Detection of Referable Diabetic Retinopathy From Fundus Image Using Convolutional Neural Network

Sandep Gupta1, Ashim Kahdka2, Bidur Devkota3

[1] *Department of Software Engineering, Gandaki College of Engineering and Science, Pokhara University, Nepal*
[2]*Department of Electronics and Communication Engineering, NCIT, Kathmandu, Nepal*
[3]*Department of Electronics and Computer Engineering, IOE, Pashchimanchal Campus, Nepal*

(Manuscript Received 000 0, 2022; Revised 000 0, 2022; Accepted 000 0, 2022)

Abstract

Diabetic retinopathy (DR) is an eye disease which is caused by high blood sugar and high blood pressure and damage the blood vessels in the back (retina) of the eye. DR is not a reversible process and treatment only sustain vision. The number of ophthalmologists cannot meet the growing demands around the world. This study focuses on the automatic diagnosis of the disease through deep learning. Convolutional neural network (CNN) are more widely used as a deep learning method in medical image analysis. Binary and multiclass image classification of images into non referable DR and referable DR has been done using proposed CNN model. The Binary classification prediction result gives sensitivity, specificity and F2 score of 90 %, 93.56 % and 90.65 % respectively using 10 fold cross validation. For multiclass classification problem, the sensitivity, specificity and F2 score value for class 0 (no DR) are 81.75 %, 91.06 % and 81.80 % respectively. Whereas for class 1 (non severe DR),sensitivity, specificity and F2 score values are 71.28 %, 78.52 % and 70.01 % respectively. Similarly for class 2 (sever DR), sensitivity, specificity and F2 score values are 73.03 %, 93.39 % and 75.08 % respectively. The images in dataset has poor contrast and consists of impulse noises. Contrast limited adaptive histogram equalization (CLAHE) method is used to improve the contrast of the image followed by median filter to remove noise present image.

Keywords: DR; CNN; Deep Learning, CLAHE, Median filter;

---

1. Introduction

Diabetes is a disease that increases the amount of glucose in the blood caused by a lack of insulin[1]. Diabetes influences the eyes and vision in various ways, including visual impairment, cataract glaucoma, impact on optic nerve, temporary paralysis of the muscles on the outside of the eye, and double vision. But the most common and most important of these artifacts is the impact on the retina[1]. Diabetic Retinopathy (DR) is a complication of diabetes that causes the blood vessels of the retina to swell and to leak fluids and blood [2,3]. DR is one of the main causes of reduced vision and the probability of DR increases with longer duration of being affected with diabetes. People with untreated diabetes are more likely to lose their eyesight than normal people[1].

The early screening of DR remains a challenge for several reasons. First diabetes are generally treated at endocrinology department in hospitals and fundus examination has long been ignored resulting in delayed treatment for many patients. Second, the process of DR screening is time consuming. Consequently, only a certain number of patients can be processed each day. Third, the number of ophthalmologists cannot meet the growing demands around the world[5].

Automated DR screening through deep learning is an alternative solution to the above problems, with the

Corresponding author. Sandep Gupta
E-mail address:kasaudhansandeep48@gmail.com ,

used for medical image analysis process and can help in clinical diagnosis. Unlike conventional machine learning methods that rely heavily on feature engineering, deep learning algorithms automatically learn the most predictive representations in a manner of layer by layer combination. This proposal deals with the referable DR detection technique through binary and multiclass image classification namely referable DR and non-referable DR. This technique allows to triage patients who require referral to the ophthalmologist from those who can wait until the next screening.

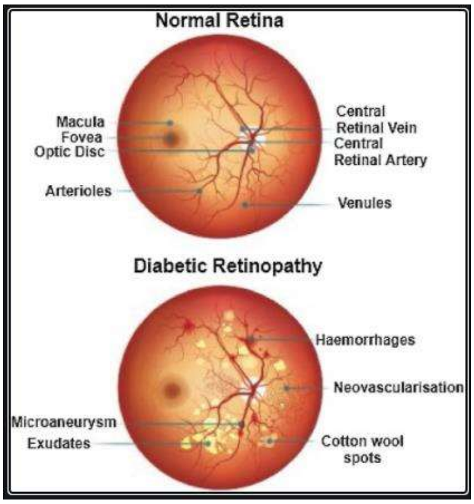

Figure1: Difference between normal retina and diabetic retinopathy

DR is detected by the appearance of different types of lesions on a retina image. These lesions are microaneurysm, hemorrhages, soft exudates and hard exudates. Red lesions are microaneurysm and hemorrhages, while bright lesions are soft and hard exudates[5]. Figure1 shows the fundus images for normal retina and DR including all those lesions[6]. There are five stages of DR depending on the presence of these lesions, namely, no DR, mild DR, moderate DR, severe non proliferative DR and proliferative DR. No DR, mild DR, moderate DR are categorized in non referral DR while severe DR and proliferative DR are referral DR.

Many datasets such as Messidor, IDRiD etc consists of high quality image which were captured under controlled, non standard conditions (i.e. similar environmental and hardware conditions across captures). Thus it can be argued that the algorithms trained on such datasets will perform poorly under typical practical solutions where the image may not be directly compatible and the environmental and hardware details may differ. Also the dataset Kaggle Eyepacs and APTOS consists of images which were captured from a variety of camera models, under

the analysis. Also some datasets including EYEPACS and APTOS 2019 have been graded by only one experts, which can lead to an annotation bias[9]. The objectives of the study is to classify the fundus images into referable and non-referable image using binary as well as multi-class

1.1 Objective and Contributions

The objectives of this research is

- to classify the fundus image into referable and non-referable DR images} using binary as well as multi-class classification

The contribution of the thesis are given below.

- Classification of DR images has been done into non-referable and referable DR images. Images such as no DR, mild DR and moderate DR belong to non-referable DR class whereas severe non-proliferative DR and proliferative DR belong to referable DR class.
- Multiclass image classification was also done in order to provide more variation in the referral assessment of DR images. Such as images that belong to class 2 (severe non-proliferative DR and proliferative DR) should be referred earlier than images belonging to class 1 (mild DR and moderate DR). Whereas images that belong to class 0 are non referable as they absence any DR lesions.
- Total image dataset was prepared using five different datasets such as IDRiD, Messidor, EYEPACS, kaggle APTOS and DDR reflecting the diversity in the image dataset.
- Datset were highly preprocessed through a series of multiple tasks to obtain good quality image. The images were cropped in order to remove unnecessary background. Then CLAHE operation has been applied followed by the median filtering technique in order to improve the poor contrast of image and remove the noise present in image.
- The customized CNN algorithm was used as deep learning methodology.

2. Materials and Method

The block diagram for system model for the classification of fundus imges is shown in Figure 2

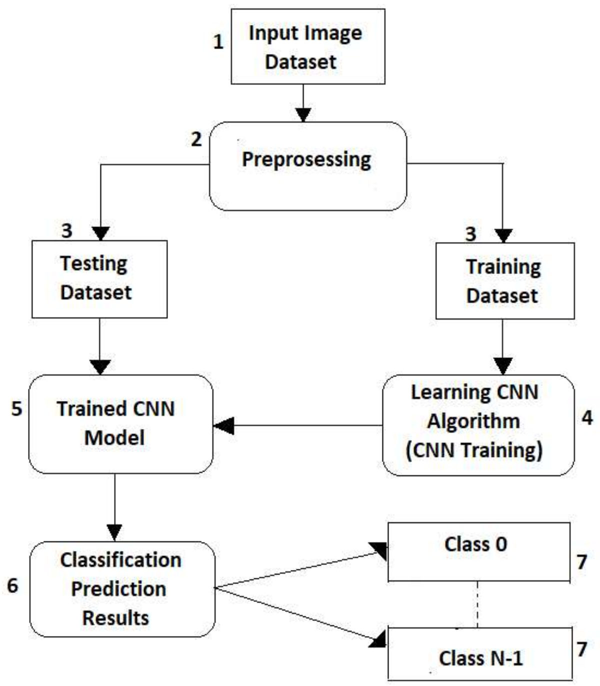

Figure2: System Model

2.1 Data Collection

Fundus image dataset has been used for the proposed method such as IDRiD, MESSIDOR, Kaggle Aptos 2019, DDR, Kaggle EYEPACS[1,8,9]. Table1 shows the number of fundus images taken for the binary classification. Total number of fundus images taken is 19020. among them non-referable DR image class includes 13334 images while referable DR image class includes only 5686 images. Hence the dataset is highly imbalanced. For multiclass classification, class 0 (no DR) includes 4676 images, class 1 (mild DR and moderate DR) includes 8658 images and class 2 (severe non-proliferative DR and proliferative DR) includes 5,686 images. The total number of image dataset used is given in table 1.

Table1: Total number of images taken

Non referable dr image	Referable dr image	Total
13334	5686	19020

2.2 Preprocessing

Retinal images are normally affected with impulse noise during image acquisition. Noise detection and removal is an important process as the image are corrupted by impulse noise because of transmission and acquisition. Also Non -uniform illumination and poor contrast due to the anatomy of fundus image, opaque media, and wide angle optics of the camera, insufficient pupil size and sensing array geometry, and the are the major causes of the low quality retinal images. The signs of DR are micro-aneurysm, hemorrhages, edema, hard exudates and cotton wool spots. The exterior of those DR lesions varies. For example microaneurysm and hemorrhages are red dark spots and are mostly undividable from the background while exudates are a high contrast yellow color. The accuracy of all these symptoms(lesions) depends on the quality of acquired retinal image. Before the detection of abnormalities and feature selection in retinal images, it is must to remove the different noise present in the retina, which will automatically increase the quality of the image. Therefore the use of good quality retinal images is very essential for accurate detection, diagnosis and damage assessment of retina. The aim of prepossessing is to increase the quality of an image by reducing the amount of noise appearing in the image and highlighting features that are used in image classification. Figure 3 shows the series of tasks which are carried out for prepossessing the fundus image dataset to improve the quality of the image[7,10].

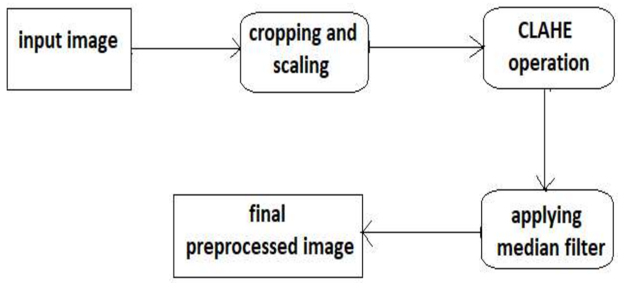

Figure3: Prepossessing steps

2.3 CNN algorithm

The architecture being utilized is shown in figure 4. This methodology is utilizing a 21 layered approach with each layer having its own specification. Algorithm consists of feature extracting layers and classification layers. Five sets of convolution layer has been used as feature extracting layer. In each set two convolution layers having same number of filter and same kernel size has been used. The number of filter for convolution layers goes on increasing as we move deeper into the network. Given the input image, first convolution layer convolves 32 different kernel of size 3×3 to extract low level features from the input image like edges. Also the image is zero padded along each size of image by 1. Convolution layer uses the

ReLU activation function to produce non linear output feature map. Convolution layer is followed by the batch normalization. The use of batch normalization is important in deep learning. Training deep neural network with tens of layers is challenging as they can be sensitive to the initial random weights and configuration of the learning algorithm. This problem is known as internal covariat shift. Batch normalization solves this major problem of internal covariat shift. It is a technique for training very deep neural networks that normalizes the contribution to a layer for every mini-batch. This has the impact of settling the learning proces and drastically decreasing the number of training epochs required to train deep neural networks.

Second convolution layers also have the same kernel size of 32 but is followed by the batch normalization and max pooling layer of kernel size 3×3 with strides 2. This combination of convolution + batch normalization + convolution + batch normalization + max pooling is repeated for the entire feature extracting layers. One dimensional flattening of the CNN is accomplished by the Global average pooling layer followed by the 5 fully connected layers (1024, 256, 64, 32 and 8) which helps to identify very small lesion like micro-aneurysm, hemorrhages. The classification output is predicted using softmax activation function and dropout is performed after each dense layer to reduce the chances of over-fitting.

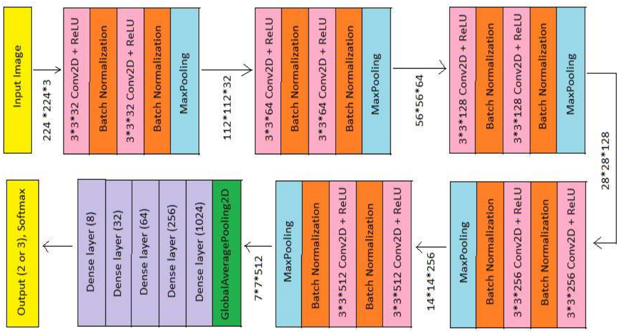

Figure4: Architecture of proposed CNN model

3. Results and Discussion

Deep learning algorithms such as proposed CNN model and pre-trained CNN model called densenet-121 have been implemented during the experimentation. To measure the model performance, evaluation metrics like sensitivity, specificity, precision factor and F2 score value are computed from the confusion metric which includes TP, TN, FP and FN. Such type of DR classification problem for medical diagnosis should give more concern on FN rather than FP. For this concern F2 score was chosen in order to give more weights on recall than precision. Also ROC curve has been used to analyze the model performance.

3.1 Binary Classification

Binary classification of DR images into non-referable DR (no DR, mild DR and moderate DR) and referable DR (severe non-proliferative DR and proliferative DR) has been done using proposed CNN model.

Table2 shows the results of proposed CNN model for binary classification. The model was trained taking different dropout values from 0 to 0.5 with increment 0.1. Performance measures such as sensitivity, specificity and F2 score was taken to compare the performance of the experiments. It can be observed that the optimal result is obtained when 0.5 drop out is used in the proposed CNN model.

Table2: Performance of model for different split ratio

Drop Out Value	Sensitivity	Specificity	F2 Score
No	0.8	0.97	0.82
0.1	0.8	0.96	0.82
0.2	0.8	0.96	0.83
0.3	0.82	0.95	0.84
0.4	0.84	0.94	0.86
0.5	**0.89**	**0.92**	**0.89**

10 Fold Cross Validation

10 fold cross validation has been performed using different training- testing split ratio. Table2 shows optimal learning rate and corresponding test accuracy for different split ratio. Optimal test accuracy is obtained for 70:30 split ratio when learning rate is 0.0008. Table 3 shows the score table for binary classification using 10 fold cross validation. Sensitivity, specificity and F2 score value are 90 %, 93.56 % and 90.65 % respectively. Figure 5 shows the confusion metrcis and ROC curve whereas figure6 shows the accuracy and loss curve for binary classification using 10 fold cross validation. The ROC curve shows a plot between True Positive Rate (TPR) and False Positive Rate (FPR). The threshold line is a 45° angle line shown by dashed line. The binary classification results of proposed CNN is above the threshold value of ROC curve. Hence it can be concluded that the test accuracy is good as curve for each classification are near to 1 value of ROC.

Table3: Score table for binary classification

Senisitivity	Sepecificity	F2 Score
0.90	0.94	0.91

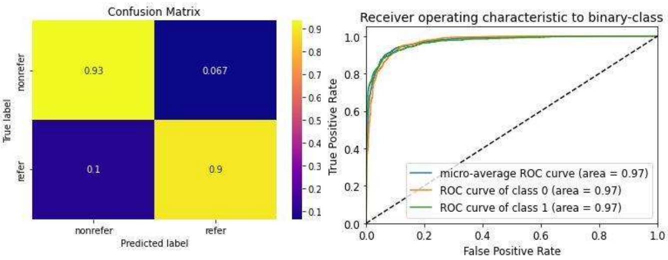

Figure5: Confusion metrics and ROC curve for binary classification (learning rate: 0.0008 and batch size:32)

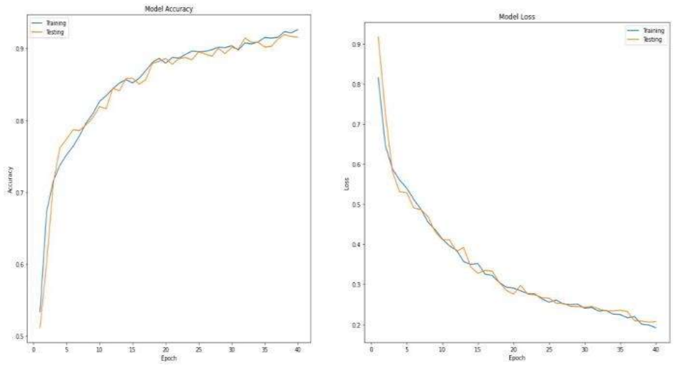

Figure6: Accuracy and loss curve for binary classification (learning rate:0.0008 and batch size:32)

3.2 Multiclass Classification

10 fold cross validation has been used for hyperparameter tuning purpose. Learning rate and train-test split ratio of dataset are the hyper-parameters which are tuned through 10 fold cross validation for the proposed model. 10 fold cross validation has been used for hyperparameter tuning purpose. Learning rate and train-test split ratio of dataset are the hyper-parameters which are tuned through 10 fold cross validation for the proposed model. Tab \ref{compsplit} shows the results of 10 fold cross validation for different training-testing split ratio using different learning rate from 0.0001 to 0.001 with increment of 0.0001. Optimal result is obtained for 70:30 training-testing split ratio when learning rate is 0.0004.

Table4: Performance of model for different spli ratio

Training-testing ratio	Optimal learning rate	Testing Accuracy
80:20	0.0008	92.58%
70:30	**0.0008**	**92.58%**
60:40	0.0006	87.34%
50:50	0.0006	86.63%

Hence the proposed CNN model is trained with optimal hyperparameter set that is learning rate of 0.0004 and 70:30 train-test split ratio. Table5 shows the performance result of the model for multiclass image classification problems. Sensitivity, specificity and F2 score value for level 0 are 81.75 %, \$91.06 % and 81.8 % respecitively. For level 1, sensitivity, specificity and F2 score value are 71.28 %, 78.52 % and 70.01 % respectively. Similarly for level 3, sensitivity, specificity and F2 score value are 73.03 %, 93.39 % and \$75.08 %\$. Figure 7 shows the confusion metrics and ROC curve for the multiclass classification problems. All the classification curves are near to 1 value of ROC curve. Figure8 shows the accuracy and loss curve for multiclass classification problem. It can be observed that testing accuracy and loss curve converge and follows the training accuracy and loss curve which indicates the good fitting of the model.

Table5: Score table for multiclass classification

Class Level	Sensitivity	Specificity	F2 Score
Level 0	0.8175	0.91.6	0.8180
Level 2	0.7128	0.7852	0.7009
Level 3	0.7303	0.9339	0.7508

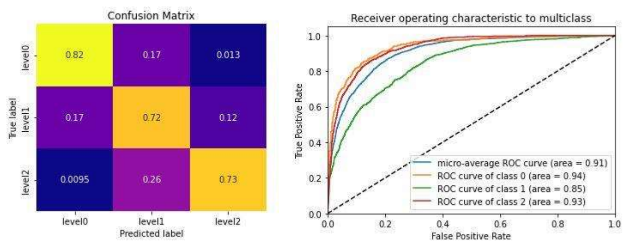

Figure7: Confusion metrics and ROC curve for multiclass classification(learning rate:0.0004 and batch size: 32.

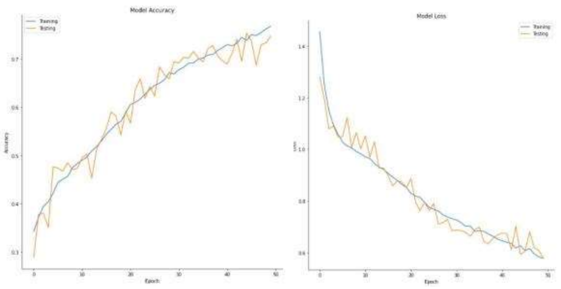

Figure8: Accuracy and loss curve for multiclass classification(learning rate: 0.0004 and batch size:32

The pretrained model such as Densenet121, InceptionV3 and VGG16 has been applied on the given dataset using transfer learning methods to compare result with that of proposed CNN model. Table 6, table 7 and table 8 shows the performance of for each class level in terms of sensitivity, specificity and F2 score value using Densenet121, InceptionV3 and VGG16 pretrained model respectively. Using Densenet121 model, F2 score value for class levels 0, 1 and 2 are obtained as 76.53 %, 66.70 % and 62.00 % respectively. Using InceptionV3 model, F2 score value for class levels 0 1 and 2 are obtained as 37.41 %, 69.39 % and 53.70 % respectively. Whereas for VGG16 model, F2 score value for class levels 0 1 and 2 are obtained as 70.58 %, 66.34 % and 68.08 % respectively. From the results it can be concluded that the performance of the proposed CNN model is better than that of pretrained models on the given dataset.

Table6: Score table for Densenet121 model

Class Level	Sensitivity	Specificity	F2 Score
Level 0	0.7625	0.8907	0.7653
Level 2	0.7114	0.6811	0.6670
Level 3	0.5811	0.9475	0.6200

Table7: Score table for Inception model

Class Level	Sensitivity	Specificity	F2 Score
Level 0	0.3277	0.9740	0.3741
Level 2	0.8258	0.4360	0.6939
Level 3	0.5015	0.9169	0.5370

Table8: Score table for VGG16 model

Class Level	Sensitivity	Specificity	F2 Score
Level 0	0.6934	0.8918	0.7058
Level 2	0.7029	0.6987	0.6634
Level 3	0.6519	0.9326	0.6808

4. Conclusions

Binary classification of DR images into referral and non referral DR images was done with good result using the proposed CNN model. The sensitivity, specificity and F2 score value for binary classification was 90 %, 93.56% and 90.65% respectively using 10 fold cross validation. Multiclass classification of DR images was done into class level 0 (no DR), class level 1 (mild DR and moderate DR) and class level 2 (severe non proliferative DR and severe proliferative DR) using the proposed CNN model. 10 fold cross validation was done in order to get the optimal learning rate and train-test split ratio which are 0.0004 and 70:30 respectively. Sensitivity, specificity and F2 score value for level 0 are 81.75 %, 91.06 % and 81.8 % respectively. For level 1, sensitivity, specificity and F2 score value are 71.28 %, 78.52 % and 70.01 % respectively. Similarly for level 3, sensitivity, specificity and F2 score value are 73.03 %, 93.39 % and 75.08 %. Since both results for binary classification problem and multiclass problem are good, the proposed methodology can work better for referral assessment of DR patients. It can be claimed that it can allows to triage patients who require referral to the ophthalmologists form those who can wait until the next screening and hence can be applied in clinical applications for screening DR patients. In future, more number of dataset can be included for training the algorithm to achieve more better prediction results for multiclass classification as well as for binary classification. Since the available images are captured under only non typical standard conditions, experimental images can be collected to train the model.

References

[1] Alyoubi W.L. and Shalash W. M. and Abulkhair M.F. Diabetic retinopathy detection through deep learning techniques: *A review,Informatics in Medicine Unlocked* 20 (2020).

[2] Torabian M. and Ghaderi M. and Keafiyeh R. Classification of diabetic and normal fundus image using deep learning method, *Leonardo Electronic Journal of Practices and Technologies,* (2018)

[3] Shanthi and Sabeenian R.S. Modified Alexnet architecture for classification of diabetic

retinopathy images, *Computers and Electrical Engineering* 76 (2019) 56-–64.

[4] Pirese R. and Avilaa S. and Wainera J. and Vall E.and Abramoff M.D. and Rocha A. A data driven approach to referable diabetic retinopathy detection, *Artificial Intelligence In Medicine,* 96 (2019) 93-106

[5] Li T., and Gao Y., Wang K., Guo S., Liu H.and Kang H. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening*, Information Sciences*, 501 (2019) 511–-522.

[6] Panel R., and Patre B. Algorithms for red lesion detection in Diabetic Retinopathy: A review, *Biomedicine and Pharmacotherapy* 107 (2018) 681-688.

[7] Momeni A. P., Seyedarabi H., Abbasi Jahromi S. H. and Javadzadeh A. Automatic Detection and Monitoring of Diabetic Retinopathy Using Efficient Convolutional Neural Networks and Contrast Limited Adaptive Histogram Equalization, 8 (2020) 136668-136673.

[8] Etienne D. Feedback on a publicly distributed image database: The Messidor database, *In: Image Analysis Stereology,* (2014).

[9] Tsiknakis N., Theodoropoulos D., and Manikis G., Ktistakis E., Boutsora O., Berto A., Scarpa F., Scarpa A., Fotiadis I. and Marias K. Deep learning for diabetic retinopathy detection and classification based on fundus images: A review, *Computers in Biology and Medicine,* 135 (2021) 104599.

[10] Vijayakumar H.S., Jayaram M.A., AshGowda K. and. Bharathi P. A. Comparative Study on Filters with Special Reference to Retinal Images, *International Journal of Computer Applications,* 138 (2018).