrast Media & Molecular Imaging*, vol. 2021, article 5709257, 11 pages, 2021.
- [6] T. Y. Wong, R. Klein, B. E. Klein, J. M. Tielsch, L. Hubbard, and F. J. Nieto, "Retinal microvascular abnormalities and their relationship with hypertension, cardiovascular disease, and mortality," *Survey of Ophthalmology*, vol. 46, no. 1, pp. 59–80, 2001.
- [7] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, vol. 3, pp. 169–208, 2010.
- [8] N. Patton, T. M. Aslam, T. MacGillivray et al., "Retinal image analysis: concepts, applications and potential," *Progress in Retinal and Eye Research*, vol. 25, no. 1, pp. 99–127, 2006.
- [9] X. Li, X. Hu, L. Yu, L. Zhu, C.-W. Fu, and P.-A. Heng, "CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1483–1493, 2020.
- [10] F. Li, Z. Liu, H. Chen, M. Jiang, X. Zhang, and Z. Wu, "Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm," *Translational Vision Science & Technology*, vol. 8, no. 6, pp. 4–4, 2019.
- [11] V. Gulshan, R. P. Rajan, K. Widner et al., "Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India," *JAMA Ophthalmology*, vol. 137, no. 9, pp. 987–993, 2019.

- [12] G. Ramkumar and E. Logashanmugam, "Study on impulsive assessment of chronic pain correlated expressions in facial images," *Biomedical Research*, vol. 29, no. 16, 2018.
- [13] T. D. Keenan, S. Dharssi, Y. Peng et al., "A deep learning approach for automated detection of geographic atrophy from color fundus photographs," *Ophthalmology*, vol. 126, no. 11, pp. 1533–1540, 2019.
- [14] S. Keel, P. Y. Lee, J. Scheetz et al., "Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study," *Scientific Reports*, vol. 8, no. 1, pp. 1–6, 2018.
- [15] S. Karthikeyan, G. Ramkumar, S. Aravindkumar, M. Tamilselvi, S. Ramesh, and A. Ranjith, "A novel deep learning-based black fungus disease identification using modified hybrid learning methodology," *Contrast Media & Molecular Imaging*, vol. 2022, article 4352730, 11 pages, 2022.
- [16] R. Sayres, A. Taly, E. Rahimy et al., "Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy," *Ophthalmology*, vol. 126, no. 4, pp. 552–564, 2019.
- [17] C. Wilkinson, Ferris FL 3rd, R. E. Klein et al., "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.
- [18] D. Xiao, A. Bhuiyan, S. Frost, J. Vignarajan, M.-L. Tay-Kearney, and Y. Kanagasingam, "Major automatic diabetic retinopathy screening systems and related core algorithms: a review," *Machine Vision and Applications*, vol. 30, no. 3, pp. 423–446, 2019.
- [19] P. Vora and S. Shrestha, "Detecting diabetic retinopathy using embedded computer vision," *Applied Sciences*, vol. 10, no. 20, p. 7274, 2020.
- [20] M. Tamilselvi and G. Ramkumar, "Non-invasive tracking and monitoring glucose content using near infrared spectroscopy," in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1–3, Madurai, India, 2015.
- [21] C. Lam, D. Yi, M. Guo, and T. Lindsey, "Automated detection of diabetic retinopathy using deep learning," *AMIA summits on translational science proceedings*, vol. 2018, p. 147, 2018.
- [22] S. Mahesh and G. Ramkumar, "Smart face detection and recognition in low resolution images using Alexnet CNN compare accuracy with SVM," *Alinteri Journal of Agriculture Sciences*, vol. 36, no. 1, pp. 721–726, 2021.
- [23] R. Gargya and T. Leng, "Automated identification of diabetic retinopathy using deep learning," *Ophthalmology*, vol. 124, no. 7, pp. 962–969, 2017.
- [24] N. Datta, R. Banerjee, H. Dutta, and S. Mukhopadhyay, "Hardware based analysis on automated early detection of diabetic retinopathy," *Procedia Technology*, vol. 4, pp. 256–260, 2012.
- [25] J. Beagley, L. Guariguata, C. Weil, and A. A. Motala, "Global estimates of undiagnosed diabetes in adults," *Diabetes Research and Clinical Practice*, vol. 103, no. 2, pp. 150–160, 2014.
- [26] M. N. Ozieh, K. G. Bishu, C. E. Dismuke, and L. E. Egede, "Trends in health care expenditure in U.S. adults with diabetes: 2002–2011," *Diabetes Care*, vol. 38, no. 10, pp. 1844–1851, 2015.
- [27] L. Sellahewa, C. Simpson, P. Maharajan, J. Duffy, and I. Idris, "Grader agreement, and sensitivity and specificity of digital photography in a community optometry-based diabetic eye screening program," *Clinical Ophthalmology (Auckland, NZ)*, vol. 8, p. 1345, 2014.
- [28] L. Guariguata, D. R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, and J. E. Shaw, "Global estimates of diabetes prevalence for 2013 and projections for 2035," *Diabetes Research and Clinical Practice*, vol. 103, no. 2, pp. 137–149, 2014.
- [29] M. R. K. Mookiah, U. R. Acharya, C. K. Chua, C. M. Lim, E. Ng, and A. Laude, "Computer-aided diagnosis of diabetic retinopathy: a review," *Computers in Biology and Medicine*, vol. 43, no. 12, pp. 2136–2155, 2013.
- [30] M. D. Abràmoff, Y. Lou, A. Erginay et al., "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning," *Investigative Ophthalmology & Visual Science*, vol. 57, no. 13, pp. 5200–5206, 2016.
- [31] R. J. Winder, P. J. Morrow, I. N. McRitchie, J. Bailie, and P. M. Hart, "Algorithms for digital image processing in diabetic retinopathy," *Computerized Medical Imaging and Graphics*, vol. 33, no. 8, pp. 608–622, 2009.
- [32] D. Sidibé, I. Sadek, and F. Mériaudou, "Discrimination of retinal images containing bright lesions using sparse coded features and SVM," *Computers in Biology and Medicine*, vol. 62, pp. 175–184, 2015.
- [33] Y. Han, M. A. Bearse, M. E. Schneck, S. Barez, C. H. Jacobsen, and A. J. Adams, "Multifocal electroretinogram delays predict sites of subsequent diabetic retinopathy," *Investigative Ophthalmology & Visual Science*, vol. 45, no. 3, pp. 948–954, 2004.
- [34] A. Bora, S. Balasubramanian, B. Babenko et al., "Predicting the risk of developing diabetic retinopathy using deep learning," *The Lancet Digital Health*, vol. 3, no. 1, pp. e10–e19, 2021.
- [35] Y. Han, M. E. Schneck, M. A. Bearse Jr. et al., "Formulation and evaluation of a predictive model to identify the sites of future diabetic retinopathy," *Investigative Ophthalmology & Visual Science*, vol. 45, no. 11, pp. 4106–4112, 2004.
- [36] A. Sungheetha and R. Sharma, "Design an early detection and classification for diabetic retinopathy by deep feature extraction based convolution neural network," *Journal of Trends in Computer Science and Smart technology*, vol. 3, no. 2, pp. 81–94, 2021.
- [37] F. Arcadu, F. Benmansour, A. Maunz, J. Willis, Z. Haskova, and M. Prunotto, "Deep learning algorithm predicts diabetic retinopathy progression in individual patients," *NPJ Digital Medicine*, vol. 2, no. 1, pp. 1–9, 2019.
- [38] S. K. Somasundaram, "A machine learning ensemble classifier for early prediction of diabetic retinopathy," *Journal of Medical Systems*, vol. 41, no. 12, pp. 1–12, 2017.
- [39] V. Gulshan, L. Peng, M. Coram et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.
- [40] T. Araújo, G. Aresta, L. Mendonça et al., "DR|GRADUATE: uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images," *Medical Image Analysis*, vol. 63, article 101715, 2020.
- [41] M. Voets, K. Møllersen, and L. A. Bongo, "Replication study: development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," 2018, <https://arxiv.org/abs/1803.04337>.

- [42] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Ijcai*, vol. 14, no. 2, pp. 1137–1145, 1995.
- [43] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [44] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with LSTM," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [45] M. Golabbakhsh and H. Rabbani, "Vessel-based registration of fundus and optical coherence tomography projection images of retina using a quadratic registration model," *IET Image Processing*, vol. 7, no. 8, pp. 768–776, 2013.
- [46] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Cham, 2015.
- [47] A. M. Omer and M. Elfadil, "Preprocessing of digital mammogram image based on otsu's threshold," *American Academic Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 37, no. 1, pp. 220–229, 2017.