

Severity Grading of Diabetic Retinopathy using Deep Convolutional Neural Network

Md. Robiul Islam
Dept. of Electrical and Computer Engineering
Rajshahi University of Engineering & Technology (RUET)
Rajshahi, Bangladesh

Md. Nahid Hasan
Dept. of Computer Science and Engineering
Varendra University (VU)
Rajshahi, Bangladesh

Md. Nahiduzzaman
Dept. of Electrical and Computer Engineering
Rajshahi University of Engineering & Technology (RUET)
Rajshahi, Bangladesh

Abstract:- Diabetic Retinopathy (DR) is a major impediment of eye that is now one of the prominent sources of impaired vision due to long-term diabetes. We can save many people by early diagnosis of DR which is a reliable test that will remind patients with DR to pursue medical treatment in time. Diagnosis is a complicated process and manually fundus images are used to detect DR stages. Various computerized approaches have been proposed also. Various deep learning model showed significant performance in this context. In this study, a novel deep convolutional neural network was developed for detecting the severity stages of DR after some preprocessing of the dataset. As the dataset was imbalanced, we followed downsampling technique. Various hyper parameters performance were also studied. The results of our experiment showed that the proposed model can detect all the grades and overhead the conventional methods.

Keywords:- APTOS 2019 BLINDNESS; Diabetic Retinopathy; CNN; Deep Learning.

I. INTRODUCTION

Diabetic Retinopathy (DR) is a chronic eye condition that is now one of the leading causes of blindness around the world due to long-term diabetes. 2-6 million cases were caused by DR among the global population with mild or

extreme vision impairment in 2015 [1]. It is projected that the number of DR patients globally will rise from 382 million to 592 million by 2025 [2].

The DR is distinguished by two principal types: proliferative retinopathy (PDR) and non-proliferative diabetic retinopathy (NPDR) [3]. The NPDR is the early stages, and is split into mild stage, moderate stage and severe stage. The symptoms of the NPDR include micro aneurysms, exudates, hemorrhages, cotton wool spots, intra-retinal microvascular abnormalities, neovascularization etc. At the end of the blood vessels there is some symptoms like small red circular dots named Microaneurysm (MA) at the mild stage. In the moderate stage the MA rapidly reaches deeper layers and forms a flame-formed hemorrhages in the retina; hard exudates also appear. The severe stage not only includes both mild and moderates NPDR features (in each of the four quadrants there are more than 20 intraretinal hemorrhages), but also microvascular abnormalities with venous bleeding. The PDR is the advancing DR stage leading to neovascularization the normal development, in the form of microvascular functional structures, of new blood vessel. Only after a DR status has reached a seriously affected level (severe NPDR or PDR stage), most people who are affected by DR approach a specialist or ophthalmologist. This is because symptoms do not appear in earlier phases of DR. Therefore, automated screening systems are essential that are easily accessible to monitor the presence of DR in former stages.

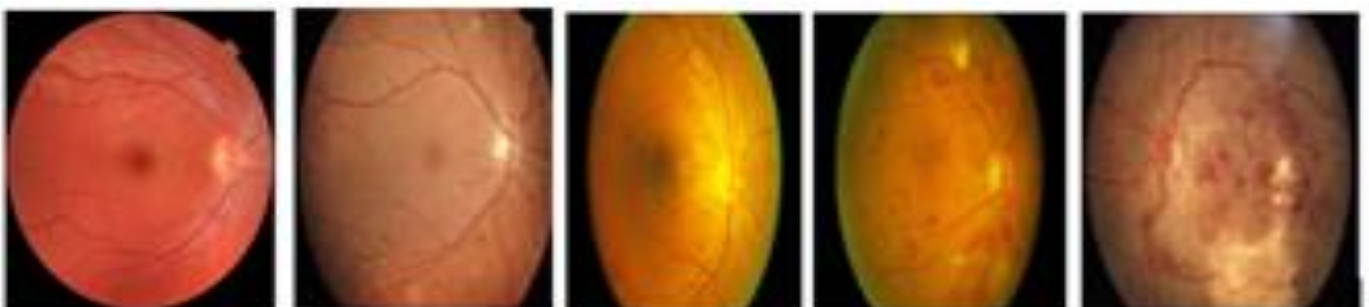


Fig. 1: Fundus Images (from normal to PDR)

For the diagnosis of DR, the colour fundus images are used. The physical analysis is carried out by highly qualified field experts and therefore is costly in case of time and expensive. It is therefore necessary to practice computer vision techniques for analyzing the fundus images automatically and support the doctors/ radiologists.

The usage of the convolutional neural networks or CNN for classification of image has become very popular in recent years. Using CNNs, large-scale image recognition tasks have been simplified [4]. CNNs often have less connections computationally than fully connected architectures, making it simpler and faster to train.

In this research, a deep convolutional neural network was developed for the detection of diabetic retinopathy severity levels. It was an image-level observation. In this analysis, an imbalanced dataset was categorized based on severity via a downsampling technique, and all DR stages were classified by our proposed method. The rest of the article is organized as follows: Section 2 defines the related work. The third section presents the preprocessing techniques and content associated with network structure design. Section 4 presents research findings and a summary in section 5.

II. RELATED WORK

Early work on the grading of diabetic retinopathy was focused on hand-crafted features such as hemorrhages, microaneurysms, hard exudates, soft exudates, optic discs etc. and blood vessel calculation. Then the grading was performed by various machine learning methods using these extracted features [5-8], i.e. K-nearest neighbor (KNN), support vector machine (SVM) and Gaussian mixture model. The main issue of these traditional machine learning approaches is that they need to extract the complicated characteristics from the fundus images manually using image processing technologies. But it is troublesome to retrieve all the complex characteristics from the Fundus pictures by this manual feature extraction. And the accuracy of the diagnosis is most often poor.

For example, Acharya et al.[5] applied morphological image processing techniques for extracting four prominent features from raw images of blood vessels, microaneurysms, exudates, and hemorrhages and fed them to the support vector machine (SVM) for automatically diagnosis eye health. Casanova [6] used methods of Random Forest and logistic regression to detect fundus images as DR or Non-DR and determine the probability of DR. E. V. Carrera et al. [7] exploited image processing techniques in order to distinguish features-- micro-aneurysms, blood vessels, and hard exudates and provided to support vector machine (SVM) for grading DR. The prime objective is to identify the severity of non-proliferative diabetic retinopathy in any retinal photo automatically. In order to detect the features which are later fed as an input to an artificial neural network (ANN) for an automated classification, Nayak et al. [8] utilized morphological image processing and texture analysis techniques on the fundus images.

Deep learning algorithms for DR grading have become popular in the last few years. For the recognition of DR severity, there are primarily two types of deep learning approaches. The first category is the lesion-level method [8, 11, 12] that uses tiny lesions' location information, such as microaneurysms, hemorrhage, to assess the output of DR grading.

Quelleg et al. [9] used CNN methods for detecting DR and segment DR lesions automatically by generating heatmaps of the convolutions layer. In order to achieve an optimal DR solution, Mansour, Romany F [10] applied AlexNet DNN. The DR model used a multilevel optimization measure which includes pre-processing, Gaussian mixture model (GMM) for segmenting concept region, connected component analysis for localizing the region of interest (ROI), high-dimensional features extraction from AlexNet DNN, linear discriminant analysis (LDA) and principal component analysis (PCA) for the selection from extracted features and support-vector-machine as classifier to ensure optimal five-class DR classification. Jiang et al. [11] developed a D-Net framework using encoder-decoder technique, which segment the blood vessels. Sheik Muhammad Saifu et al. [12] developed a novel deep convolutional neural network that detects at an early level, recognizing all DR sign microaneurysms (MAs, as well as properly assigning label images to retinal fundus images, grading into five classes. Pratt et al. [13] has built a CNN architecture and data augmentation which can recognize the complicated features involved in the Retina classification task and thus provide automated and user-free diagnosis. Different ensemble models were also suggested. Ensemble method combine several machine techniques into one predictive model [16, 17, 18].

The second category is image level observation to train a classifier model for differentiating directly between DR scores [19, 20]. V. Gulshan et al. [19] suggested an inception-v3 architecture to grade DR. Gargeya et al. [20] developed a CNN network for DR grading measurements.

Again, previous works can be categorized into two categories: first, binary classification [8, 14, 15] where the images of the fundus are classified as either Normal or DR. The severity of DR is overlooked here, another is multi-class classification where the images of the fundus are classified into five, sometimes four classes according to their phases of severity [13, 16].

III. PROPOSED MODEL

A. Dataset Description

We utilized the new Kaggle dataset named "APTOS 2019 Blindness Detection" which was given by Aravind Eye Hospital for diagnosing [21] and to prevent the disease of diabetic retinopathy among country people where medical screening is difficult to implement. 5590 high-resolution fundus images are available in the dataset. As it is a competition dataset, 3662 images come with labels of their ground truth. Each picture was scored the severity of the

diabetes retinopathy at a scale between 0 and 4: Normal is 0, Mild is 1, Moderate is 2, Severe is 3, Proliferative DR is 4.

TABLE I. DATASET CLASS DISTRIBUTION

DR Grade	Number of Examples
normal→0	1805
mild→1	999
Moderate→2	370
Severe→3	295
PDR→4	193

B. Preprocessing

Prior to the model deployment, data preprocessing or data cleaning is a key step. Prior to our model implementation we conducted multiple preprocessing measures on input data sets. The goal of pre-processing is to optimize image data to remove unnecessary distortions or to enhance those image characteristics appropriate for further processing and analysis.

Images from different sources came in the dataset. Images are subjected to a diversity of lighting environments. Some photographs are very darkened and hard to envisage. Various color lighting is often unclear. Another issue is that we can have the useless dark regions for those pictures, as seen in the figure 1(c).

1) Cropping:

It is necessary to cut out the useless regions, because when we decrease the size of the picture, the informative regions become too tiny. First of all, we cut off the useless dark areas from the photos by means of the crop circle. Here after cropping, the resultant image is a color image, e.g. we do a colour version of the crop circle because in this problem the colour image provides better information of demonstration than the gray scale image. Most of the previous works developed have modified over the grayscale image to escape computational costs, but the color version offers better demonstration information that offer better accuracy.

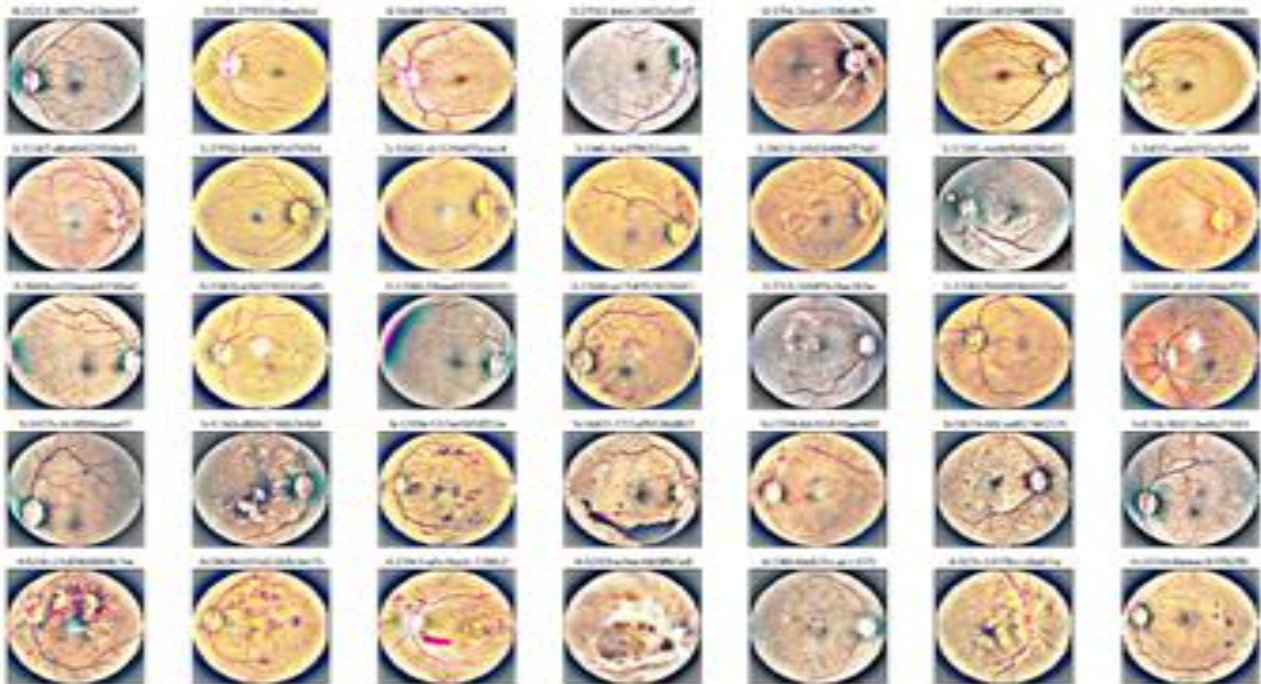


Fig. 2: Fundus Images after preprocessing

2) Resizing:

The images are high-resolution fundus images in the dataset. They have a variety of resolutions such as 1050 x 1050, 3888 x 2951, 2416 x 1736 and so on. For implementation into a deep convolutional neural network (CNN), all input images should be of the same resolutions. The lesions, the features of diabetic retinopathy in the images are very small in size. When photographs are reduced to very smaller resolution, these small lesions can often disappear. Furthermore, the parameter size can be significantly increased with a much higher resolution input. So, to unify the size of images we resize our images into a moderate size of 412 x 412 pixels.

3) Image Enhancement and Noise removal:

Images also contain various forms of noise attributable to various sources (measuring device). Next, we blur the pictures for sharpening details. Image-blurring techniques lead to noise reduction. The smoothing of the picture can be accomplished in several ways. To smooth the image, we used a Gaussian filter. The smooth picture was then separated from the original image (the resultant distinction is denoted as a mask). This means that a maximum of high-frequency components that were canceled by smoothing filters are present in the resulting picture (mask). This mask was added to the main input images to increase the high-frequency constituents.

4) **Downsampling:**

From the table I we can see that our dataset is an imbalanced dataset. This will hamper our model to learn all the grading properly. So, we downsampled the two major labels (Normal and Mild) to average size of the dataset. The rest of the labels remain the same. The downsampling process was done by random selection.

TABLE II. DATASET DISTRIBUTION AFTER DOWNSAMPLING

DR Grade	Number of Examples
normal→0	700
mild→1	700
Moderate→2	370
Severe→3	295
PDR→4	193

C. **Model Architecture Design**

We developed a deep convolution neural network shown in the fig. 4 to measure diabetic retinopathy. Convolution Neural Network (CNN) has demonstrated noteworthy computer vision and image recognition results recently. CNNs is made easy to use for large scale image recognition [4]. The ImageNet dataset was used for training a deep convolutional neural network to classify 1.2 images [22]. The building blocks of the CNN are three fundamental ones: convolution-layers, on-site activation and pooling layers. There can be several fully connected layers at the end of the classification task, along with a classification layer.

For extracting different features from the input images convolution layer is used. The convolution operation between the input image and a filter of a given size of (k x k) is done in this layer. The dot product is taken between the filter and the portions of the input image by sliding the filter over the input image with respect to the filter size of (k x k). The output is referred to as the feature map that provides us with image details, such as the corners and edges. Then, to learn some other features of the input image, this feature map is fed to other layers. By splitting the input image into small patches with a sliding window manner and calculating the max or min or average of each patch, down-sampling is performed by the pooling layer on a given input image or feature map. The activation function determines what model information should be fired forward and which information should not be given at

the end of the layer. It makes the network non-linear. There are a number of activation functions commonly used including ReLU, Softmax, tanH and Sigmoid.

Deep Convolutional Neural Networks are data driven or data consumer. As our dataset size was moderate in size, we used transfer learning technique for a good initialization of our network. Vgg19 model was used as the pre-trained model where the weights were trained on large ImageNet dataset previously. The extracted features from vgg19 were deployed to a channel-wise attention like module shown in Fig. 4. These extracted features from the vgg19 model contain the high-level demonstration of the input image and it's tough to learn the detailed feature for each grade of DR. Actually, we were inspired by the attention module of the work [23]. They proposed a novel disease-specific attention module to learn detailed semantic features of DR and DME (Diabetic Mecular Edema) diseases. From this attention module which has channel-wise attention module followed by spatial-wise attention module, we used only a kind of channel-wise attention followed by a GlobalAveragePooling layer and two final fully connected layers. We squeezed the spatial information from the extracted features of vgg19 to get two kinds of global spatial features (max and average). After that, these were fed into two dense layers (not shared as in [23]) to produce the channel-wise map. After obtaining the attention map, we multiplied (element-wise multiplication) it with the original feature maps extracted from vgg19. It provides severity-specific features

D. **Training Process**

After downsampling our imbalanced dataset we got a moderately balanced dataset. After that we split our dataset to train set for training our model and test set for testing at a ratio of (4:1). We folded our training dataset to five folds with a stratified K-fold cross validation to make sure each set contains around the same proportion of examples from every objective class as the whole set. The remaining folds were used to train the model when a fold was applied for validation. Cross Validation of stratified K-fold helped prevent over-fitting issues. After the learning was done, we calculated our model's accuracy against our test set. Again, the rate of the missed diagnosis is significant for clinical uses. So, accuracy is not enough. We further calculated the precision and recall.

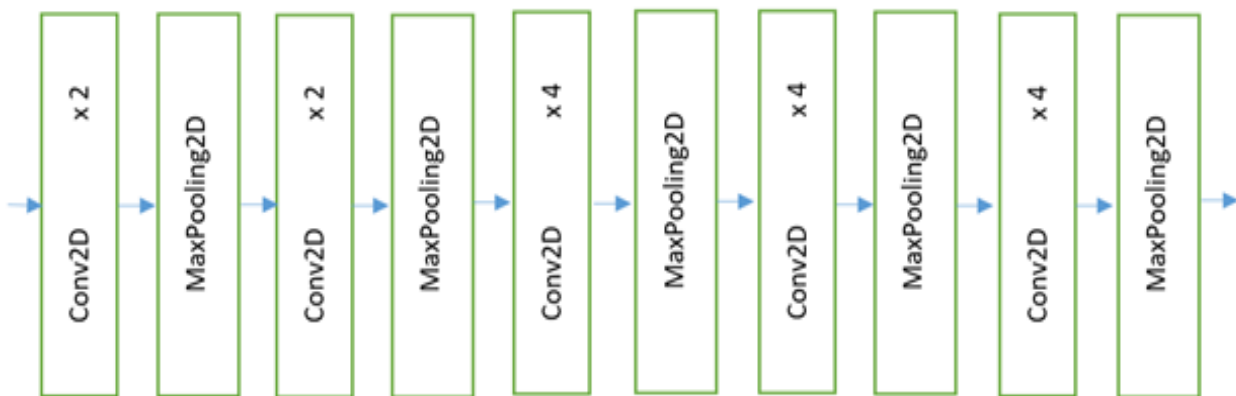


Fig. 3: Vgg19 model

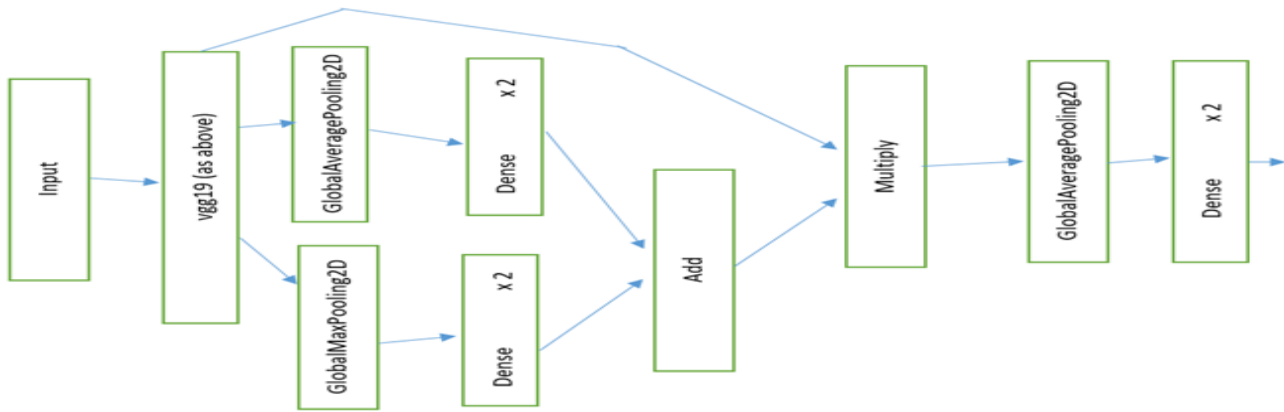


Fig. 4: proposed Model

IV. RESULT ANALYSIS

A. Experimental Setup

We ran our developed CNN model in Google Colab Pro with GPU T4 support using Tensorflow python library. One of the key hyperparameters in modern deep learning systems is Batch size. Practitioners also choose to use a larger batch size to train their model because it permits machine speedups from GPU parallelism. It is well recognized, however, that too big a batch size can result in poor generalization. We checked our model for different batch sizes. It provided best result for the batch size of 32.

Another hyper-parameter that controls how often we change our network's weights with respect to the loss gradient is the learning rate. The lower the value, the slower the downward slope we move along. We set the rate of learning to 0.01.

In addition, to minimize the loss function, optimizers update the weight parameters. The loss function serves as a reference to the terrain reminding the optimizer whether it moves in the correct direction to hit the bottom of the valley, the global minimum. For the training part of a deep learning model, it is necessary to choose an optimizer. A better optimizer will dramatically speed up the training process, avoid the low local optima, and give a better learning outcome. We used Root Mean Square Propagation (RMSProp) as the optimizing algorithm for our model.

B. Evaluation Metrics

To measure the quality of a model, various evaluation metrics are used. Using accuracy, precision and recall assessment metrics, we assessed our proposed model.

Accuracy: On a given test dataset, the accuracy of a classifier is the percentage of the test set samples correctly identified by the classifier:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{1}$$

Precision: It tells us, how many are really positive out of all the predictions that our model says are positive? Or, with all the positive predictions, what proportions are really positive? Precision may be called a measure of reliability, i.e. what

percentage is actually positive for samples labelled as positive.

$$\text{Precision} = \frac{TP}{(TP+FP)} \tag{2}$$

TP is True Positive in the formula above, and FP is False Positive. Wherever TP (True Positive) is the quantity of properly classified occurrences of the observed class, TN (True Negative) is the quantity of negative samples correctly classified by the model, FP (False Positive) is the quantity of negative occurrences improperly classified as positive, and FN (False Negative) is the quantity of positive samples inaccurately classified as negative.

Recall: Recall is a metric of completeness, i.e. what proportion is classified as such by the positive samples.

$$\text{Recall} = \frac{TP}{P} = \frac{TP}{(TP+FN)} \tag{3}$$

Confusion Matrix: A confusion matrix is a table often used to characterize a classifier's performance on a test data set for which the ground truth values are specified.

C. Results and Discussion

Table III shows the model evaluation accuracy for varying sizes of batch. We can see that with the increasing number of batch size the model's accuracy also increase but to a certain limit. We got best accuracy when the batch size was 32.

TABLE III. MODEL EVALUATION FOR VARIOUS BATCH SIZES

Batch Size	Accuracy
8	85.3%
16	88.6%
32	95.4%
40	91.2%

Fig. 5 shows the overall result of our model's evaluation. We can see that model is slightly biased towards the negative classes (No DR and Moderate) which has more training samples than others during training. Nevertheless, for highly imbalanced dataset accuracy is a deceptive metric. Although we have downsampled our dataset, it is not totally

free of imbalance. So besides accuracy we measured precision and recall as shown in table IV. Besides our model, we showed a comparison with some of the conventional methods like ResNet50, ResNet-101, and InceptionV3. We showed the average precision and recall of the five severity labels in table

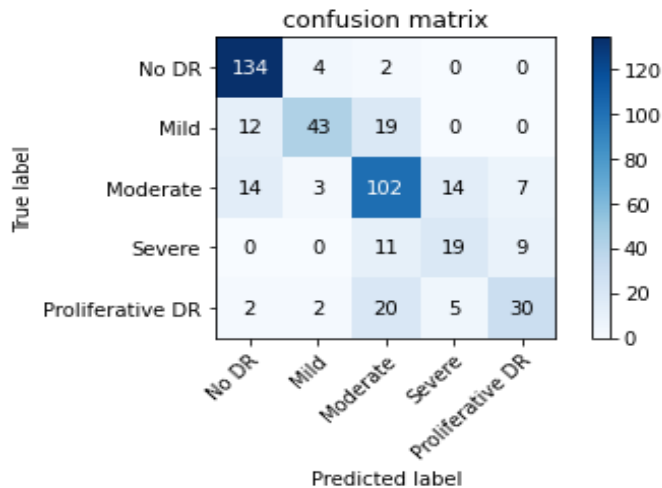


Fig. 5: Confusion Matrix for the Test Set

TABLE IV. MODEL PRECISION AND RECALL METRIC

Severity Grade	Precision	Recall
0 (No DR)	0.83	0.96
1 (Mild)	0.81	0.58
2 (Moderate)	0.66	0.73
3 (Severe)	0.5	0.49
4 (PDR)	0.65	0.51

TABLE V. MODEL COMPARISON

Models	Accuracy	Precision	Recall
Resnet-52	83.23%	58.23%	55.8%
Resnet-101	85.6%	62.36%	64.6%
Inception v3	88.4%	65.1%	66.3%
Proposed Model	95.4%	0.69	0.65%

V. CONCLUSION

Diabetic Retinopathy is a chronic eye disease that causes permanent blindness finally. Early diagnosis is can save one from blindness. For manual diagnosis, color fundus images are used but require highly domain experts and time consuming and expensive also. Different computerized methods has been developed. In this paper a novel deep convolutional neural network was developed. The model used the kaggle latest dataset on diabetic retinopathy. The color fundus images were preprocessed and downsampled as the dataset is imbalanced before feature extracting and classification using the proposed model. Study results showed that proposed model got the best evaluation accuracy of 95.4%. In case of average precision and recall our model overhead the state-of-art methods. In future, we will improve the average precision and recall metrics by embedding more fundus images.

REFERENCES

- [1]. Flaxman, Seth R., et al. "Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis." *The Lancet Global Health* 5.12 (2017): e1221-e1234.
- [2]. Jan, Sanaullah, et al. "Status of diabetic retinopathy and its presentation patterns in diabetics at ophthalmology clinics." *Journal of Postgraduate Medical Institute (Peshawar-Pakistan)* 32.1 (2018).
- [3]. Porta, Massimo, and F. Bandello. "Diabetic retinopathy." *Diabetologia* 45.12 (2002): 1617-1634.
- [4]. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [5]. Acharya, U. R., et al. "Computer-Based Detection of Diabetes Retinopathy Stages Using Digital Fundus Images." *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 223, no. 5, July 2009, pp. 545–553, doi:10.1243/09544119JEIM486.
- [6]. Casanova, Ramon, et al. "Application of random forests methods to diabetic retinopathy classification analyses." *PloS one* 9.6 (2014): e98587.
- [7]. Carrera, Enrique V., Andrés González, and Ricardo Carrera. "Automated detection of diabetic retinopathy using SVM." *2017 IEEE XXIV international conference on electronics, electrical engineering and computing (INTERCON)*. IEEE, 2017.
- [8]. Nayak, Jagadish, et al. "Automated identification of diabetic retinopathy stages using digital fundus images." *Journal of medical systems* 32.2 (2008): 107-115.
- [9]. Quellec, Gwenolé, et al. "Deep image mining for diabetic retinopathy screening." *Medical image analysis* 39 (2017): 178-193.
- [10]. Mansour, Romany F. "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy." *Biomedical engineering letters* 8.1 (2018): 41-57.
- [11]. Jiang, Yun, Ning Tan, Tingting Peng, and Hai Zhang. "Retinal vessels segmentation based on dilated multi-scale convolutional neural network." *IEEE Access* 7 (2019): 76342-76352.
- [12]. Islam, Sheikh Muhammad Saiful, Md Mahedi Hasan, and Sohaib Abdullah. "Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images." *arXiv preprint arXiv:1812.10595* (2018).
- [13]. Pratt, Harry, et al. "Convolutional neural networks for diabetic retinopathy." *Procedia computer science* 90 (2016): 200-205.
- [14]. Gondal, Waleed M., et al. "Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images." *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017.
- [15]. Yang, Yehui, et al. "Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks." *International conference on medical image computing and computer-assisted intervention*. Springer, Cham, 2017.

- [16]. Qummar, Sehrish, et al. "A deep learning ensemble approach for diabetic retinopathy detection." *IEEE Access* 7 (2019): 150530-150539
- [17]. Gao, Z., Li, J., Guo, J., Chen, Y., Yi, Z., & Zhong, J. (2018). Diagnosis of diabetic retinopathy using deep neural networks. *IEEE Access*, 7, 3360-3370.
- [18]. Jiang, Hongyang, et al. "An interpretable ensemble deep learning model for diabetic retinopathy disease classification." 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019.
- [19]. Gulshan, Varun, et al. "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs." *Jama* 316.22 (2016): 2402-2410.
- [20]. Gargeya, Rishab, and Theodore Leng. "Automated identification of diabetic retinopathy using deep learning." *Ophthalmology* 124.7 (2017): 962-969.
- [21]. Kaggle, "APTOS 2019 Blindness Detection", <https://www.kaggle.com/c/aptos2019-blindness-detection/data>, August 29, 2019, accessed January 01, 2020.
- [22]. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012): 1097-1105.
- [23]. Li, Xiaomeng, et al. "CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading." *IEEE transactions on medical imaging* 39.5 (2019): 1483-1493.