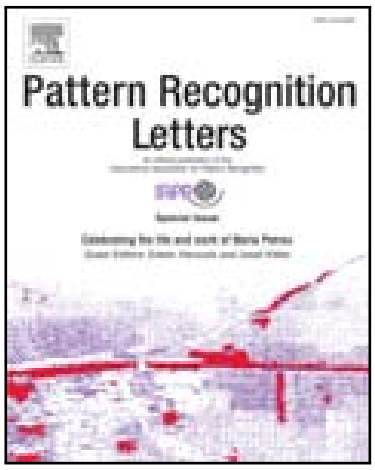

Deep Learning Frameworks for Diabetic Retinopathy Detection with Smartphone-based Retinal Imaging Systems

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

Pattern Recognition Letters
journal homepage: www.elsevier.com

Deep Learning Frameworks for Diabetic Retinopathy Detection with Smartphone-based Retinal Imaging Systems

[a]Recep E. Hacisoftaoglu, [a]Mahmut Karakaya[*], [b]Ahmed B. Sallam

[a]*Dept. of Computer Science, University of Central Arkansas, Conway, AR, 72035, USA*
[b]*Jones 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.

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.

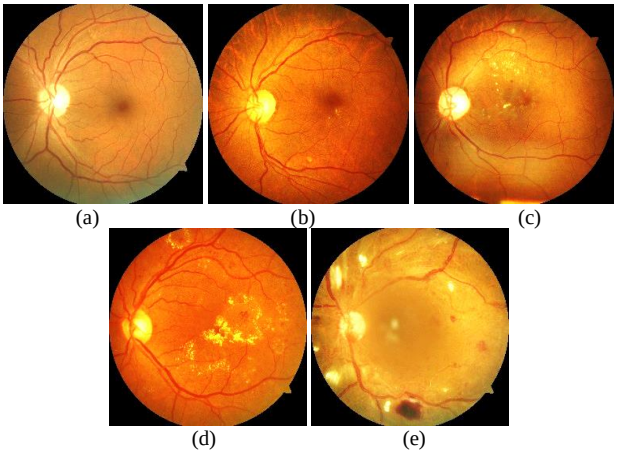

Fig. 1. Retina images from the UoA-DR dataset with different DR levels, (a) normal, (b) mild, (c) moderate, (d) severe, and (e) proliferative.

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 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

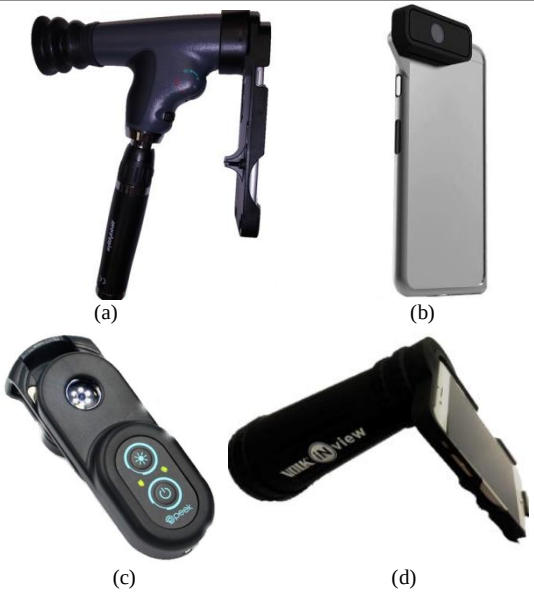

Fig. 2. Smartphone-based retinal imaging systems available in the market, (a) iExaminer (b) D-Eye, (c) Peek Retina, and (d) iNview.

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.

TABLE I
SPECIFICATIONS OF SMARTPHONE-BASED RETINAL IMAGING SYSTEMS

	iExaminer	**D-Eye**	**Peek Retina**	**iNview**
Light Source	Self	Phone	Self	Phone
Degree of Retinal View	25	6-20	20-30	50
Working Distance (mm)	22	22	22	65
Size (mm)	70/220/162	68/135/7	25/75/35	180/76/180
Weight (g)	390	25	43	332
Price ($)	750	400	235	995

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. Gargeya 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-Optha 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 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).

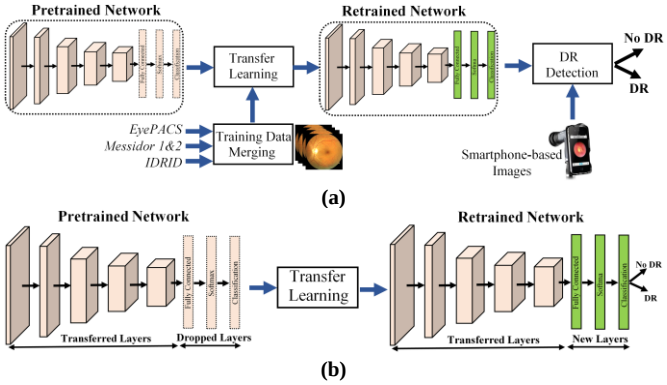

Fig. 3. (a) Flow chart of the proposed smartphone-based DR detection method and (b) Transfer learning approach for deep learning architectures.

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
EyePACS	25810	2443	5292	873	708	35126
EyePACS-u[*]	9895	899	2175	568	317	13624
Messidor	547	149	240	251	-	1187
Messidor-2	1017	270	347	75	35	1748
IDRiD	168	25	168	93	62	516
UoA-DR	56	9	50	55	30	200

[*]*EyePACS-u:* EyePACS-Updated

TABLE III
LABEL ASSIGNMENTS FOR DR SEVERITY

Datasets	Label0	Label1	Label2	Label3	Label4
EyePACS	No	Mild	Moderate	Severe	Prolif.[*]
Messidor	No	Mild	Mod-Sev.[+]	Prolif.[*]	-
Messidor-2	No	Mild	Moderate	Severe	Prolif.[*]
IDRiD	No	Mild	Moderate	Severe	Prolif.[*]
UoA-DR	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 Mesidor-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, Mesidor-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 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).

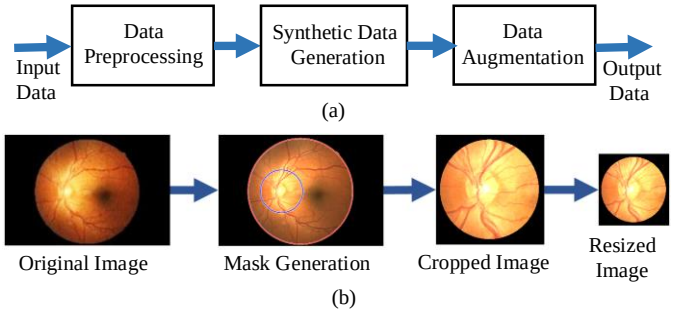

Fig. 4. (a) Workflow of synthetic retina image generation for smartphone-based retinal imaging systems and (b) Steps of synthetic data generation with masking, cropping, and resizing the input images.

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).

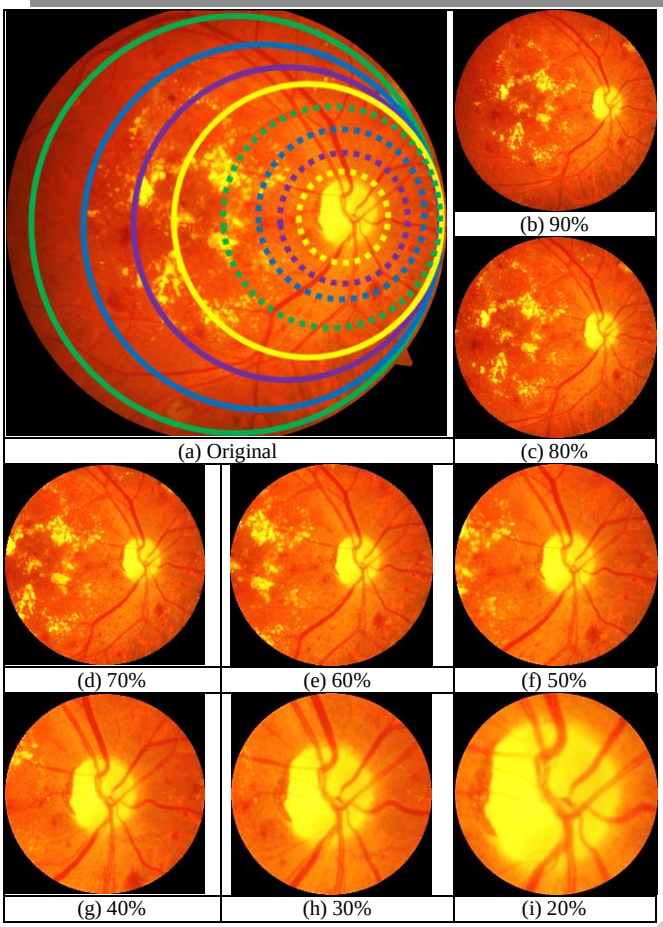

Fig. 5. (a) Comparison of the FoV of synthetic images with different percentages w.r.t the original image where the solid green, blue, purple, and yellow lines represent 90%, 80%, %70, and %60 FoV, respectively. The dotted green, blue, purple, and yellow lines represent 50%, 40%, 30%, and 20% FoV, respectively. 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, Messidor, and IDRiD). In this paper, the deep neural network was trained, validated, and tested with images from single, crossed, and merged datasets.

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

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.

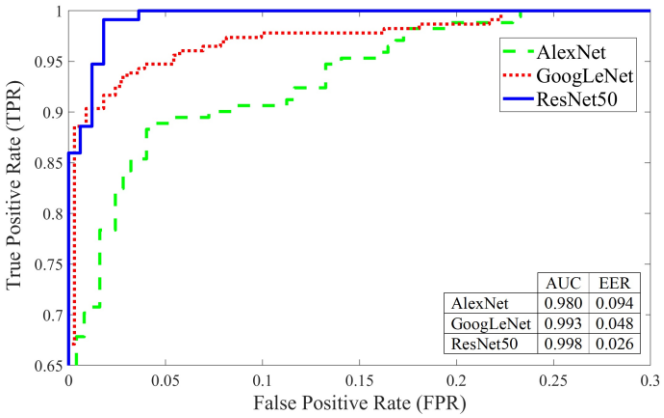

Fig. 6. Performance analysis using ROC for AlexNet, GoogLeNet, and ResNet50 frameworks from original retina images.

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

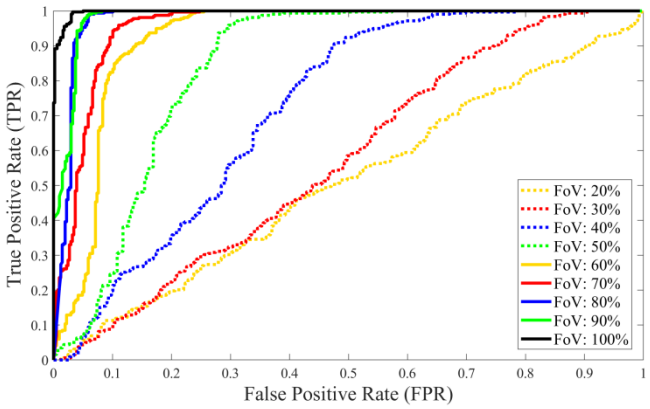

Fig. 7. Performance analysis using ROC for different testing images from original and synthetic smartphone-based retinal images with various percentage of FoV compared with original retina images.

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 iNview 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 (σ = 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 (σ = 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] Petrushkin, 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] Gargeya, 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] Hacisoftaoglu, 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.