

**DEEP LEARNING FRAMEWORKS FOR DIABETIC RETINOPATHY  
DETECTION USING SMARTPHONE-BASED RETINAL IMAGING SYSTEMS**

by

Recep Emre Hacisoftoglu

A thesis presented to the Department of Computer Science  
and the Graduate School of University of Central Arkansas in partial  
fulfillment of the requirements for the degree of

Master of Science

in

Computer Science

Conway, Arkansas

December 2019

TO THE OFFICE OF GRADUATE STUDIES:

The members of the Committee approve the thesis of Recep Emre Hacısoftaoglu  
presented on November 14, 2019.



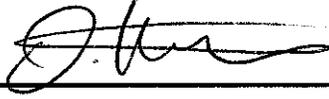
---

Dr. Mahmut Karakaya, Committee Chairperson



---

Dr. M. Emre Celebi, Committee Member



---

Dr. Olcay Kursun, Committee Member

## PERMISSION

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

Department    Computer Science

Degree         Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of Central Arkansas, I agree that the Library of this University shall make it freely available for inspections. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work, or, in the professor's absence, by the Chair of the Department or the Dean of the Graduate School. It is understood that due recognition shall be given to me and to the University of Central Arkansas in any scholarly use which may be made of any material in my thesis.



---

Recep Emre Hacisoftaoglu

November 14, 2019

## ABSTRACT

Diabetic Retinopathy, the most common reason of vision loss, is caused by damage to the small blood vessels in the retina. If untreated, it may result in varying degrees of vision loss and even blindness. Since Diabetic Retinopathy is a silent disease that may cause no symptoms or only mild vision problems, annual eye exams are crucial for early detection to improve the chances of effective treatment where fundus cameras are used to capture the retinal images. However, fundus cameras are too big and heavy to be transported easily and too costly to be purchased by every health clinic, so fundus cameras are an inconvenient tool for widespread screening. Recent technological developments have enabled using smartphones in designing small-sized, low-power, and affordable retinal imaging systems to perform Diabetic Retinopathy screening and automated Diabetic Retinopathy detection using machine learning and image processing methods. However, Diabetic Retinopathy detection accuracy depends on the image quality and it is negatively affected by several factors such as Field of View. Since smartphone-based retinal imaging systems have much more compact designs than the traditional fundus cameras, the retina images captured are likely to be low quality with smaller Field of View.

In this thesis, we first investigate the smartphone-based portable ophthalmoscope systems available on the market and compare their Field of View and image quality to determine if they are suitable for Diabetic Retinopathy detection during a general health screening. Then, we propose automatic Diabetic Retinopathy detection algorithms for smartphone-based retinal images using deep learning frameworks, AlexNet and GoogLeNet. To test our proposed methods, we generate smartphone-based synthetic retina images by simulating the different Field of View with masking the original image around

the optic disk and cropping it. Using transfer learning, we retrain the pretrained networks with retina images from several datasets including EyePACS, Messidor, IDRiD, and Messidor-2. Then, we compare our smartphone-based results with the original retina images from the University of Auckland Diabetic Retinopathy dataset. As a main contribution to the existing literature, we study the effect of the Field of View in smartphone-based retinal imaging and improve Diabetic Retinopathy detection accuracy by training network with merged publicly available datasets.

Based on the results, iNview retinal imaging system has the largest Field of View and better image quality compared with iExaminer, D-Eye, and Peek Retina systems. Using AlexNet, the overall classification accuracies of smartphone-based systems are sorted as 61%, 62%, 69%, and 75% for iExaminer, D-Eye, Peek Retina, and iNview images, respectively. We observed that the network DR detection performance decreases as the Field of View of the smartphone-based retinal systems get smaller where iNview is the largest and iExaminer is the smallest. Finally, we observed the best results for vision threatening Diabetic Retinopathy detection with GoogLeNet as 95% testing accuracy and 95% sensitivity for original images, while 89% testing accuracy and 94% sensitivity for smartphone-based synthetic images. Although we have used a smaller number of images in the training set compared with the existing deep networks in the literature, we obtained considerably acceptable higher accuracies for validation and testing data. As a result, the smartphone-based retina imaging systems can be used as an alternative to the direct ophthalmoscope once it tested in the clinical settings. However, the Field of View of the smartphone-based retina imaging systems plays an important role in determining the automatic Diabetic Retinopathy detection accuracy.

# TABLE OF CONTENTS

ABSTRACT.....	iv
LIST OF FIGURES .....	viii
LIST OF TABLES .....	xi
1. INTRODUCTION .....	1
2. RELATED WORKS .....	7
3. METHODS .....	11
3.1. Smartphone-based Portable Retinal Imaging Systems .....	11
3.1.1. 20D Lens .....	16
3.1.2. iExaminer .....	16
3.1.3. D-Eye .....	19
3.1.4. Peek Retina .....	19
3.1.5. Volk iNview.....	20
3.2. Field of View (FoV) and Image Quality for Smartphone-based Retinal Imaging Systems .....	21
3.2.1. Circular Test Patterns.....	21
3.2.2. Best Distance Calculation for Environment Setting .....	24
3.3. Deep Learning Architectures for Diabetic Retinopathy Detection.....	32
3.3.1. AlexNet .....	33
3.3.2. GoogLeNet.....	35
4. EXPERIMENTAL SETUP AND DATASETS.....	38
4.1. Datasets .....	38
4.2. Dataset Preparation .....	41
4.3. Dataset Augmentation.....	43
4.4. Data Merging and Splits for Training and Validation .....	44
5. RESULTS AND DISCUSSION .....	46
5.1. Diabetic Retinopathy Detection using Single Dataset.....	46
5.1.1. Results for Synthetic Retina Box Images .....	46
5.1.2. Results for Smartphone-based Images.....	49
5.2. Diabetic Retinopathy Detection using Multiple Datasets .....	56
5.2.1. Results for Original Images .....	57
5.2.2. Results for Smartphone-based Images.....	60
6. CONCLUSIONS.....	65

REFERENCES ..... 67

## LIST OF FIGURES

Figure 1. Retinal images from the UoA – DR dataset with different DR levels, (a) normal, (b) mild, (c) moderate, (d) severe, and (e) proliferative ..... 2

Figure 2. Smartphone-based retinal imaging systems available in the market, (a) Fundus photography method used by Harvard Medical School (b) iExaminer developed by Welch Allyn, (c) D-Eye, (d) Peek Retina, and (e) iNview by Volk..... 5

Figure 3. The general optical design of the smartphone-based retinal imaging devices.. 12

Figure 4. (a) Peek Retina synthetic eye model in a box for data collection and (b-c) Synthetic retina images. .... 14

Figure 5. Experimental setup retinal imaging with (a) iExaminer, (b) D-Eye, (c) Peek Retina, and (d) Volk iNview..... 18

Figure 6. The circular test pattern of the Peek Retina synthetic eye model box..... 22

Figure 7. Circular test pattern images captured by iExaminer from (a) 150mm, (b) 100mm, (c) 75mm, (d) 65mm, (e) 50mm, (f) 35mm, (g) 30mm, (h) 22mm in distance. .... 25

Figure 8. Circular test pattern images captured by D-Eye from (a) 150mm, (b) 100mm, (c) 75mm, (d) 65mm, (e) 50mm, (f) 35mm, (g) 22mm in distance. .... 26

Figure 9. Circular test pattern images captured by Peek Retina (a) 150mm, (b) 100mm, (c) 75mm, (d) 65mm, (e) 50mm, (f) 35mm, (g) 22mm in distance. .... 28

Figure 10. (a) Circular test pattern images captured by iNview from (a) 150mm, (b) 100mm, (c) 75mm, (d) 65mm, (e) 50mm, (f) 35mm, (g) 30mm, (h) 22mm in distance.. 30

Figure 11. (a) Circular test pattern image (b) Comparison of the FoV of each smartphone-based retinal imaging system where the solid black line is for iNview, the green dashed-dotted line is for D-Eye, the purple dashed line is for Peek Retina, the red dotted line is for

iExaminer. Captured images of circular test patterns at the best distance by (c) iExaminer, (d) D-Eye, (e) Peek Retina and (f) iNview. ....	31
Figure 12. Flow chart of original retina conversion images for GoogLeNet.....	41
Figure 13. Comparison of the FoV of synthetic retina images with different percentages with respect to the original image where the solid green, blue, purple, and white lines represent 90%, 80%, 70%, and 60% FoV, respectively. The dotted green, blue, purple, and white lines represent 50%, 40%, 30%, and 20% FoV, respectively. ....	42
Figure 14. Input images for deep learning architecture with (a) original size and synthetic images with a smaller FoV where their percentage w.r.t the original image as (b) 90%, (c) 80%, (d) 70%, (e) 60%, (f) 50%, (g) 40%, (h) 30%, (i) 20%. ....	43
Figure 15. Retina images. (1st row) Original printed real retina images and their captured versions using (2nd row) iExaminer, (3rd row) D-Eye, (4th row) Peek Retina, (5th row) iNview.....	48
Figure 16. Comparison of the FoV of each smartphone-based retinal imaging system using printed normal retina image with FoV markers where the solid black line is for iNview, the purple dashed line is for Peek Retina, the green dashed-dotted line is for D-Eye, the red dotted line is for iExaminer.....	49
Figure 17. (a) Flow chart of original retina conversion images for AlexNet, (b) Flow chart of synthetic retina image generation for smartphone-based retinal imaging systems. Example of images, (c) Original image from EyePACS dataset and generated synthetic images for (d) iExaminer, (e) D-Eye, (f) Peek Retina, (g) iNview.....	50
Figure 18. Generation of the customized dataset of retinal fundus images and its label distribution. ....	52

Figure 19. Classification accuracy of deep learning network for different testing images from original and synthetic smartphone-based retinal images. ....	54
Figure 20. Performance analysis using ROC for different testing images from original and synthetic smartphone-based retinal images. ....	55
Figure 21. Overall classification accuracy of deep learning networks for different smartphone-based synthetic retina images with respect to their FoV percentage .....	62
Figure 22. Performance analysis using ROC for different testing images from original and synthetic smartphone-based retinal images with various percentages of FoV compared with original retina images.....	63

## LIST OF TABLES

Table 1. Comparison of smartphone-based retinal imaging systems available in the market .....	15
Table 2. AlexNet Architecture .....	34
Table 3. Retina Image Datasets with DR Severity Labels.....	39
Table 4. Label Assignments for DR Severity .....	40
Table 5. Classification accuracy of the deep learning network. Note that each result shows the mean and its standard deviation as mean $\pm$ std. ....	53
Table 6. Area Under Curve (AUC), Equal Error Rate (EER), and Selected Threshold for EER parameters at ROC curve for different testing images from original and synthetic smartphone-based retinal images.....	56
Table 7. Classification accuracy of deep learning frameworks .....	58
Table 8. Accuracy for rDR detection using two operating points .....	59
Table 9. Accuracy for vtDR detection using two operating points.....	59
Table 10. Classification accuracy of deep learning frameworks .....	61
Table 11. Area Under Curve (AUC), Equal Error Rate (EER), and Selected Threshold for EER results of deep learning frameworks .....	64

## 1. INTRODUCTION

The World Health Organization estimates that 347 million people have diabetes worldwide and the number will increase to 552 million by the year 2030. In the United States, more than 9 percent of Americans (29 million) have the disease and 8 million of those are undiagnosed. A diabetic person is at high risk of eye disease including diabetic retinopathy (DR), diabetic macular edema (DME), cataract, and glaucoma. The most common cause of vision loss, DR is caused by bleeding of the small blood vessels in the retina. If these bloody retina veins are untreated, it may cause varying degrees of vision loss and even blindness. The signs of DR can be listed as including but not limited to the existence of microaneurysms, hard exudates, vitreous hemorrhages, and retinal detachments. Figure 1 shows retina images with different DR levels such as (a) normal, (b) mild, (c) moderate, (d) severe, and (e) proliferative. As shown in Figure 1(a), there is no abnormal lesions in the normal retina. Based on the severity and amount of the microaneurysms, hard exudates, vitreous hemorrhages, and retinal detachments, Figure 1(b-d) are labeled differently from mild to proliferative using clinical DR disease severity scales (See Section 4.1. Datasets). Readers can easily observe the differences, especially in the yellow and red regions.

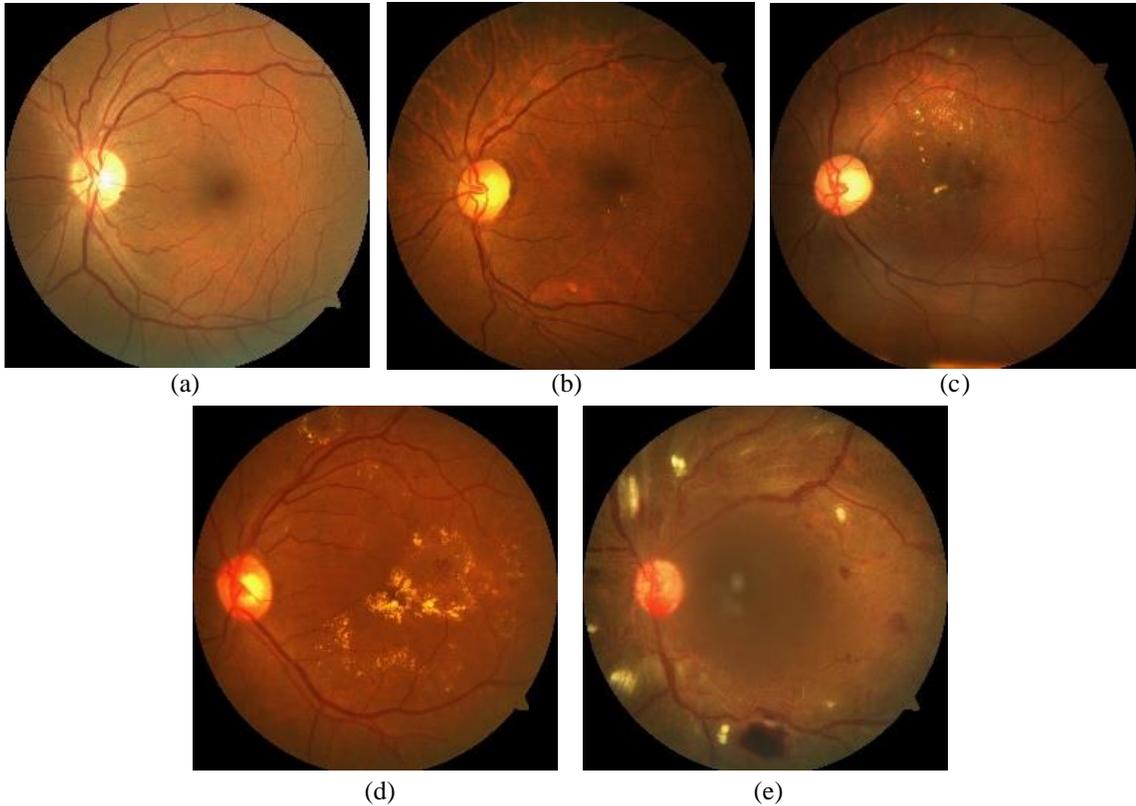


Figure 1. Retinal images from the UoA – DR dataset with different DR levels, (a) normal, (b) mild, (c) moderate, (d) severe, and (e) proliferative

In the US, more than 4.4 million people aged 40 and older had DR problems at different stages. Due to its silent nature, DR may cause no symptoms or only mild vision problems [1]. Since early detection could improve the chances of effective treatment to prevent blindness, doctors suggest annual eye exams for diabetic patients. With early diagnosis and accurate evaluation of DR severity, it is possible to coordinate diabetic eye care and prompt appropriate treatment for the prevention of blindness and visual loss [2]. However, current studies show that access to such medical care in developed countries ranges between 60% and 90%, with significantly lower rates in developing countries [3]. Patients without eye care access do not benefit from the potential of early detection and

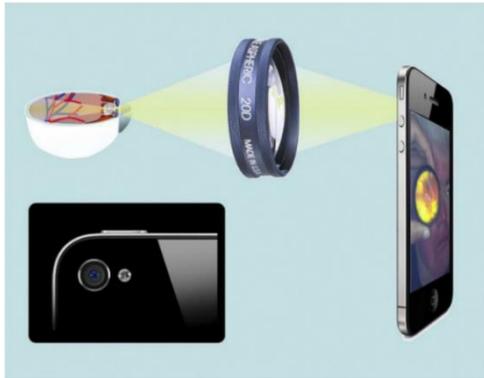
getting effective treatment in a timely manner. Among rural and minority populations, there exists a significant disparity in early diagnosis and access to eye care.

In the retina examination, doctors use special optical devices such as an ophthalmoscope, a 20D (diopter) lens, and a fundus camera. Due to the easier storage, better image quality, and faster electronic transfer, fundus cameras are widely used to detect eye diseases with their digital imaging features. However, retinal imaging with a fundus camera is a time-consuming and manual process. Some certain level of expertise is required to capture a retinal image and an expert evaluation report may take a few days to submit. Since fundus cameras are also too big and heavy to be transported easily and too costly to be purchased by every health clinic, fundus cameras are inconvenient tools for widespread screening. In addition, it is very hard to find equipment and expert in rural areas that have a high diabetes rate [4]. The need for infrastructure for DR screening will become even more insufficient as the number of diabetic people increases. Therefore, there is a growing demand for small, portable, and inexpensive retinal imaging systems to perform DR screening and automated DR detection methods using data science and image processing techniques.

Recent technological developments have enabled using of smartphones in designing small-sized, low-power, and affordable biomedical imaging devices. These systems are capable of imaging, onboard processing, and wireless communication. Therefore, smartphone-based systems are very popular in several applications ranging from health care to entertainment since they make existing systems small and portable. Since fundus cameras are large-size, heavy-weight, and high-price devices, they are a good candidate to be transformed into a portable device to perform fast DR screening.

Developing new portable retinal imaging systems using smartphones is an emerging research and technology area that attracts several universities and companies.

Using 20D lens with a smartphone is one of the early designs to record video on human and rabbit eyes [5] as shown in Figure 2(a). This early design used the Filmic Pro mobile application to adjust focus, exposure, and light intensity manually. This allows having better quality fundus images in clinic settings for both under anesthesia and awake conditions. Another very simple smartphone-based design is the iExaminer system of Welch Allyn [6]. It is developed by attaching a smartphone to a PanOptic ophthalmoscope as shown in Figure 2(b). These systems require attaching the smartphone to an existing medical device. Recently, new smartphone-based retinal imaging systems are released to the market including D- Eye, Peek Retina, and iNview. These smartphone-based retinal imaging systems are bundled with associated secure HIPPA-compliant mobile application for image capture and transmission. In addition, iExaminer, D-Eye, and iNview have the Food and Drug Administration (FDA) approval for their devices. However, Peek Retina is currently waiting for the approval from FDA.



(a)



(b)



(c)



(d)



(e)

Figure 2. Smartphone-based retinal imaging systems available in the market, (a) Fundus photography method used by Harvard Medical School (b) iExaminer developed by Welch Allyn, (c) D-Eye, (d) Peek Retina, and (e) iNview by Volk.

Russo et al. [7] developed the D-Eye, a small, portable, and inexpensive retinal imaging system to capture retina images as an attachment to a smartphone. Using the cross-polarization technique to reduce corneal reflections and integrated with the smartphone's autofocus feature to prevent the patient's refractive error, D-Eye allows retinal eye screening even for undilated eyes. To illuminate the retina, D-Eye reflects the smartphone flashlight as shown in Figure 2(c). Peek Retina imaging system [8] includes its own light to illuminate the retina and the image is captured by a smartphone as shown in Figure 2(d). To capture wide-angle retinal images, Volk Optical developed iNview smartphone-based retinal imaging system [9] as shown in Figure 2(e). It illuminates the retina by reflecting the smartphone flashlight and does not require pupil dilation to acquire 50 degrees of retinal view to visualize the entire posterior pole in a single image. All these smartphone-based imaging systems are capable of capturing retina images, but none of these smartphone-based systems offer a way to analyze and evaluate eye disease by using image processing techniques.

## 2. RELATED WORKS

In this section, we introduce solutions for the detection of DR using conventional and deep learning methods. There exist several telemedicine solutions for retinal image analysis for DR screening [10]. These solutions in the literature need manual grading. However, they have to be fully automated to accelerate the diagnosis of retinal diseases using predictive models, especially for patients in rural areas. Researchers presented automatic retinal image analysis (ARIA) tools including iGradingM [11], The TRIAD Network [12, 13], IDx-DR [14], RetmarkerDR [15], and Retianalyze [16]. However, these methods are semi-automated and require some expert control to decide the existence of the retinal disorder. This is the main obstacle to apply them automatically for large datasets.

Kaggle's competition is one of the attempts to attract researchers to present solutions for DR detection providing the EyePACS retinal images. EyePACS contains 78685 retinal images to assess the DR existence and severity recommendation in different studies. This competition gives an opportunity to researchers to design their deep learning networks and try them with a publicly available dataset. It also accelerates the research findings for automatic DR detection and usage of the deep networks for classification based medical research with higher detection performance. Graham is the winner achieving the accuracy score of 0.84957. He first preprocessed the retina images to remove the illumination difference and used a convolutional neural network, SparseConvNet and random forest for classification by augmenting the retina images to increase the number of images in the training set [17].

With the improvement of computational power and advances in neural networks, deep learning algorithms, especially Convolutional Neural Networks (CNNs), have been

widely used in different applications including retinal imaging. Rajalakshmi et al. [18] proposed using Remidio Fundus on Phone (FOP) device to capture high-quality retina images compared with the traditional fundus devices. The FOP is a high-quality portable fundus camera that is capable of capturing wide retinal color photography covering macula, nasal to the optic disc, superior-temporal and inferior-temporal quadrants. After data capturing, retina images are graded with EyeArt [19, 20], a deep learning method to determine the existence of the DR. It is a cloud-based retinal image assessment tool to grade DR development using deep learning methods that were trained with EyePACS dataset [21]. The EyeArt system also offers image processing and machine learning techniques such as image gradeability, image enhancement, image restoration, interest region detection, and descriptor computation. Compared with manually graded results by two ophthalmologists, deep artificial neural network method shows high sensitivity and specificity for retina images captured by FOP device. FOP proves the concept of smartphone-based designs and shows the technological and economic feasibility of the portable retinal imaging systems. However, due to their fewer controllable parameters such as aperture and inexpensive lenses, smartphone-based systems have a lower image quality compared to the fundus camera and FOP. Therefore, the existing algorithms could not be applied directly to the retinal images captured with smartphone-based retinal imaging systems because the quality of captured retina images plays an important role in the accuracy of the deep learning techniques.

Gulshan et al. used specific deep learning methods for automated DR detection [22]. They have used CNN methods for image classification and trained their algorithm with Inception-v3 architecture [23] with about 128175 images using EyePACS and

Messidor-2 datasets. They have mentioned the variety of DR problems such as referable Diabetic Retinopathy (rDR), vision-threatening Diabetic Retinopathy (vtDR), and referable Diabetic Macular Edema (rDME). Based on the results, they suggest that such a CNN network gives the best results when the network trained with 60000 images in terms of sensitivity and specificity. For the sensitivity, they reached about 97%. In order to apply these results to the clinical environment, more research is needed. They obtained an accuracy of 97.5% and 96.1% choosing two different threshold values.

Abramoff et al. [24] developed Iowa detection program for detecting rDR and they have used their own DR database and publicly available Messidor-2 dataset for training and testing, respectively. Based on the test results on the Mesidor-2 dataset, they achieved 96.8% sensitivity and 59.4% specificity for detecting rDR. After that, they improved their results and reported 96.8% sensitivity and 87% specificity for detecting rDR using machine learning methods. [25]

Gargeya et al. [26] used a customized CNN technique to detect DR. They trained their system with 75137 fundus images from their own dataset and tested with Messidor-2 and E-Optha datasets. They classified images into two categories, one with the healthy eyes, the other with any DR stage, in other words, mild and worse DR. They acquired 94% sensitivity and 98% specificity from their own dataset. Also, they tested their model with Messidor-2 dataset, and they achieved 93% sensitivity and 87% specificity.

Philip et al. [27] developed a DR assessment system based on healthy and disease conditions, also known as mild DR and worse. They trained their algorithm with 1067 images and tested with 14406 images. The performance of their algorithm was 86.2% and 76.8% with sensitivity and specificity, respectively.

Carson et al. [28] introduced a CNN based deep learning techniques to detect DR using various classification models including but not limited to 2-ary, 3-ary, and 4-ary. Pretrained AlexNet and GoogLeNet models were investigated and transfer learning approaches were applied using Kaggle EyePACS and Messidor-1 dataset. They suggested using image processing techniques to increase validation accuracy especially for the detection of mild DR such as image normalization and contrast adjustments using histogram equalization. They augmented the retina images to increase the number of images in the training set and prevent overfitting. They received 95% sensitivity in the validation set and its accuracy in the testing set 74.5%, 68.75%, and 51.25% for 2-ary, 3-ary, and 4-ary models, respectively.

Pires et al. proposed a solution for detecting rDR using data-driven approaches [29]. They used transfer learning techniques by applying to CNN. They applied on their training stage to data augmentation, multi-resolution, feature extraction, per patient analysis, and testing their solution with cross dataset logic by using Kaggle EyePACS dataset as a training, Messidor-2 dataset for testing. Based on the results, they obtained a 98.2% Receiver Operating Characteristic (ROC) curve for predicting rDR.

Since these methods in the literature focus on fundus images, they cannot be easily applied to smartphone-based images. For this reason, we need to investigate and create our own synthetic DR dataset using the FoV approach.

### 3. METHODS

In this section, we first present the details of smartphone-based portable retinal imaging systems available on the market to compare their features and image qualities. Second, we introduce the Field of View (FoV) determination process of each smartphone-based retinal imaging system using a circular test pattern. Third, we introduce the layout of the adopted deep learning architecture for DR detection.

#### 3.1. Smartphone-based Portable Retinal Imaging Systems

There are four smartphone-based retinal imaging systems available in the market including iExaminer, D-Eye, Peek Retina, and iNview. First, we attach each smartphone-based retinal imaging devices to the smartphone to capture retinal images. Since the retina is located at the back of the eye and light does not reflect back through the retina, it is not possible to take a picture of the retina directly without any external illumination. Therefore, these imaging systems attached to the smartphone needs to illuminate the dark retina by reflecting the smartphone flashlight on the retina or using its own light to brighten the retina. Figure 3 shows the general optical design of the smartphone-based retinal imaging devices. Based on their optical design, each smartphone-based portable retinal imaging system has different degrees in angles of retinal views (AoV) to capture retinal images. It suggests that this difference in angles determines the FoV of different imaging systems. For example, the distance of the device to the eye, (i.e.,  $x$ ), the length of the device (i.e.,  $y$ ), and the number of lenses in the optic design determine the AoV in different imaging systems.

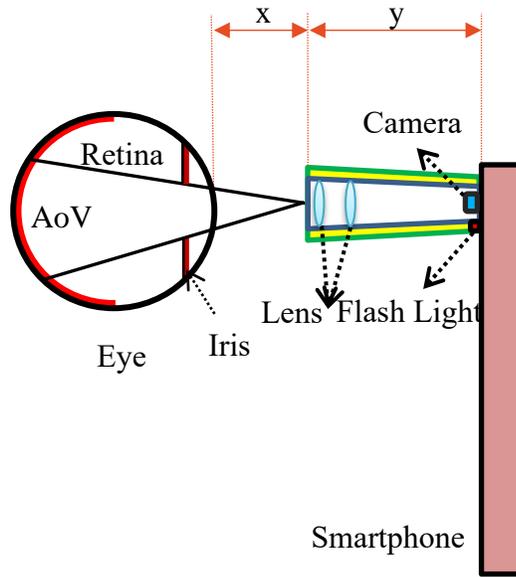


Figure 3. The general optical design of the smartphone-based retinal imaging devices.

Even in a controlled environment, pupil dilation level of the human subject and the eye gaze between a human subject and the retinal imaging system changes easily. Since the visible portions of the retina will change, the retina images captured from a human subject changes over time for different smartphone-based retinal imaging systems. To test the visible retina, a synthetic eye model can be used. For our experimental setup, we used the synthetic eye model box provided with Peek Retina smartphone-based retinal imaging system as shown in Figure 4(a). It replicates the eye structures including pupil, lens, and retina. The dimensions of the synthetic eye model are 53 mm x 53 mm x 22 mm in width, length, and depth. The opening on the box replicates the pupil of the eye. Under the pupil opening, there is a lens to illustrate the lens of the eye. Since the distance from the front surface of the cornea to the retina is approximately 24 mm on average, a printed test image is placed inside the box at the bottom that gives the same distance. As shown in Figure 4(b-

c), the printed real retina images show related retina tissues including the optic nerve, macula, and blood vessel. There are eight different high-quality printed real retina images provided with Peek Retina system including (1) normal retina, (2) glaucoma, (3) age-related macular degeneration, (4) diabetic retinopathy - clinically significant macular edema, (5) branch retinal vein occlusion, (6) diabetic retinopathy - ghost vessel, (7) papilloedema optic disc swelling, and (8) diabetic retinopathy - proliferative.



(a)



(b)

(c)

Figure 4. (a) Peek Retina synthetic eye model in a box for data collection and (b-c) Synthetic retina images.

In the following subsections, we present the details about the publicly available smartphone-based portable retinal imaging systems and their features. Table 1 summarizes and compares the important features of smartphone-based retinal imaging systems publicly available in the market including their price, size, weight, compatible smartphones, illumination source, pupil dilation dependency, degree of retinal view, type of captured data with its own mobile application, image size, and the maximum number of images.

Table 1. Comparison of smartphone-based retinal imaging systems available in the market

<b>Properties</b>	<b>iExaminer</b>	<b>D-Eye</b>	<b>Peek Retina</b>	<b>iNview</b>	<b>20D Lens</b>
<b>Compatible Smartphones</b>	iPhone 6	iPhone 6/7	Universal	iPhone 5/6/6s	Universal
<b>Illumination Source</b>	Inside	Flashlight	Inside	Flashlight	Flashlight
<b>Dilation Dependency</b>	Not Required	Required	Required	Not Required	Required
<b>Degree of Retinal View</b>	25	6-20	20-30	50	46
<b>Working Distance (mm)</b>	22	22	22	65	50
<b>Size (mm)</b>	70/220/162	68/135/7	27/75/35	180/76/180	50/50/10
<b>Weight (gr)</b>	390	25	43	332	50
<b>Price (\$)</b>	750	400	235	995	113
<b>Own Mobile Application</b>	Yes	Yes	Yes	Yes	No
<b>Type of Captured Data</b>	Image	Video	Image / Video	Image	N/A
<b>Images/Video Duration</b>	5 Images	30 Seconds	N/A	9 Images	N/A
<b>Size of Images</b>	320x280	640x480	640x480	640x480	N/A
<b>Data Capture Modes</b>	Manual	Manual	Manual	Manual/Auto	Manual

### **3.1.1. 20D Lens**

One of the early and simple designs for smartphone-based retinal imaging is developed using 20D (diopter) lens at Harvard Medical School and the Massachusetts Eye Hospital [5]. An operator holds the 20D lens directly in front of the eye and captures the retinal images with a smartphone. The flashlight of the smartphone or an external light source attached to the doctor's head is used to illuminate the retina. After capturing images using any smartphone application, retinal images are extracted using the virtue of MovieToImage and Video2Photo applications. In addition, light intensity, focusing, and exposure can be adjusted before data capture. In order to acquire the retinal images for optimal quality, not only is the pupil dilation suggested but also it requires a certain level of expertise to capture retinal images.

### **3.1.2. iExaminer**

iExaminer is another simple early design for a fundus photography method using a smartphone [6]. It is developed by attaching a smartphone to a Welch Allyn PanOptic ophthalmoscope. PanOptic Ophthalmoscope has been used for several years by ophthalmologists to provide convenient eye exams and detect eye diseases, including DR. In the iExaminer system, the eyecup of the ophthalmoscope where the doctor looks removed and the smartphone is attached using a special attachment to capture the retina images using the smartphone's camera. The iExaminer system involves three main pieces, including a smartphone application, iExaminer adapter, and PanOptic Ophthalmoscope. However, it is only compatible with the iPhone 4 and 6 since there is no smartphone adapter available in the market for other devices.

To capture retina images, iExaminer provides up to 25-degree FoV for dilated eyes and allows adjusting focus ranging from -20 to +20 diopter. It offers several apertures and filter options including small spot, large spot, microspot, slit aperture, red-free filter, cobalt blue filter, half-moon, and fixation aperture. Its optic design generates its own light, converges it to a point at the cornea, and diverges around the retina. This allows easy entry into small pupils and wide area illumination of the fundus. Therefore, it does not require pupil dilation for retinal imaging. The operator can also control the amount of illumination manually. It also has an eyecup at the patients' side that helps stabilization for the view and occludes ambient light to prevent the interference from outside light. The smartphone application allows capturing up to five images from a patient. After capturing retina images, the iExaminer system can send them to the doctor via an e-mail without applying any image processing algorithms.

To capture images with iExaminer, we attached the phone to the adapter on the PanOptic Ophthalmoscope. Since iExaminer uses its own light, we disabled the flashlight of the phone and set the carousel settings for large light in the ophthalmoscope. Then, we positioned the iExaminer on the synthetic eye model box and touched the eyecup to the box as shown in Figure 5(a). In order to use the auto-focus property, we set the diopter to zero and waited for a few seconds before capturing the images.



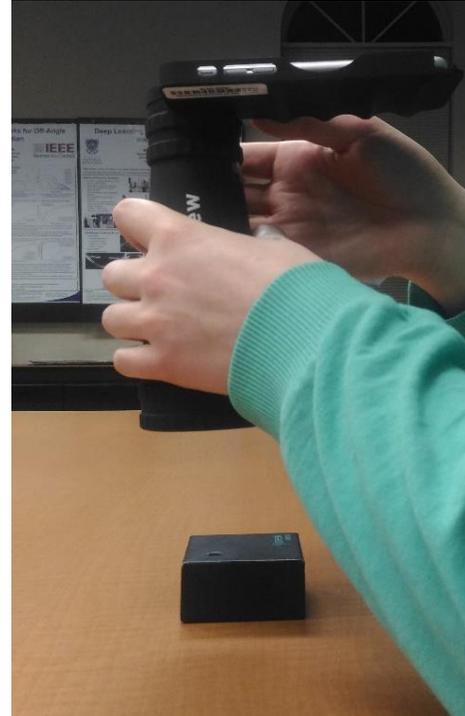
(a)



(b)



(c)



(d)

Figure 5. Experimental setup retinal imaging with (a) iExaminer, (b) D-Eye, (c) Peek Retina, and (d) Volk iNview.

### **3.1.3. D-Eye**

D-Eye retinal imaging system [7, 30] illuminates the retina by reflecting the smartphone flashlight to capture magnified retinal images up to 20 degrees in angle using its optics design and the smartphone camera. The D-Eye adapter is designed to attach it to various compatible iPhone models using compatible bumpers to magnetically attach. Since D-Eye reflects the flashlight of the smartphone to the retina using its optic design with mirrors and lenses, it does not require additional external power and light sources. Due to this design constraint, D-Eye is compatible with iPhone 5, 5s, 6, 6Plus, 6s, 6sPlus, and 7. D-Eye also provides an iOS application that gives the opportunity to both recording videos and shooting multiple images. Even if the system does not require pupil dilation to capture retina images, pupil dilation might be helpful where a larger FoV is needed such as analyzing retinal tissue structures and screening a complete retina. For undilated eyes, D-Eye system can capture retina up to a 6-degree angle of view, while dilated eyes allow the 20-degree angle of view. To capture images with D-Eye, we position the D-Eye on the synthetic eye model box as close as possible as shown in Figure 5(b).

### **3.1.4. Peek Retina**

Peek Retina is another plug-in imaging system [8] that has its own adjustable light and power source to illuminate the retina. Users can change the amount of light to illuminate the retina by choosing one level out of three. Since it has its own light source, Peek Retina has a universal clip to attach any smartphone. It also has own Android application, named Peek Retina Camera, that enables capturing photos and recording videos. However, we can use any camera application with iPhone and Windows Phone

models to capture images by adjusting manual settings in terms of autofocusing, clarity, and brightness.

Since Peek Retina system is still waiting for approval from the Food and Drug Administration in the US, it can only be used for educational purposes, not for medical purposes. However, there is a synthetic eye model box in its package for synthetic data capture. Therefore, we can only collect synthetic data provided in its package with eight different retina images that have different eye disease conditions from the normal eye to proliferative DR. For data capture with Peek Retina, we positioned the Peek Retina adapter to align with the camera and tighten the knob to hold the device in the correct position as shown in Figure 5(c). Since Peek Retina has its own light source, we disabled the flashlight of the smartphone.

### **3.1.5. Volk iNview**

The Volk iNview [9] smartphone-based retinal imaging system is developed by Volk Optical to capture wide-angle retinal images. It has three core components, including a mobile application, an indirect ophthalmoscopy lens, and an attachment adaptor for iPhone or iPod. This system captures the high-resolution retinal images using the camera and flashlight of the smartphone. The system does not require pupil dilation and it is able to capture 50 degrees of retinal view to visualize the entire posterior pole in a single image. In order to magnify the retina and to visualize it in a wider FoV, iNview Retinal Imager used 25D (diopter) lens as the primary imaging lens in its design. Its optical design generates a working distance of approximately 65mm. Due to the size of the 25D lens in the optical design, it has a large end tube size that requires the operator to use the patient's forehead as a steady reference to stabilize the device.

Volk iNview provides the widest FoV compared to other smartphone-based retinal imaging systems but it is also the biggest and heaviest design (see Table 1). Volk iNview is compatible with iPhone 5s, 6, 6s and iPod Touch. The free mobile application provides automatic retinal image capture and good auto-focus features. It also allows the user to encrypt the captured images using a password key for data security. This adds an extra layer of security to send the retinal images through an e-mail. Since iNview does not have its own light, we enabled the phone flashlight and positioned the iNview on the synthetic eye model box. In order to use the auto-focus property, we placed the iNview to about 150mm away from the box and make it closer to the box to set the distance around 65mm (the best distance for data acquisition) as shown in Figure 5(d). Waiting for a few seconds before data capture helps the auto-focus feature to stabilize the image.

## **3.2. Field of View (FoV) and Image Quality for Smartphone-based Retinal Imaging Systems**

### **3.2.1. Circular Test Patterns**

In order to compare the FoV and image quality and show the difference between smartphone-based retinal imaging systems, we set up our experiments using a circular test pattern as shown in Figure 6. We collected retina images and videos using four smartphone-based retinal imaging systems available on the market including iExaminer, D-Eye, Peek Retina, and Volk iNview. To make the fair comparison between different retinal imaging systems, we captured retinal images and recorded videos with iPhone 6 using their compatible adapters and bumpers. We first attached each smartphone-based retinal imaging system to iPhone 6 to capture retina images. Since the retina is located at the back of the eye and light does not reflect back through the retina, it is not possible to take a

picture of the retina directly without any external illumination. Therefore, we illuminated the dark retina by reflecting the smartphone flashlight or using the light source on the smartphone-based retinal imaging system. The amount of light reflected on the retina is adjusted using the mobile application or controller on the smartphone-based retinal imaging system. The retina images were captured using the smartphone's camera and saved to its memory.



Figure 6. The circular test pattern of the Peek Retina synthetic eye model box

Even if in a controlled environment, pupil dilation level of the human subject and the eye gaze between a human subject and the retinal imaging system changes easily. Since the visible portions of the retina will change, the retina images captured from a human subject changes over time for different smartphone-based retinal imaging systems. To fix the visible retina, a synthetic eye model can be used. For our experimental setup, we used

the synthetic eye model box provided with Peek Retina smartphone-based retinal imaging system. It replicates the eye structures including pupil, lens, and retina. The dimensions of the synthetic eye model are 53 mm x 53 mm x 22 mm in width, length, and depth. The opening on the box replicates the pupil of the eye. Under the pupil opening, there is a lens to illustrate the lens of the eye. Since the distance from the front surface of the cornea to the retina is approximately 24 mm on average, a printed test image is placed inside the box at the bottom that gives the same distance.

In our experiments, we used a circular test pattern to capture images using different smartphone-based retinal imaging systems. Therefore, we placed the circular test pattern at the bottom of the Peek Retina synthetic eye model box as shown in Figure 6. With this set of experiments, we compared the FoV, illumination distribution, depth of focus blur, environmental setup for each smartphone-based retinal imaging system.

Before comparing the smartphone-based retinal imaging systems, we first investigated the data collection environment to figure out what condition is more suitable for capturing retina images such as ambient light in the environment, mobile application, auto-focus, zoom, etc. For example, we captured retina images in an office with different light setup and in outdoor to test the effect of ambient light in the environment. We captured images with four smartphone-based retinal imaging systems using different mobile applications such as iPhone's camera app, Pro-Movie for video recording app, and other apps in App Store by changing the focus and zoom settings. Based on our observations, even if there might have small variations in each system due to its own features, the best experimental setup for data collection, in general, is to use iPhone's camera app with the flashlight on and auto-focus in a dark office environment.

### **3.2.2. Best Distance Calculation for Environment Setting**

After finding the best setup, we investigated the effect of the distance between the retina and the smartphone-based retinal imaging system. We captured retina images and videos at a different distance including 150mm, 100mm, 75mm, 65mm, 50mm, 35mm, 30mm, and 22mm. For example, when the distance between retina and imaging system is fixed at 22mm, it means that the imaging device touching to the synthetic eye model box. To find the optimum distance between the imaging system and the eye, we compared the FoV and depth of field blur by counting the number of visible circles, and the sharpness of each circle in the captured images from different distances.

Figure 7 shows the images captured with iExaminer using circular test patterns from 150mm, 100mm, 75mm, 65mm, 50mm, 35mm, 30mm, and 22mm in distance. We observed that the number of circles increases from one to six as the distance between the retina and the imaging device decreases.

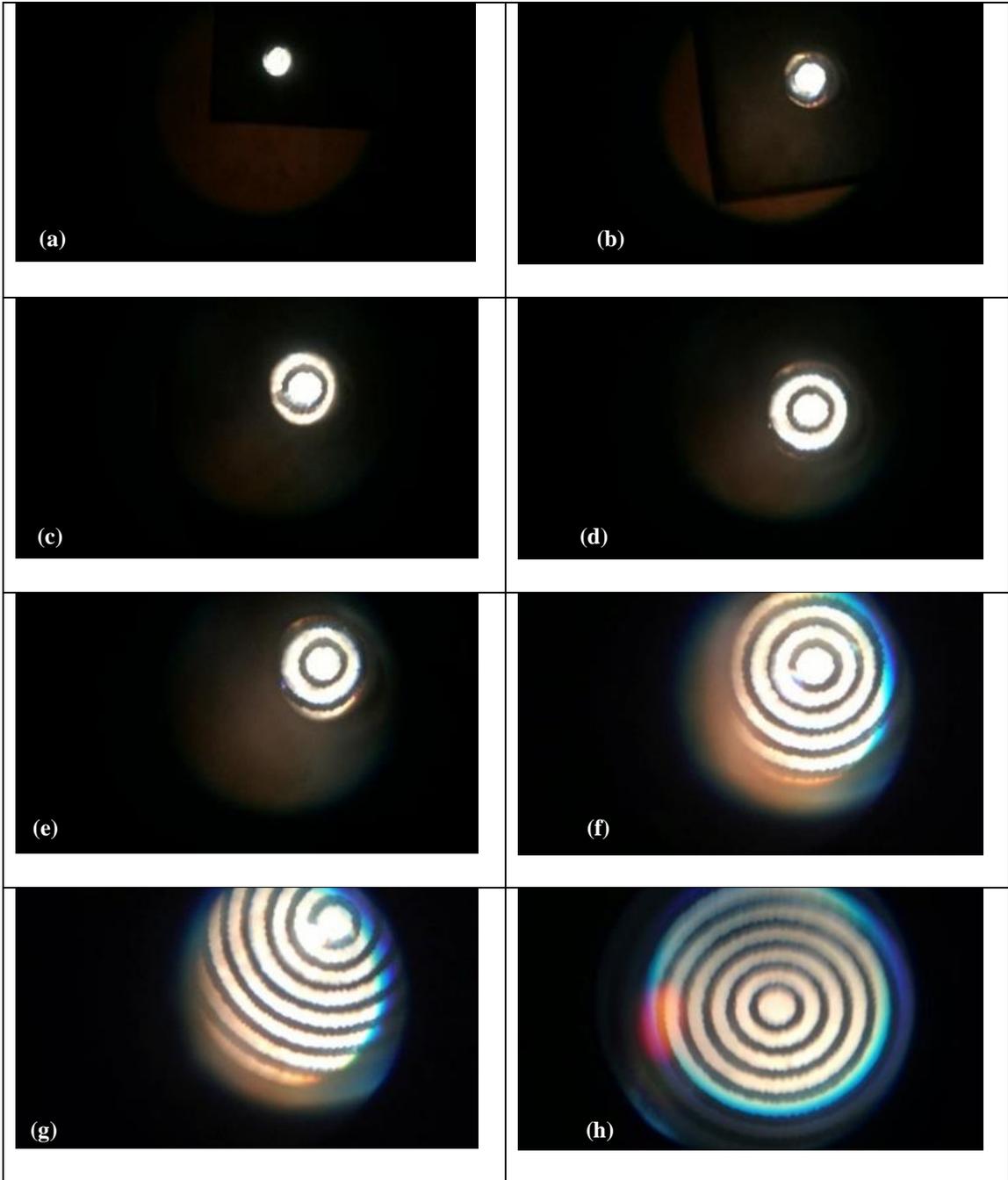


Figure 7. Circular test pattern images captured by iExaminer from (a) 150mm, (b) 100mm, (c) 75mm, (d) 65mm, (e) 50mm, (f) 35mm, (g) 30mm, (h) 22mm in distance.

Using D-Eye, we captured the circular test patterns from 150mm, 100mm, 75mm, 65mm, 50mm, 35mm, and 22mm in the distance as shown in Figure 8. The number of

visible circles in the test pattern increases when the imaging system is closer to the synthetic eye model box. There are seven visible circles at the closest 22mm distance.

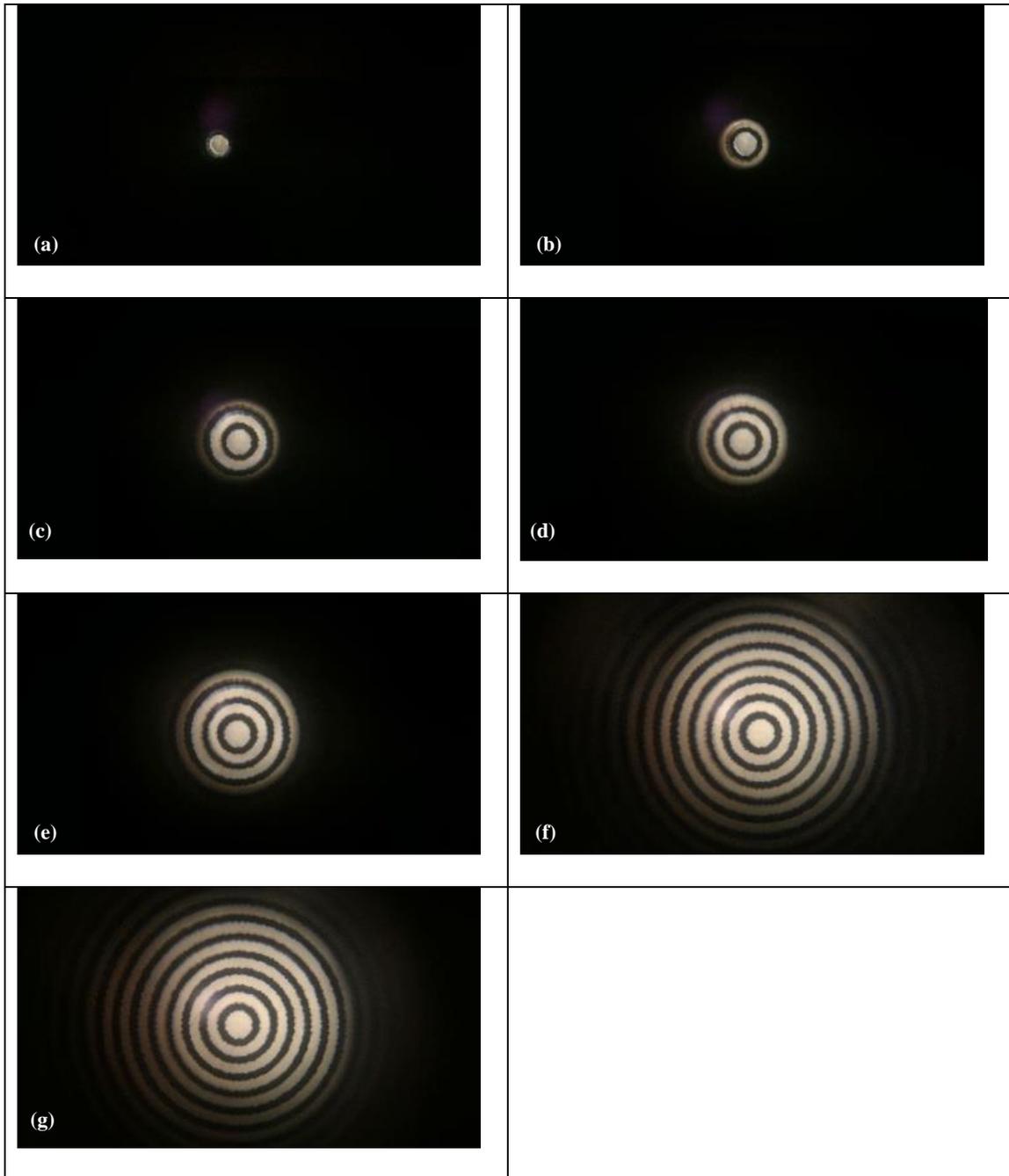


Figure 8. Circular test pattern images captured by D-Eye from (a) 150mm, (b) 100mm, (c) 75mm, (d) 65mm, (e) 50mm, (f) 35mm, (g) 22mm in distance.

Figure 9 shows the images captured with Peek Retina using circular test patterns from 150mm, 100mm, 75mm, 65mm, 50mm, 35mm, and 22mm in distance. We observed that the number of circles increases from one to seven as the distance between the retina and the imaging device decreases.

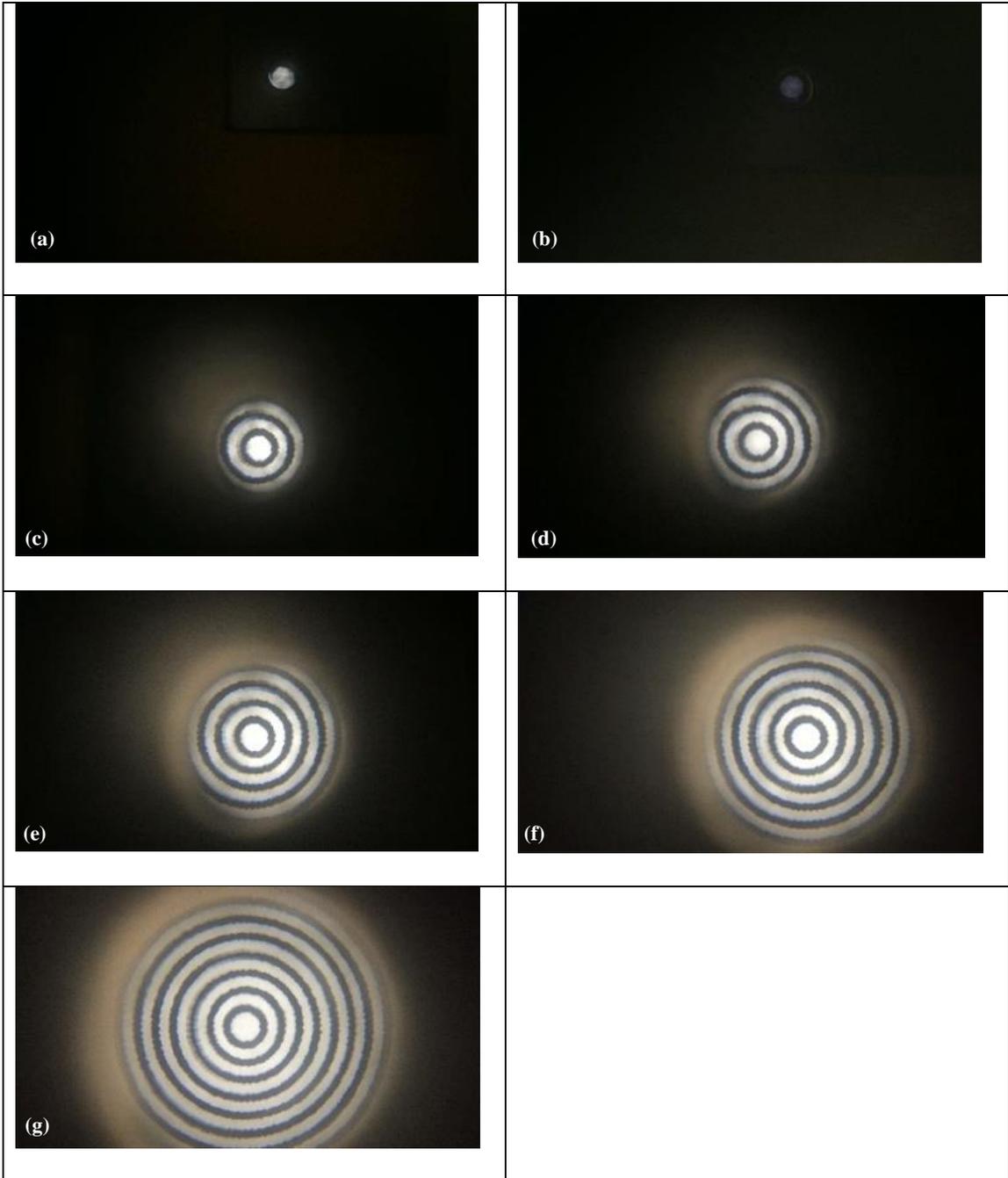


Figure 9. Circular test pattern images captured by Peek Retina (a) 150mm, (b) 100mm, (c) 75mm, (d) 65mm, (e) 50mm, (f) 35mm, (g) 22mm in distance.

Finally, we captured the circular test patterns using iNview from 150mm, 100mm, 75mm, 65mm, 50mm, 35mm, 30mm, and 22mm in the distance as shown in Figure 10.

Unlike other retinal imaging systems, a number of visible circles in the circular test pattern first increases then decreases when the imaging system is closer to the synthetic eye model box. The maximum number of circles is from 65mm where 10 to 15 circles are visible. This is the best seen for data acquisition distance for iNview.

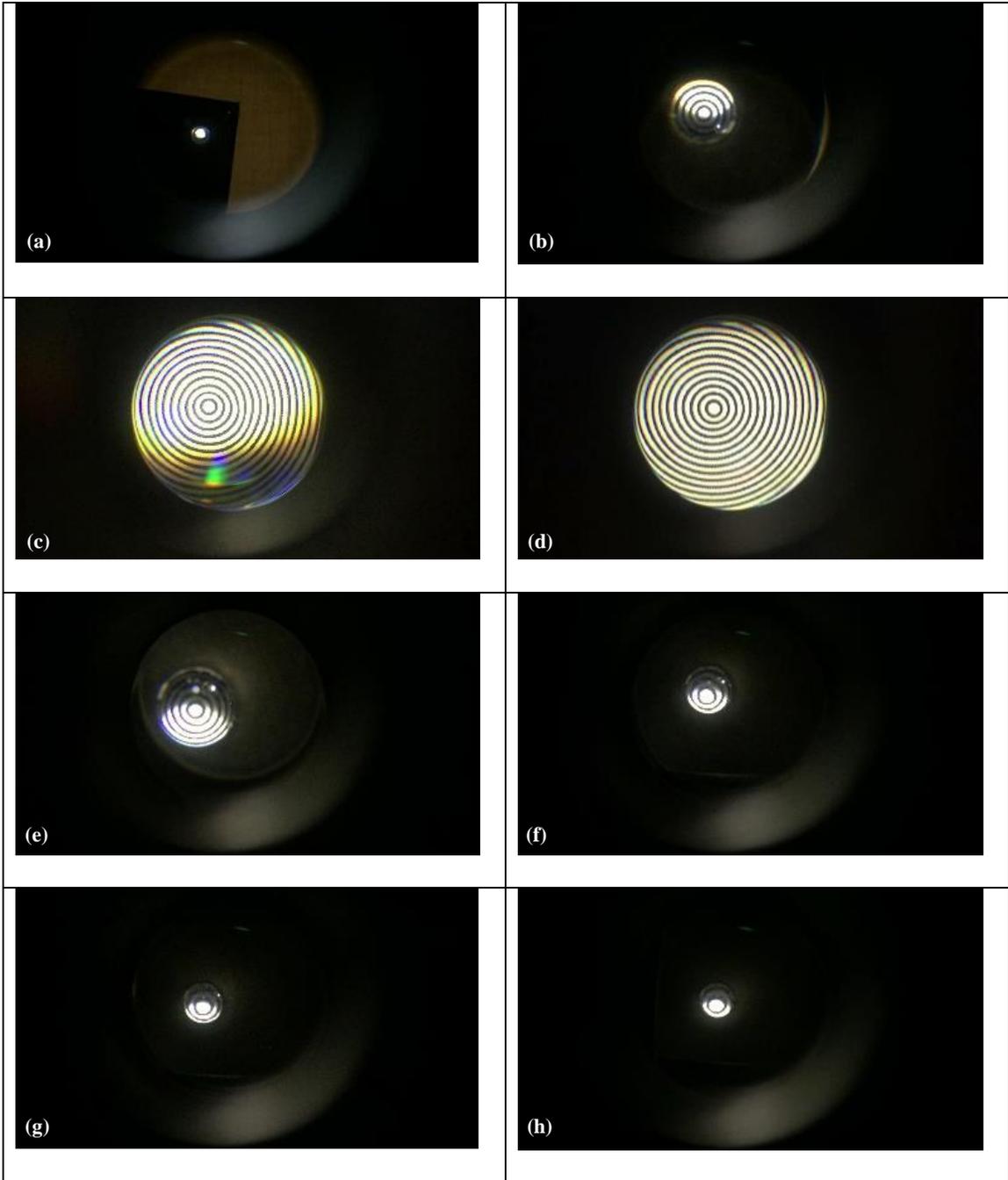


Figure 10. (a) Circular test pattern images captured by iNview from (a) 150mm, (b) 100mm, (c) 75mm, (d) 65mm, (e) 50mm, (f) 35mm, (g) 30mm, (h) 22mm in distance.

To compare the FoV of the retinal imaging systems, Figure 11(c-f) shows the captured circular test pattern images by iExaminer, D-Eye, Peek Retina, and iNview at

their best distance, respectively. Even if there are 15 circles in the test pattern, none of the smartphone-based retinal imaging systems can see the entire test pattern.

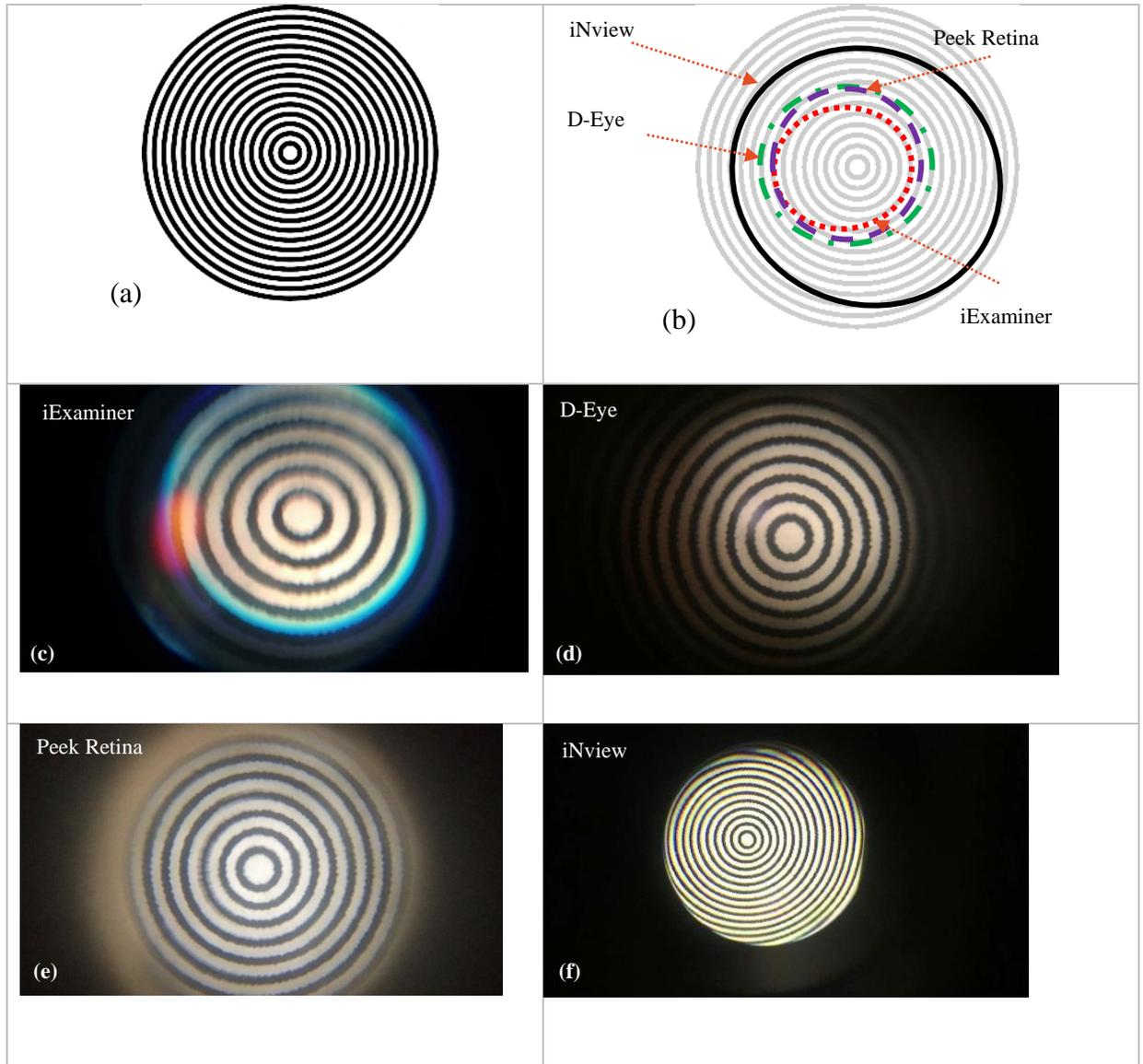


Figure 11. (a) Circular test pattern image (b) Comparison of the FoV of each smartphone-based retinal imaging system where the solid black line is for iNview, the green dashed-dotted line is for D-Eye, the purple dashed line is for Peek Retina, the red dotted line is for iExaminer. Captured images of circular test patterns at the best distance by (c) iExaminer, (d) D-Eye, (e) Peek Retina and (f) iNview.

Figure 11(b) shows the marks for the FoV of each imaging system. iNview has the largest FoV compared with others where its FoV is marked with solid black line ranging from 10 to 15 circles. Due to its optical design, its FoV is shifted to the right. It has a very homogenous illumination distribution. D-Eye has the second-largest FoV with 7-8 visible circles that are marked with a green dashed-dotted line. However, its brightness in the captured image does not distribute evenly where it gradually decreases from the center to sides. Peek Retina has almost the same size FoV with D-Eye where seven circles are visible that is marked with a purple dashed line. However, its brightness distribution is better than D-Eye where its distribution is almost homogeneous. The smallest FoV belongs to iExaminer where it is only possible to see six circles marked with a red dotted line. In addition, there exists a color artifact at the edge of the captured image and the sharpness of the image also decreases at the larger circles close to the boundaries.

Based on the results, iNview retinal imaging system has the largest FoV and better image quality compared with iExaminer, D-Eye, Peek Retina retinal imaging systems. D-Eye and Peek Retina has a similar size in the FoV that is half of iNview. iExaminer has the smallest FoV. The best FoV is reached when images captured the closest distance (22mm) for iExaminer, D-Eye and Peek Retina. However, iNview captures the best quality images at 65mm in distance.

### **3.3. Deep Learning Architectures for Diabetic Retinopathy Detection**

This section provides the layout of the utilized deep learning framework. For image classification, Convolution Neural Network (CNN) is one of the most popular deep learning frameworks so this work adopts the CNN based AlexNet [31, 32] and GoogLeNet [33] using the transfer learning.

### 3.3.1. AlexNet

Due to its simple design and high classification accuracy compared with other CNNs, AlexNet is very popular in different research communities and used in several applications. Since 1.2 million high-resolution images are used to train AlexNet, it classifies images into 1000 different classes with a very low error rate.

AlexNet first uses five convolutional layers to extract low-level features. Then, it maps the final features to the set of the predetermined number of classes using three fully connected layers and a softmax layer. In order to transfer the network knowledge, we use the general purpose features learned previously to retrain the softmax layer on a different dataset with different classes. In each layer, network architecture contains different layers such as convolutions, fully connected, activation functions, dropouts, and max pooling as can be seen in Table 2. More complex features are extracted gradually in each convolution layer using the features from previous layers in the network. To introduce nonlinearity into the network, we use Rectified Linear Units (ReLU) as an activation function after each convolutional layer and fully-connected layer. Max pooling layers are used to keep the number of parameters required to be learned by outputting the maximum value of each square kernel and discarding the rest. This helps to reduce the redundancies among the features learned by the network. Dropout layers are introduced after the last two ReLU activations during training to force the network learning more robust features from training samples and avoid the overfitting problem. Using the output of the last fully connected layer, softmax classifies images into different classes based on the highest probability.

Table 2. AlexNet Architecture

Name	Type/Stride	Description
Data	Input Image	227x227x3 images
conv1	Convolution/4	96 11x11x3 convolutions
relu1	ReLU	ReLU
norm1	Normalization	Cross channel normalization
pool1	Max Pooling/2	3x3 max pooling
conv2	Convolution/1	256 5x5x48 convolutions
relu2	ReLU	ReLU
norm2	Normalization	Cross channel normalization
pool2	Max Pooling/2	3x3 max pooling
conv3	Convolution/1	384 3x3x256 convolutions
relu3	ReLU	ReLU
conv4	Convolution/1	384 3x3x192 convolutions
relu4	ReLU	ReLU
conv5	Convolution/1	256 3x3x192 convolutions
relu5	ReLU	ReLU
pool5	Max Pooling/2	3x3 max pooling
fc6	Fully Connected	4096 fully connected layer
relu6	ReLU	ReLU
drop6	Dropout	50% dropout
fc7	Fully Connected	4096 fully connected layer
relu7	ReLU	ReLU
drop7	Dropout	50% dropout
fc8	Fully Connected	52 fully connected layer
Prob	Softmax	softmax
Out	Output	Classification

Since AlexNet contains nearly 60 million parameters in its architecture, training it from scratch requires a very large amount of input images and computational power. When we do not have millions of labeled images, the best approach is to use transfer learning. In this paper, we removed the last three layers of the AlexNet (i.e., 1000-class fully connected, softmax, and classification). Then, we replaced them with a new two-class fully connected layer, softmax, and classification layers for transfer learning. Finally, we retrained the new network with the retina images from the EyePACS dataset using Stochastic Gradient Descent (SGD) algorithm with a minibatch size of 2, 4, 8 examples, the learning rate of  $1e-5$ , and a momentum of 0.9. During the retraining, the network updates the parameters of each layer for every iteration with each training sample.

### **3.3.2. GoogLeNet**

GoogLeNet framework gets popularity from different research communities and applied for different applications. GoogLeNet is trained on ImageNet [34] with millions of high-resolution images to classify them into 1000 different classes such as a keyboard, mouse, and many species of animals with a very low error rate by tuning nearly seven million parameters doing 1.5 billion operations.

This network consists of 144 layers in MATLAB which 57 convolutional layers followed by activation layers and a softmax layer to map the final features to the set of classes of interest. For knowledge transfer, this softmax layer can be retrained on a different dataset with different classes using the general-purpose features learned up to this layer. Alternatively, these features can be fed into other machine learning classifiers such as Naïve Bayes, Random Forest, and Support Vector Machines (SVM). Each convolution layer extracts gradually more complex features from a larger area using features from its

deeper layers to the network. As an activation function, Rectified Linear Units (ReLU) are used after each convolutional layer and fully-connected layer to introduce nonlinearity into the model. Some RELU functions are followed by a depth concatenation and normalization layer to preserve the generalization capabilities of the network by regularizing the responses. To lower the number of parameters to learn, max-pooling regularization layers are used; it outputs the maximum value of each square kernel discarding the rest, thus reducing the redundancies among the features learned by the network.

For every training sample, the network updates the weights and biases. To force the network to learn more robust features from inputs by randomly setting the output of each hidden neuron to zero with a certain probability to contribute to the forward pass and back-propagation, dropout layers follows the first two ReLU activations. While testing the network all units are used without dropout. Using dropout layers also helps to reduce the risk of overfitting. The image classification is performed at the softmax layer using the output of the last fully connected layer. It is based on the highest probability among different classes.

Since training a CNN network from scratch is overwhelming that requires a very large amount of input and computational power, we adopt the idea of using transfer learning in this work. Before transferring the network, we first freeze 110 layers and removed the last three layers from the GoogLeNet including the last fully connected (fc8), softmax (prob), and classification (output) layers and replaced with two fully connected layers, softmax, and two classification layer. The new network is retrained with the retina images using Stochastic Gradient Descent (SGD) algorithm with a minibatch size of 4, 8, 16 examples, the learning rate of  $1e-5$ , and a momentum of 0.9. The number of max epoch

number in experiments was set to 50, 75, and 100 depending on the number of images in training and validation sets.

For all our experiments, we developed our programs and CNN frameworks using the MatConvNet [35] and image processing toolboxes in MATLAB 2018. The experiments are performed on a DELL Alienware Aurora R8 workstation with 8 core Intel Core i7 9700K processor at 4.6GHz, NVIDIA GeForce RTX 2080 with 8GB GPU, and 32GB memory.

## 4. EXPERIMENTAL SETUP AND DATASETS

### 4.1. Datasets

To investigate the automatic DR detection accuracy for smartphone-based retinal imaging systems and compare them with traditional fundus imagery, we set up two sets of experiments using synthetic and original retina images. For our experiments, we used several publicly available retina image datasets including EyePACS [36], Messidor [37], Messidor-2 [38], IDRiD [39], and University of Auckland Diabetic Retinopathy (UoA-DR) dataset [40, 41]. EyePACS is the largest publicly available dataset that offered during Kaggle's competition with 35126 retina images that include five different DR severity labels. IDRiD has 271 retinal images and its DR severity assigned based on five classes. Messidor DR database images have four labels and their ground truths available with Diabetic Macular Edema grades [42]. Messidor-2 dataset is an extension of Messidor with its ground truth labels. UoA-DR has 200 retina images and has detailed DR and DME severity scales as well as information about neovascularization, hemorrhage, and microvascular abnormalities. We converted all this information to regular five different DR categories. Table 3 shows detailed information for each data set that we have used in this study.

Table 3. Retina Image Datasets with DR Severity Labels

<b>Datasets</b>	<b>Label0</b>	<b>Label1</b>	<b>Label2</b>	<b>Label3</b>	<b>Label4</b>	<b>Total</b>
<b>EyePACS</b>	25810	2443	5292	873	708	35126
<b>EyePACS-u</b>	9895	899	2175	395	260	13624
<b>Messidor</b>	547	149	240	251	-	1187
<b>Messidor-2</b>	1368	-	-	-	380	1748
<b>IDRiD</b>	168	25	168	93	62	271
<b>UoA-DR</b>	56	9	50	55	30	200

Retina images in these datasets are graded according to the International Clinical DR scale [43] except Messidor and Messidor-2. This scale classifies the retina images into five classes including none, mild DR, moderate DR, severe DR, and Proliferative DR. Abramoff [44] graded Messidor-2 retina images based on the rDR standards that apply for the moderate, severe, or proliferative DR level and/or rDME. Table 4 shows the available data labels in each dataset. The image acquisition method was different for most of the datasets where images are captured using Canon, Centervue DRS, Optovue iCam, Topcon NW cameras, and pupil dilation levels might be different for each image.

Table 4. Label Assignments for DR Severity

Datasets	Label0	Label1	Label2	Label3	Label4
<b>EyePACS</b>	No DR	Mild	Moderate	Severe	Proliferative
<b>Messidor</b>	No DR	Mild	Moderate Severe	Proliferative	-
<b>Messidor-2</b>	No DR	-	-	-	Others
<b>IDRiD</b>	No DR	Mild	Moderate	Severe	Proliferative
<b>UoA-DR</b>	No DR	Mild	Moderate	Severe	Proliferative

Since some of the images include darkness, reflections, lack of contrasts, and even lack of optic nerve, we needed to preprocess the images, especially for the EyePACS dataset. For the data preprocessing stage, we removed the images from the data set if the optic disk could not be detected. In addition, since 90% of the dataset consists of healthy retina images, it causes a bias in the network retraining. Therefore, 21,502 images removed from the original EyePACS data set. Besides, according to the new findings, Messidor data set updated since there were some grading inconsistencies so that the number of images decreased to 1,187 images by removing 13 images from the dataset.

Originally, there are four different DR classes in Messidor datasets and five classes in the other datasets. However, the same person can make mistakes when classifying images into five different groups especially for the mild DR and moderate DR. In order to remove the inconsistencies and transfer the problem into the easier two-class domain, we conducted our experiments by classifying two different classes. For that purpose, we tested different options including (1) healthy retina vs proliferative DR, (2) healthy retina vs severe and proliferative DR, (3) referable diabetic retinopathy (rDR), and (4) vision-

threatening DR (vtDR). Referable DR (rDR) consists of the images for normal retina vs Moderate DR, Severe DR, Proliferative DR, ungradable images and referable DME while vtDR is dropped out moderate DR and ungradable images from the rDR.

## 4.2. Dataset Preparation

The original retina images in the dataset are originally color images and their resolutions vary since they captured by different fundus cameras. However, GoogLeNet Deep Learning framework requires the inputs to have 224x224x3 pixels as color images. Therefore, we first cropped pixels from the right and left sides of each original image to make it a square shape. Then, we down-sampled the cropped square images into 224x224x3 as can be seen in Figure 12.

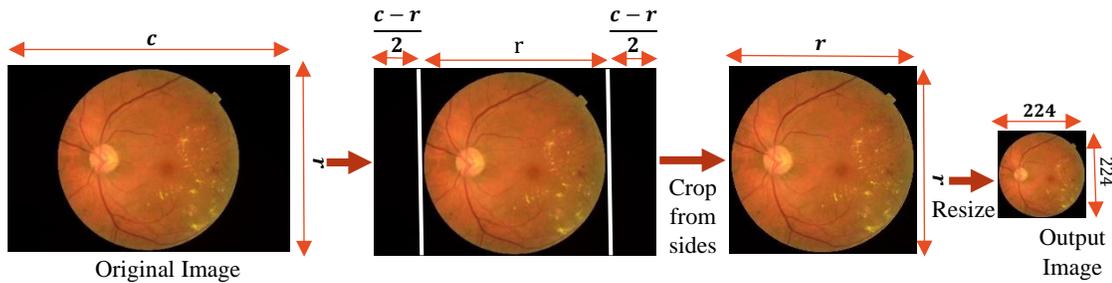


Figure 12. Flow chart of original retina conversion images for GoogLeNet

In order to train and test the deep learning network, we need to feed the CNN frameworks with retina images captured by different smartphone-based devices. However, there is no available real data captured by the smartphone-based retinal imaging device in the literature. Therefore, we generate retina images by simulating the FoV for each device using the original retina images from the UoA-DR dataset. We generate a circular mask around the optic nerve center based on the different ratios of FoV ranging from 20% to

100% 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.

Figure 13 shows each circular mask representing the different FoV to compare the difference in smartphone-based retinal imaging systems. Finally, the original image is cropped with respect to the mask. Examples of generated smartphone-based synthetic retina images are shown in Figure 14 for different FoV.

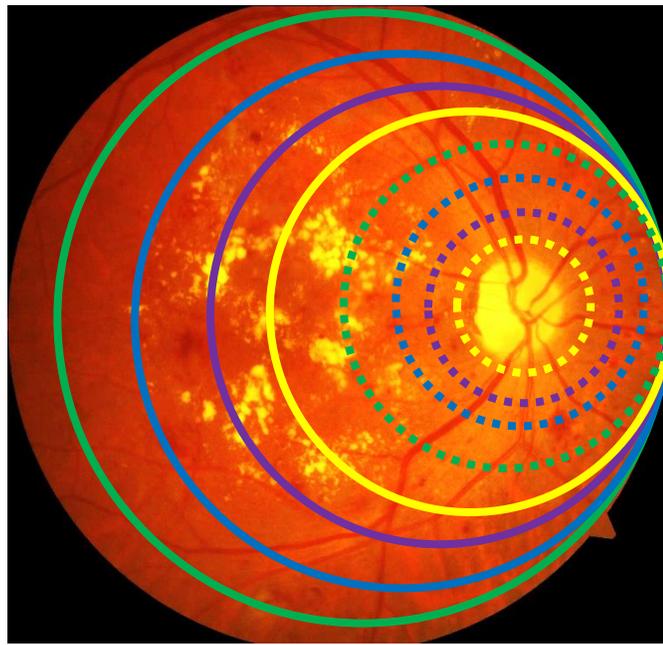


Figure 13. Comparison of the FoV of synthetic retina images with different percentages with respect to the original image where the solid green, blue, purple, and white lines represent 90%, 80%, 70%, and 60% FoV, respectively. The dotted green, blue, purple, and white lines represent 50%, 40%, 30%, and 20% FoV, respectively.

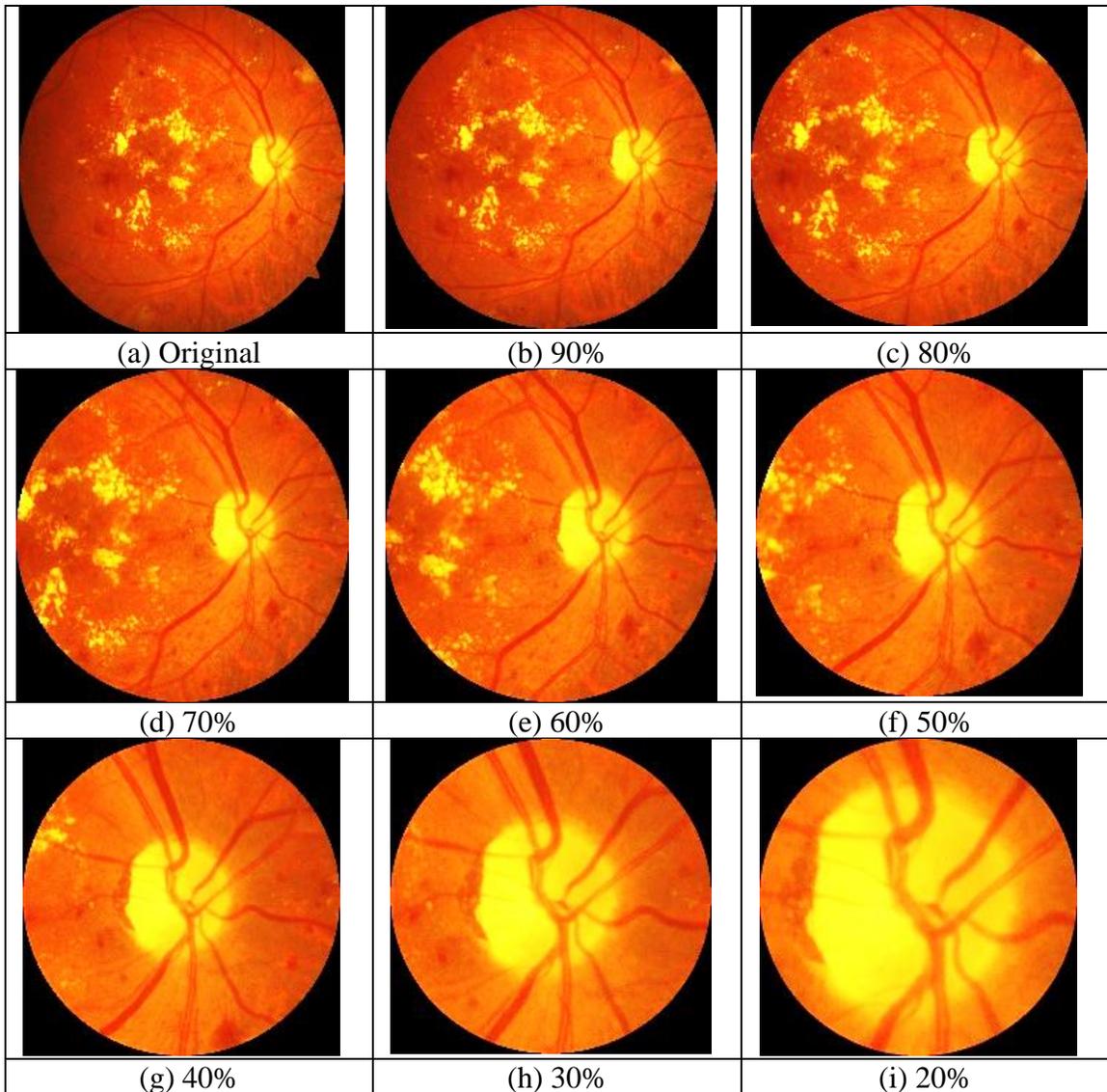


Figure 14. Input images for deep learning architecture with (a) original size and synthetic images with a smaller FoV where their percentage w.r.t the original image as (b) 90%, (c) 80%, (d) 70%, (e) 60%, (f) 50%, (g) 40%, (h) 30%, (i) 20%.

### 4.3. Dataset Augmentation

Data augmentation is a very useful technique to prevent bias especially for the deep learning algorithms that require large datasets. Therefore, it is crucial especially for a small number of training datasets. By performing augmentation, researchers acquire more data

by applying affine transformations, hue shifts, and cropping to the existing images. Some of the data augmentation operations including but not limited to filling value, hue shifts, cropping, random rotation, random reflection, random scaling, random shearing, and random translation. However, most of these operations more useful when applying to object recognition tasks. For example, when finding a human, the algorithm might need to take look at the different locations to images. A person can be any part of the image even just human head can be found in the images. For these tasks, most of the listed data augmentation is essential for object recognition. However, our specific DR detection task needs to hall retina image especially the surroundings of the optic disc, fovea, and macula so that we did not apply most of the augmentation process since we did not want to lose the lesions that include hemorrhages, exudates, and microaneurysms. Hence, we applied data augmentation process just flipping the images, in other words, obtaining the mirror images for each image sample.

#### **4.4. Data Merging and Splits for Training and Validation**

For DR detection performance analysis of the deep learning framework, we designed several experiments using seven different combinations of retina datasets in training and testing. We feed the GoogLeNet network and test it with images from single, crossed and merged datasets. Table 8 shows details of these seven sets of experiments with (1) Train and validation with EyePACS, (2) Train and validation 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 Merged dataset (EyePACS, Messidor, Messidor-2, and IDRiD), and (7) vtDR detection with

Merged dataset (EyePACS, Messidor, and IDRiD). Note that non-healthy images (Label 1) are used from the Messidor-2 dataset for (6).

After data conversion, data augmentation, and dataset generation, we split our retina images in each dataset into two categories: training and validation sets with a ratio of 0.8. The training sets include a maximum of 2,000 images for each experiment.

## 5. RESULTS AND DISCUSSION

### 5.1. Diabetic Retinopathy Detection using Single Dataset

To compare the smartphone-based retinal imaging systems, we set up two sets of experiments using synthetic retina images and real fundus images. We first collected retina images using iExaminer, D-Eye, Peek Retina, and iNview. To make the fair comparison between different retinal imaging systems, we captured retinal images with iPhone 6 using their compatible adapters and bumpers. In order to capture an image of the dark retina, we first needed to use a light source for illumination. For this purpose, we reflected the smartphone flashlight to the retina for D-Eye and iNview or use its own light source on iExaminer and Peek Retina. The smart phone's camera was used to capture retina images and images were saved to the smartphone's memory.

#### 5.1.1. Results for Synthetic Retina Box Images

In our first set of experiments using a single dataset (EyePACS), we captured images from the printed retina in the retina box using different smartphone-based retinal imaging systems. We placed the printed real retina images at the bottom of the Peek Retina synthetic eye model box. Then, we captured retina images from the optimum distance of each device in order to get the largest FoV. With this set of experiments, we compared the image quality and FoV of smartphone-based retinal imaging systems.

Figure 15 shows the captured printed real retina images using iExaminer, D-Eye, Peek Retina, and iNview. The original printed real retina images are shown in the first row. The image at the top left corner is a normal retina image where there is no abnormality visible. The image in the middle of the top row belongs to a DR patient where there is

clinically significant macular edema. The image on the right of the top row belongs to a proliferative DR patient where there are several visible abnormalities. The rest of the images on each row are captured using smartphone-based retinal imaging systems from retina box for each original printed real retina images in the first row. Images in the second row are captured using iExaminer. We observed that the image quality is good, and the optic nerve and some central blood veins are visible. Since some macula edema can be seen, these images can be used for DR detection. Images on the third row were captured using D-Eye where the illumination does not distribute evenly. However, we can still visualize the macular edema, so they are helpful for DR detection. Images on the fourth row of Figure 16 shows the image captured using Peek Retina. We observed that the FoV of Peek Retina is larger than iExaminer. The optic nerve and its surrounding vessels are visible, so these images can be used to detect DR. The images at the last in Figure 16 are captured using iNview. We observed that its FoV is the largest compared with other smartphone-based retinal imaging systems. The image quality is very good with evenly distributed illumination.

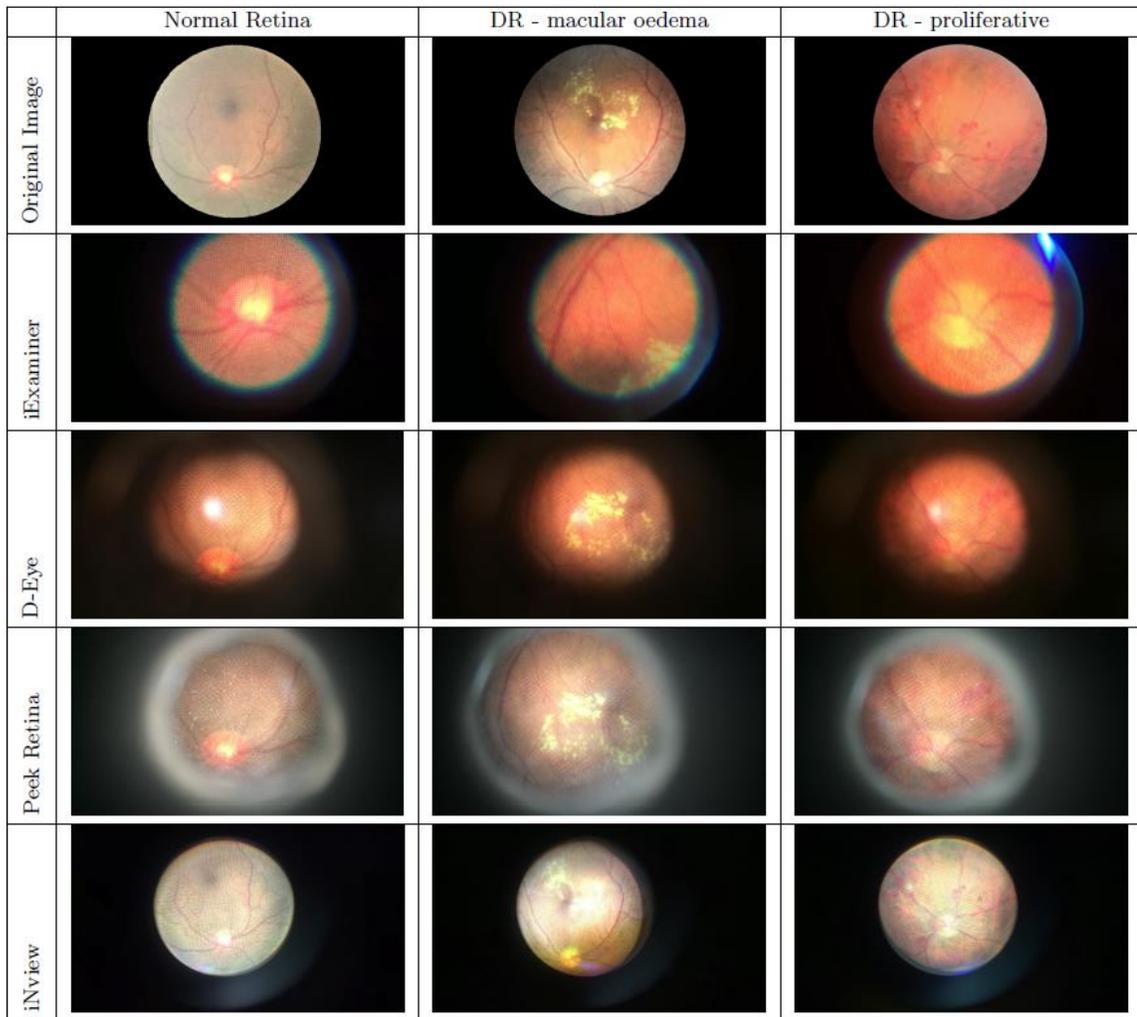


Figure 15. Retina images. (1st row) Original printed real retina images and their captured versions using (2nd row) iExaminer, (3rd row) D-Eye, (4th row) Peek Retina, (5th row) iNview.

To compare the FoV of the retinal imaging systems into one shape, Figure 16 shows the captured printed retina images using iExaminer, D-Eye, Peek Retina, and iNview, respectively. Even if the printed retina image shows a larger portion of the retina, none of the smartphone-based retinal imaging systems can capture the entire printed retina pattern. Figure 17 shows the printed image from a normal retina with the markers for the FoV of each imaging system. We observed that the radius of FoV for iExaminer, D-Eye, Peek

Retina, and iNview are 32%, 40%, 45%, and 94% of the radius of printed fundus image, respectively. iNview has the largest FoV compared with others where its FoV is marked with the solid black line. D-Eye has the second-largest FoV that is marked with a green dashed-dotted line. Peek Retina has almost the same size FoV with D-Eye marked with a purple dashed line. However, their FoV is almost half of the iNview. The smallest FoV belongs to iExaminer, where it is only possible to capture the fovea and its surroundings that is marked with a red dotted line.

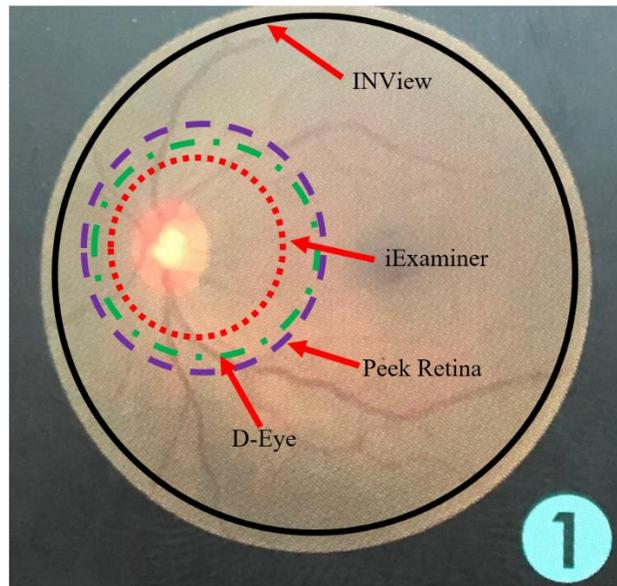


Figure 16. Comparison of the FoV of each smartphone-based retinal imaging system using printed normal retina image with FoV markers where the solid black line is for iNview, the purple dashed line is for Peek Retina, the green dashed-dotted line is for D-Eye, the red dotted line is for iExaminer.

### 5.1.2. Results for Smartphone-based Images

In our second set of experiments using a single dataset, we design several experiments for each smartphone-based retinal imaging device to analyze the DR detection

performance of the deep learning framework for smartphone-based systems and compare them with the accuracy of the original retina images. The original retina images in the dataset are originally color images and their resolutions vary since they captured by different fundus cameras. However, AlexNet Deep Learning framework requires the inputs to have  $227 \times 227 \times 3$  pixels as color images. Therefore, we first cropped pixels from the right and left sides of each original image to make it a square shape. Then, we down-sampled the cropped square images into  $227 \times 227 \times 3$  as shown in Figure 17(a).

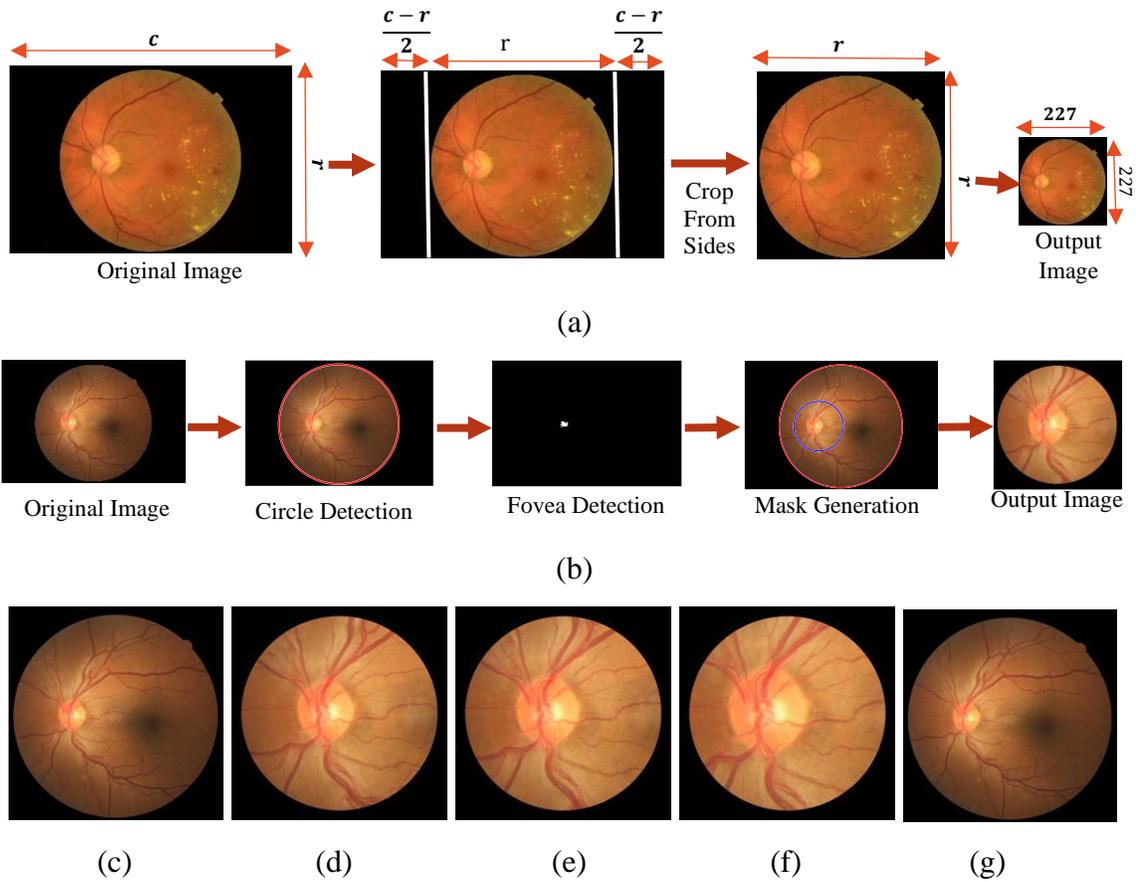


Figure 17. (a) Flow chart of original retina conversion images for AlexNet, (b) Flow chart of synthetic retina image generation for smartphone-based retinal imaging systems. Example of images, (c) Original image from EyePACS dataset and generated synthetic images for (d) iExaminer, (e) D-Eye, (f) Peek Retina, (g) iNview.

In order to train and test the deep learning network, we need to feed the CNN frameworks with retina images captured by different smartphone-based devices. However, there is no available real data captured by the smartphone-based retinal imaging device in the literature. Therefore, we generate retina images by simulating the FoV for each device using the retina images from EyePACS dataset as illustrated in Figure 17(b). We first segment the boundary of the original retina images in EyePACS and fit a circle to its boundary using a circular Hough transform. Second, the optic nerve is detected using thresholding and morphological filters since the optic nerve is the brightest area in the retina. Then, we generate a circular mask of each smartphone-based device using the center of the optic nerve as a circle center. The mask radius is calculated by multiplying the radius of the original image boundary and the percentage of the radius of FoV in Figure 6. Finally, the original image is masked and cropped with respect to the mask center. Examples of synthetic retina images for iExaminer, D-Eye, Peek Retina, iNview are shown in Figure 17(c-g), respectively.

In the original image dataset, we have 13,624 retina images from five different gradings (i.e., 0:No DR, 1:Mild, 2:Moderate, 3:Severe, and 4:Proliferative DR) where their distributions are 9898, 920, 2,108, 424, and 262, respectively. In order to simplify the problem, we dropped the grade 1 and 2 and merged grade 3 and 4. Since no DR images are more than other labels in the dataset, we randomly selected 686 no DR images and removed the rest from the dataset in order to remove the data bias. As a result, we had 686 no DR and 686 DR images remaining in each subset as shown in Figure 18. Then, we split our customized dataset with a ratio of 1/9 into two subsets where the testing subset has 1,234 images and the training set has 138 images. We selected the bigger subset of original retina

images as a training set and the second subset from each synthetic retina subsets are used as testing sets. After training the network with 1,234 original retinal images, we tested the network with 138 retinal images from each dataset including original, iExaminer, D-Eye, Peek Retina, and iNview images. In total, we used 690 retinal images from five different datasets.

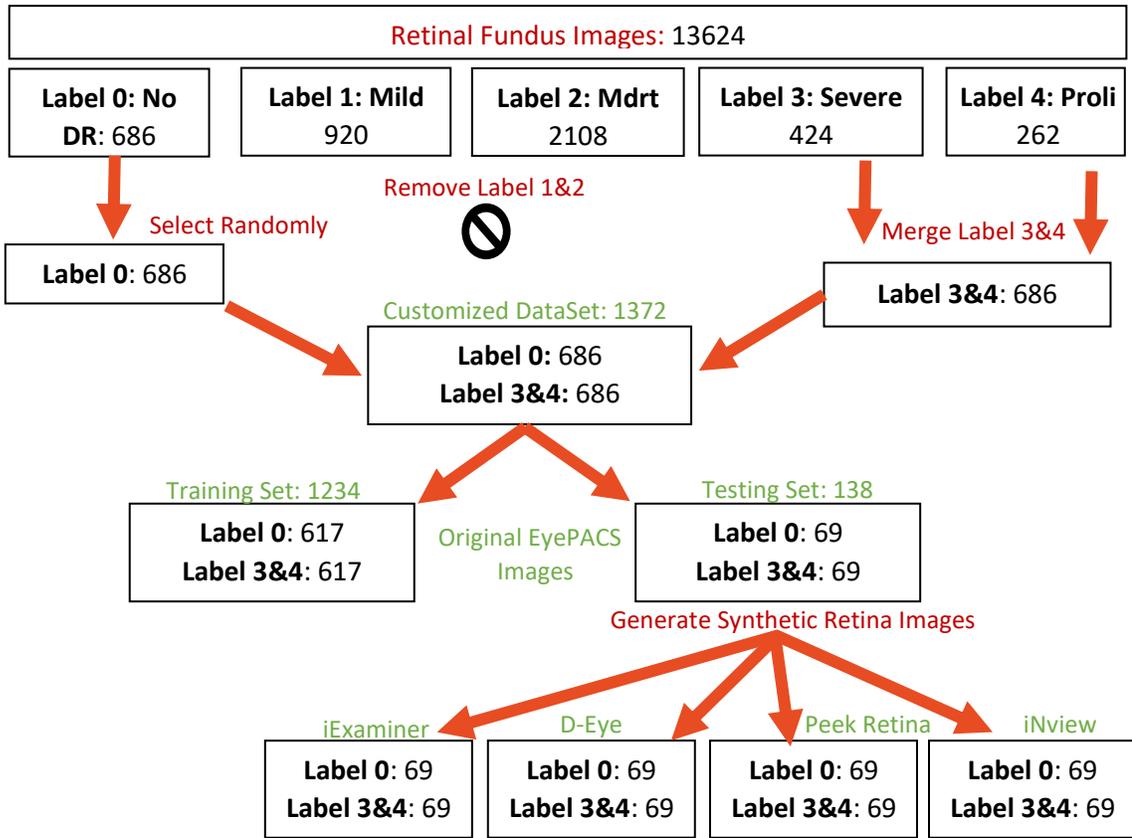


Figure 18. Generation of the customized dataset of retinal fundus images and its label distribution.

Table 5 compares the classification accuracy of the deep learning network for different testing images from original and synthetic smartphone-based retinal images. We repeated the test with different sub-sampling rates of test images and their mean and standard deviation values are given in Table 5. Since the network is trained with original

images, it shows the highest overall accuracy (82% with a standard deviation of 0.0258) for testing with original images, as expected. The overall classification accuracies of smartphone-based systems are sorted as 61%, 62%, 69%, and 75% for iExaminer, D-Eye, Peek Retina, and iNview images, respectively.

Table 5. Classification accuracy of the deep learning network. Note that each result shows the mean and its standard deviation as mean  $\pm$  std.

	<b>Overall</b>	<b>No DR</b>	<b>DR</b>
<b>Original</b>	0.8164 $\pm$ 0.0258	0.7094 $\pm$ 0.0431	0.9234 $\pm$ 0.0372
<b>iExaminer</b>	0.6144 $\pm$ 0.0552	0.7661 $\pm$ 0.0582	0.4464 $\pm$ 0.0894
<b>D-Eye</b>	0.6211 $\pm$ 0.0593	0.7371 $\pm$ 0.0566	0.4827 $\pm$ 0.0784
<b>Peek Retina</b>	0.6889 $\pm$ 0.0353	0.8397 $\pm$ 0.0451	0.5140 $\pm$ 0.0539
<b>iNview</b>	0.7531 $\pm$ 0.0311	0.6094 $\pm$ 0.0491	0.8969 $\pm$ 0.0394

We observed that the overall network performance increases as the FoV of the smartphone-based retinal systems get bigger where their radii of printed fundus images are 32%, 40%, 45%, and 94% for iExaminer, D-Eye, Peek Retina, and iNview, respectively. We also observed that the accuracy of detecting healthy retina (No DR) is higher than retina with DR, especially for iExaminer, D-Eye, and Peek Retina images. In addition, the accuracy of both labels is close to each other for iNview images. The main reason is that iExaminer, D-Eye, and Peek Retina systems can capture smaller areas of the retina and mainly their images are focused on the optic disk and its surroundings. However, lesions generally appear in areas away from the optic disk. Since iNview captures a wider FoV, where it is more probable to include lesions, it detects DR better than other smartphone-based retinal imaging systems. For better visualization, we also show the overall, No DR,

and DR classification accuracies in Figure 19 for original, iExaminer, D-Eye, Peek Retina, and iNview images.

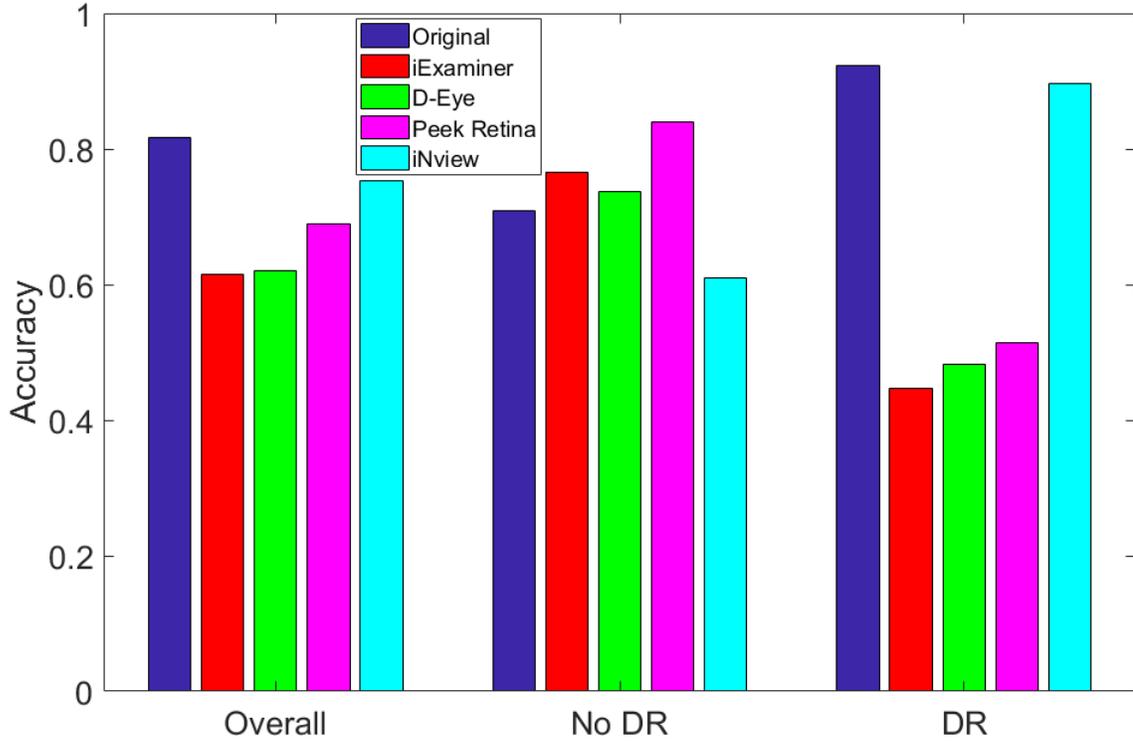


Figure 19. Classification accuracy of deep learning network for different testing images from original and synthetic smartphone-based retinal images.

Figure 20 compares the performance analysis of deep learning networks using Receiver Operating Characteristic (ROC) curve for different testing images from original and synthetic smartphone-based retinal images. To plot the ROC curve, we calculated the true positive rate (TPR) and the false positive rate (FPR) by changing the threshold value for the probability output of deep learning network. TPR is the probability of detecting No DR images as No DR, also known as sensitivity and recall. FPR is the probability of false alarm where No DR image is classified as DR and it is calculated as  $(1 - \text{specificity})$ . ROC curve analysis helps to select the optimal threshold value for a specific dataset discarding

the suboptimal solutions due to the class distributions. As threshold value changes, the TPR and FPR change accordingly. Based on the specific system requirement, we can select any threshold value to lower the false alarm rate or increase the detection accuracy.

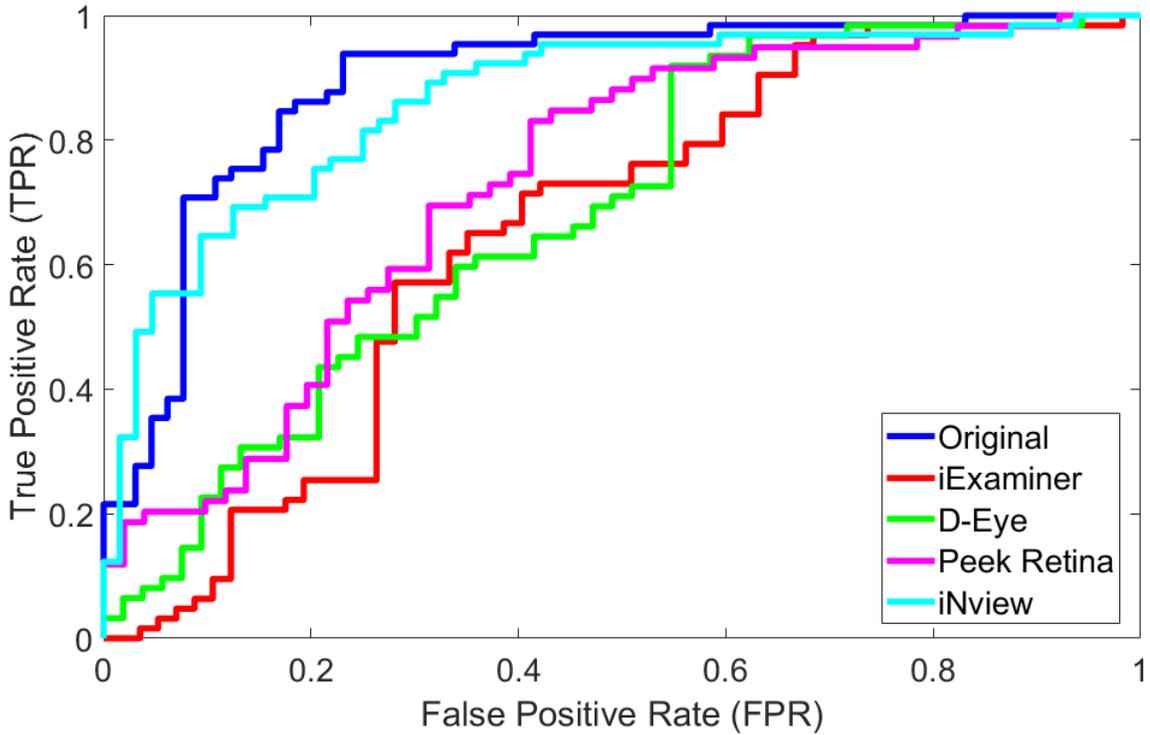


Figure 20. Performance analysis using ROC for different testing images from original and synthetic smartphone-based retinal images.

A specific threshold value in ROC generates an equal error value for false positive and false negative rate that is also known as Equal Error Rate (EER). For lower EER, the overall accuracy is higher. For original, iExaminer, D-Eye, Peek Retina, and iNview datasets respectively, EERs are calculated as 16.92%, 35.09%, 39.52%, 31.37%, and 23.44% for the following threshold values, 0.6283, 0.3915, 0.3007, 0.2625, and 0.7029. The area under the curve (AUC) values for the ROC curves are 0.8835, 0.6536, 0.6826, 0.7318, and 0.8649 for original, iExaminer, D-Eye, Peek Retina, and iNview datasets,

respectively, as shown in Table 6. As can be seen in ROC curves, original retina images marked with a solid blue line presents the best result compared with other smartphone-based retinal image datasets where all retinal structures are included in the input images, including the optic nerve, fovea, macula, and blood vessels. The network is trained with retina images that have a large FoV, where almost 100 degrees of the retina is visible. However, smartphone-based retinal systems can visualize a smaller area from the retina. iNview shows the best among others and close the results with original images because iNview covers almost the same amount of area (94%) compared to the fundus camera. When the smaller area is captured from the retina, the network performance is lower. We observed that the performance of the network depends on the FoV of the retinal imaging system.

Table 6. Area Under Curve (AUC), Equal Error Rate (EER), and Selected Threshold for EER parameters at ROC curve for different testing images from original and synthetic smartphone-based retinal images.

	Area Under Curve	Equal Error Rate	Threshold for ERR
<b>Original</b>	0.8835	0.1692	0.6283
<b>iExaminer</b>	0.6536	0.3509	0.3915
<b>D-Eye</b>	0.6826	0.3962	0.3007
<b>Peek Retina</b>	0.7318	0.3137	0.2625
<b>iNview</b>	0.8649	0.2344	0.7029

## 5.2. Diabetic Retinopathy Detection using Multiple Datasets

In our experiments, we first presented the performance results of deep learning algorithms with original retina images captured by traditional fundus cameras in seven different experiments in order to show the effect of using retina images from the single,

cross, and merged datasets in training and validation. Then, we compare these deep learning-based results with the smartphone-based synthetic retina images to investigate the effect of the FoV of smartphone-based retinal imaging devices. To show the network effectiveness with different training sets, we also compare their performance of CNNs by training and testing the networks with different types of images from different datasets. Therefore, we address the data fusion capability of CNNs with retina images from different datasets to improve the recognition performance and investigate the network behavior for untrained retina images from different datasets. In the following subsections, we first trained and tested the networks with original images. Then, we tested our network with smartphone-based synthetic retina images.

### **5.2.1. Results for Original Images**

In our first set of experiments using multiple datasets, we tested seven combinations of datasets in training and testing and their deep learning results are shown in Table 7. First, we performed our first and second sets of experiments using single datasets (1 and 2). When the network is trained and tested with images from the same datasets, the overall DR detection accuracies of the network are 0.7661 and 0.8267 for EyePACS 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 other datasets. The accuracy of the network for EyePACS training and Messidor testing dropped to 0.5734. However, we observed better accuracy results as 0.8750 for training with Messidor and validating with EyePACS. The main reason is that EyePACS has lower quality images due to the reflections and low contrast compared with images in Messidor datasets. In addition, there exist several inconsistencies in labeling in

EyePACS images. Therefore, training with only EyePACS images and testing with cross datasets results in lower accuracy. Then, we merged these two datasets for training and validation and performed experiments (5-7). We observed that the detection accuracy for rDR and vtDR are 0.9126 and 0.9460, respectively as shown in Table 7.

Table 7. Classification accuracy of deep learning frameworks

<b>Datasets</b>	<b>Type</b>	<b>Overall</b>	<b>No DR</b>	<b>DR</b>
<b>(1) EyePACS</b>	Single	0.7661	0.8065	0.7258
<b>(2) Messidor</b>	Single	0.8267	0.8933	0.7600
<b>(3) Pac_Mes</b>	Cross	0.5734	0.9908	0.1560
<b>(4) Mes_Pac</b>	Cross	0.8750	0.8796	0.8704
<b>(5) Pac-Mes</b>	Merged	0.8321	0.8467	0.8175
<b>(6) rDR</b>	Merged	0.9126	0.9235	0.9016
<b>(7) vtDR</b>	Merged	0.9460	0.9460	0.9460

Our network classifies images into two different classes based on the highest probability of using the output of the last layer. Since there are only two classes for classification, the image is classified as a healthy retina if the probability of No DR is higher than 0.5. However, equal probability might not provide the best performance. Therefore, we used ROC curves to make performance analysis in our experiments where it plots accuracy based on the various thresholds. Based on the results from the ROC curves, we selected two operating points. The first operating point is set to show the best sensitivity and the second one demonstrates to best specificity. Note that sensitivity is the most important factor for medical research since the result shows the success rates of the detection of unhealthy retinas.

Tables 8 and 9 show the accuracy of rDR and vtDR detection using two operating points for high sensitivity and high specificity. Note that, we used EyePACS (Label 0-3-4), Messidor (Label 0), Messidor-2 (Label 1), IDRiD (Label 0-2-3-4) images for rDR; and EyePACS (Label 0-3-4), Messidor (Label 0-4), and IDRiD (Label 0-3-4) images for vtDR. Using the first operating point (high sensitivity) for rDR, the DR detection accuracy is 0.9563 and the No DR detection accuracy is 0.8415. For the second operating point (high specificity), the DR detection accuracy is 0.8743 and the No DR detection accuracy is 0.94. For vtDR detection, using the first operating point (high sensitivity), the DR detection accuracy is 0.9730 and No DR detection accuracy is 0.7973. For the second operating point (high specificity), the DR detection accuracy is 0.9460 and the No DR detection accuracy is 0.8648.

Table 8. Accuracy for rDR detection using two operating points

	<b>Overall</b>	<b>No rDR</b>	<b>rDR</b>
<b>High Sensitivity</b>	0.8989	0.8415	0.9563
<b>High Specificity</b>	0.9071	0.94	0.8743

Table 9. Accuracy for vtDR detection using two operating points

	<b>Overall</b>	<b>No vtDR</b>	<b>vtDR</b>
<b>High Sensitivity</b>	0.8851	0.7973	0.9730
<b>High Specificity</b>	0.9054	0.9460	0.8648

### 5.2.2. Results for Smartphone-based Images

In our second set of experiments using multiple datasets, we investigate more in the effect of FoV of smartphone-based synthetic retinal imaging devices. Based on the previous results, we received the high detection accuracies at the merged data sets with two operating points for rDR and vtDR detection. Therefore, we trained our deep learning network with retina images from EyePACS, Messidor-2, and IDRiD datasets the same as the seventh experiment in the previous subsection. In addition, to address the cross datasets issues in deep learning, we tested our trained network with smartphone-based synthetic images generated from a completely new dataset, UoA-DR with different FoV ranging from 20% to 90% with a 10% step-size. In addition, to include synthetic images from PanOptic, D-Eye, Peek Retina, and iNview smartphone-based retinal imaging systems, we also tested synthetic images from 32%, 40%, 45%, and 94% FoV based on the calculations in this paper [45]. Using images from the UoA-DR dataset allows us to test the cross datasets without overlaps between training and testing images where they captured in totally different setups.

Table 10 presents the results of vtDR detection performance of GoogLeNet for original and synthetic images with different FoV. Since the network is trained with the same size images as original images, it shows the highest overall accuracy (94%) for testing with original images, as expected. The overall vtDR detection accuracy sorted as 89%, 88%, 86%, 82%, 77%, 74%, 70%, 69%, 61%, 60%, and 48% for synthetic retina images with 94%, 90%, 80%, 70%, 60%, 50%, 45%, 40%, 32%, 30%, and 20% FoV compared with the original images. We observed that the overall network performance decreases as the FoV of the smartphone-based images get smaller.

Table 10. Classification accuracy of deep learning frameworks

<b>Datasets</b>	<b>Overall</b>	<b>No DR</b>	<b>DR</b>
<b>Original</b>	0.9433	0.9464	0.9412
<b>94% (IN)</b>	0.8865	0.8036	0.9412
<b>90%</b>	0.8794	0.75	0.9647
<b>80%</b>	0.8652	0.7857	0.9176
<b>70%</b>	0.8227	0.7679	0.8588
<b>60%</b>	0.7714	0.6429	0.8571
<b>50%</b>	0.7376	0.5714	0.8471
<b>45% (PR)</b>	0.7021	0.5179	0.8235
<b>40% (DE)</b>	0.6934	0.5455	0.7927
<b>32% (PO)</b>	0.6071	0.3750	0.7619
<b>30%</b>	0.6058	0.3455	0.7805
<b>20%</b>	0.4752	0.4107	0.5176

For better visualization, we also showed the overall accuracy in Figure 21 for images with different FoV and marked the accuracy of PanOptic, D-Eye, Peek Retina, and iNview systems.

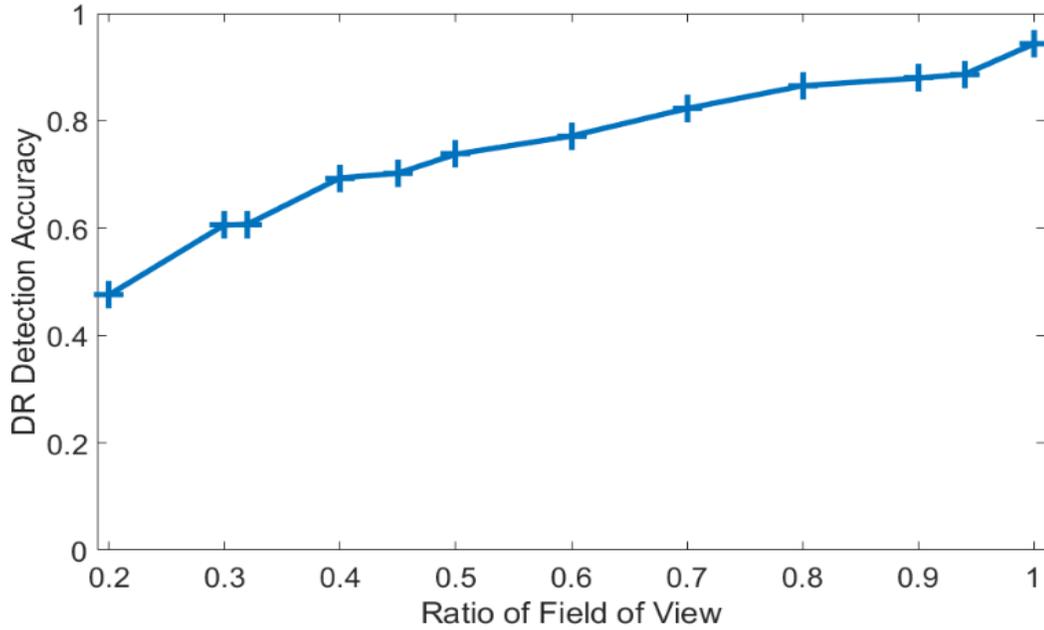


Figure 21. Overall classification accuracy of deep learning networks for different smartphone-based synthetic retina images with respect to their FoV percentage

We presented the performance analysis of deep learning networks for original and synthetic images using ROC curves in Figure 22. ROC curve is 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 DR) images as healthy. FPR is the probability of false alarm where the healthy (No DR) retina images are categorized as a disease (DR). Since TPR and FPR values change according to selected threshold changes, we can select any threshold value to lower the false alarm rate or increase the detection accuracy based on the specific system requirement.

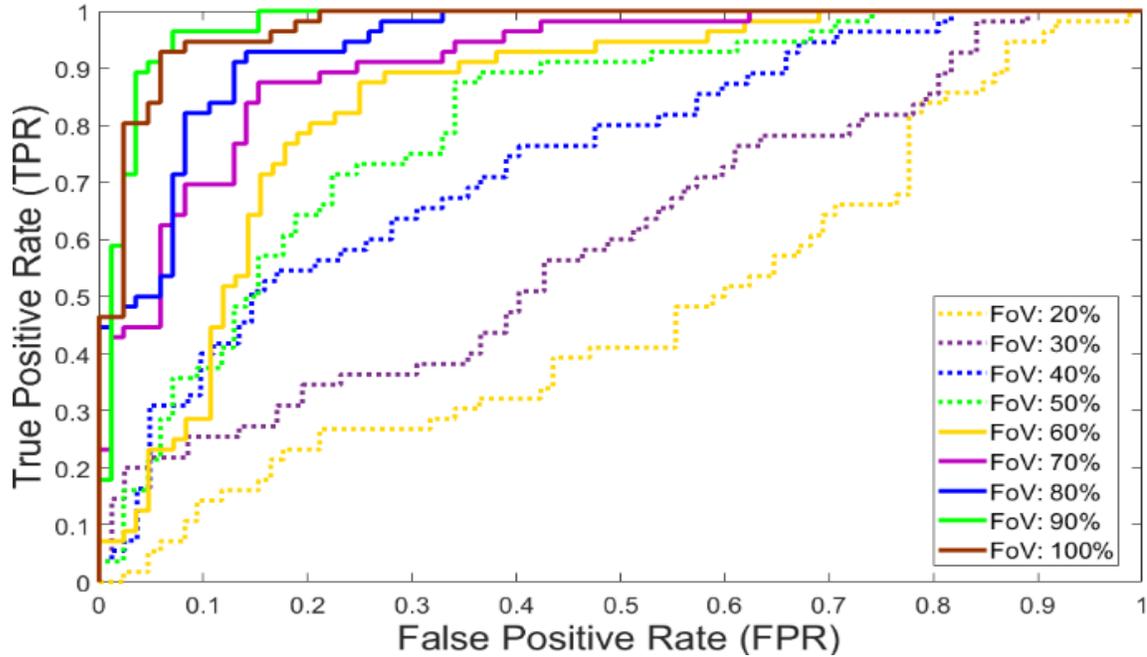


Figure 22. Performance analysis using ROC for different testing images from original and synthetic smartphone-based retinal images with various percentages of FoV compared with original retina images.

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 ROC. For lower EER, the overall accuracy is higher. For original and synthetic images, calculated EERs range from 5.6% to 55.03% for the threshold values that vary from 0.007 to 0.21. The area under the curve (AUC) values for the ROC curves changes from 0.978 to 0.475 for original and smartphone-based datasets, respectively as shown in Table 11.

Table 11. Area Under Curve (AUC), Equal Error Rate (EER), and Selected Threshold for EER results of deep learning frameworks

<b>Datasets</b>	<b>AUC</b>	<b>EER</b>	<b>THR</b>
<b>Original</b>	0.978	0.056	0.007
<b>94% (IN)</b>	0.97	0.086	0.041
<b>90%</b>	0.976	0.071	0.057
<b>80%</b>	0.942	0.127	0.057
<b>70%</b>	0.911	0.157	0.078
<b>60%</b>	0.841	0.199	0.036
<b>50%</b>	0.804	0.269	0.073
<b>45% (PR)</b>	0.768	0.305	0.055
<b>40% (DE)</b>	0.742	0.328	0.045
<b>32% (PO)</b>	0.648	0.414	0.011
<b>30%</b>	0.599	0.438	0.024
<b>20%</b>	0.475	0.553	0.210

As shown in ROC curves in Figure 22, original retina images marked with solid green line shows the best result compared with smartphone-based synthetic images. Since the network is trained with images that include all retinal structures such as the optic nerve, fovea, macula, and blood vessels, original images cover a similar area with training images. However, smartphone-based systems have a narrower FoV that cover smaller retinal areas. We observed that the network accuracy depends on the FoV and it decreases as the FoV becomes smaller as presented in ROC curves.

## 6. CONCLUSIONS

This thesis first investigated the smartphone-based portable retinal imaging systems, namely iExaminer, D-Eye, Peek Retina, and iNview to compare their image quality. Then, we adapted deep learning frameworks using transfer learning. Using smartphones is an emerging research area in designing small-sized, low-power, and affordable retinal imaging systems to perform DR screening and automated DR detection due to the size, weight, and price of fundus cameras. Smartphone-based portable retinal imaging systems available on the market are capable of capturing retina images without analyzing them using any machine learning and image processing techniques to detect DR development.

Based on the results, iNview retinal imaging system has the largest FoV and better image quality compared with iExaminer, D-Eye, and Peek Retina systems. The overall classification accuracies of smartphone-based systems using a single dataset with AlexNet are sorted as 61%, 62%, 69%, and 75% for iExaminer, D-Eye, Peek Retina, and iNview images, respectively. We also observed that the network DR detection performance decreases as the FoV of the smartphone-based retinal systems get smaller. Finally, we presented the utility of deep learning methods to improve the performance of DR detection in smartphone-based and traditional fundus camera images. This approach allows us to compare the FoV in smartphone-based retinal imaging systems and improve DR detection accuracy by training the network with merged publicly available datasets. We generated smartphone-based synthetic retina images by simulating the different FoV with masking the original image around the optic disk and cropping it. Using transfer learning of GoogLeNet, we investigate the automatic DR detection accuracy for traditional fundus

camera and smartphone-based retinal imaging systems for retina images from several datasets including EyePACS, Messidor, IDRiD, Messidor-2, and UoA-DR. Although we have used a smaller number of images in the training set compared with the existing studies, we obtained considerably acceptable higher accuracies for smartphone-based synthetic images that are comparable with the original retina images. Specifically, the overall classification accuracies of smartphone-based systems using multiple datasets with GoogLeNet are sorted as 61%, 70%, 71%, and 89% for iExaminer, D-Eye, Peek Retina, and iNview images, respectively. Furthermore, we observed that the best results for vtDR detection are 95% for accuracy and 95% for sensitivity with the original images. For smartphone-based synthetic images, vtDR accuracy is 89% and sensitivity is 94% using iNview synthetic data.

As a result, the smartphone-based retina imaging systems can be used as an alternative to the direct ophthalmoscope once it is validated in the clinical settings. However, as we have discussed the FoV of the smartphone-based retina imaging systems plays an important role in determining the automatic DR detection accuracy.

In the future, if any smartphone-based retinal imaging system device can be designed to capture the retina images without color artifacts, blurring and contrast problems, tens of thousands of unhealthy data can be collected in clinical settings from the real patients, these systems can have the potential to change DR detection procedures and save millions of people's lives in terms of vision loss and blindness.

## REFERENCES

- [1] Aiello, L.P., Beck, R.W., Bressler, N.M., Browning, D.J., Chalam, K.V., Davis, M., Ferris III, F.L., Glassman, A.R., Maturi, R.K., Stockdale, C.R. and Topping, T.M., 2011. Rationale for the diabetic retinopathy clinical research network treatment protocol for center-involved diabetic macular edema. *Ophthalmology*, 118(12), pp.e5-e14.
- [2] Herman, W.H., 2015. The cost-effectiveness of diabetes prevention: results from the Diabetes Prevention Program and the Diabetes Prevention Program Outcomes Study. *Clinical Diabetes and Endocrinology*, 1(1), p.9.
- [3] Hartnett, M.E., Key, I.J., Loyacano, N.M., Horswell, R.L. and DeSalvo, K.B., 2005. Perceived barriers to diabetic eye care: qualitative study of patients and physicians. *Archives of ophthalmology*, 123(3), pp.387-391.
- [4] Faust, O., Acharya, R., Ng, E.Y.K., Ng, K.H. and Suri, J.S., 2012. Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review. *Journal of medical systems*, 36(1), pp.145-157.
- [5] Haddock, L.J., Kim, D.Y. and Mukai, S., 2013. Simple, inexpensive technique for high-quality smartphone fundus photography in human and animal eyes. *Journal of ophthalmology*, 2013.
- [6] PanOptic + iExaminer. <http://www.welchallyn.com/en/microsites/iexaminer.html> Accessed December 1, 2019.
- [7] Russo, A., Morescalchi, F., Costagliola, C., Delcassi, L. and Semeraro, F., 2015. A novel device to exploit the smartphone camera for fundus photography. *Journal of ophthalmology*, 2015.
- [8] Maamari, R.N., Keenan, J.D., Fletcher, D.A. and Margolis, T.P., 2014. A mobile phone-based retinal camera for portable wide field imaging. *British Journal of Ophthalmology*, 98(4), pp.438-441.
- [9] Volk iNview. <https://volk.com/index.php/volk-products/ophthalmic-cameras/volk-inview.html> Accessed December 1, 2019.
- [10] Sim, D.A., Keane, P.A., Tufail, A., Egan, C.A., Aiello, L.P. and Silva, P.S., 2015. Automated retinal image analysis for diabetic retinopathy in telemedicine. *Current diabetes reports*, 15(3), p.14.
- [11] Philip, S., Fleming, A.D., Goatman, K.A., Fonseca, S., Mcnamee, P., Scotland, G.S., Prescott, G.J., Sharp, P.F. and Olson, J.A., 2007. The efficacy of automated “disease/no disease” grading for diabetic retinopathy in a systematic screening programme. *British Journal of Ophthalmology*, 91(11), pp.1512-1517.
- [12] Karnowski, T.P., Giancardo, L., Li, Y., Tobin, K.W. and Chaum, E., 2013, July. Retina image analysis and ocular telehealth: the Oak Ridge National Laboratory-Hamilton Eye Institute case study. In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 7140-7143). IEEE.

- [13] Garg, S., Jani, P.D., Kshirsagar, A.V., King, B. and Chaum, E., 2012. Telemedicine and retinal imaging for improving diabetic retinopathy evaluation. *Archives of internal medicine*, 172(21), pp.1677-1680.
- [14] IDx-DR. <https://www.eyediagnosis.net/> Accessed December 1, 2019.
- [15] Retmarker. <http://www.retmarker.com/> Accessed December 1, 2019.
- [16] <http://retinalyze.net/> Accessed December 1, 2019.
- [17] Graham B. Kaggle Diabetic Retinopathy Detection competition report. <https://www.kaggle.com/c/diabetic-retinopathy-detection/discussion/15801>, 2015. Accessed December 1, 2019.
- [18] Rajalakshmi, R., Arulmalar, S., Usha, M., Prathiba, V., Kareemuddin, K.S., Anjana, R.M. and Mohan, V., 2015. Validation of smartphone based retinal photography for diabetic retinopathy screening. *PloS one*, 10(9), p.e0138285.
- [19] Rajalakshmi, R., Subashini, R., Anjana, R.M. and Mohan, V., 2018. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*, 32(6), p.1138.
- [20] Solanki, K., Ramachandra, C., Bhat, S., Bhaskaranand, M., Nittala, M.G. and Sadda, S.R., 2015. EyeArt: automated, high-throughput, image analysis for diabetic retinopathy screening. *Investigative Ophthalmology & Visual Science*, 56(7), pp.1429-1429.
- [21] Solanki K, Bhaskaranand M, Ramachandra C, Bhat S. Comprehensive clinical validation study of a fully-automated diabetic retinopathy screening system using color fundus images against 7-field ETDRS stereoscopic reference standard. *EURETINA 2016 Abstract*
- [22] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J. and Kim, R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22), pp.2402-2410.
- [23] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).
- [24] Abramoff, M.D., Folk, J.C., Han, D.P., Walker, J.D., Williams, D.F., Russell, S.R., Massin, P., Cochener, B., Gain, P., Tang, L. and Lamard, M., 2013. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA ophthalmology*, 131(3), pp.351-357.
- [25] Abramoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C. and Niemeijer, M., 2016. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative ophthalmology & visual science*, 57(13), pp.5200-5206.
- [26] Gargeya, R. and Leng, T., 2017. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 124(7), pp.962-969.

- [27] Philip, S., Fleming, A.D., Goatman, K.A., Fonseca, S., Mcnamee, P., Scotland, G.S., Prescott, G.J., Sharp, P.F. and Olson, J.A., 2007. The efficacy of automated “disease/no disease” grading for diabetic retinopathy in a systematic screening programme. *British Journal of Ophthalmology*, 91(11), pp.1512-1517.
- [28] Carson Lam, D.Y., Guo, M. and Lindsey, T., 2018. Automated detection of diabetic retinopathy using deep learning. *AMIA Summits on Translational Science Proceedings*, 2018, p.147.
- [29] Pires, R., Avila, S., Wainer, J., Valle, E., Abramoff, M.D. and Rocha, A., 2019. A data-driven approach to referable diabetic retinopathy detection. *Artificial intelligence in medicine*, 96, pp.93-106.
- [30] D-EYE. <http://www.d-eyecare.com/> Accessed December 1, 2019.
- [31] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [32] Vedaldi, A. and Lenc, K., 2015, October. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 689-692). ACM.
- [33] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- [34] ImageNet. <http://www.image-net.org/> Accessed December 1, 2019
- [35] Vedaldi, A. and Lenc, K., 2015, October. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 689-692). ACM.
- [36] Kaggle Diabetic Retinopathy Detection Competition. <https://www.kaggle.com/c/diabetic-retinopathy-detection/data/> Accessed December 1, 2019
- [37] Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A. and Charton, B., 2014. Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology*, 33(3), pp.231-234.
- [38] Quéllec, G., Lamard, M., Josselin, P.M., Cazuguel, G., Cochener, B. and Roux, C., 2008. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE transactions on medical imaging*, 27(9), pp.1230-1241.
- [39] Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabudde, V. and Meriaudeau, F., 2018. Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. *Data*, 3(3), p.25.

- [40] Chalakkal, R.J., Abdulla, W.H. and Sinumol, S., 2017, November. Comparative analysis of university of auckland diabetic retinopathy database. In Proceedings of the 9th International Conference on Signal Processing Systems (pp. 235-239). ACM.
- [41] Chalakkal, R.J. and Abdulla, W., 2017, March. Automatic segmentation of retinal vasculature. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 886-890). IEEE.
- [42] Messidor Dataset. <http://www.adcis.net/en/third-party/messidor/> Accessed December 1, 2019
- [43] International Clinical Diabetic Retinopathy Disease Severity Scale. <http://www.icoph.org/downloads/Diabetic-Retinopathy-Scale.pdf> Accessed December 1, 2019
- [44] Messidor-2 reference standard for referable Diabetic Retinopathy. <https://medicine.uiowa.edu/eye/abramoff/> Accessed December 1, 2019
- [45] Hacisoftoglu, R.E. and Karakaya, M., 2019, May. Field of view of portable ophthalmoscopes for smartphones. In Smart Biomedical and Physiological Sensor Technology XV (Vol. 11020, p. 110200X). International Society for Optics and Photonics.