Deep learning system for assessing diabetic retinopathy prevalence and risk level estimation

Ankur Biswas¹, Rita Banik²

¹Tripura Institute of Technology, Narsingarh, West Tripura, 799009, India
²ICFAI University Tripura, Kamalghat, West Tripura, 799210, India

Abstract. Diabetic retinopathy, one of the foremost problems brought on by Diabetes Mellitus has seen an exponential rise in incidence due to the exponential growth of diabetics worldwide and causes visual issues and sightlessness owing to deformity of individual retina. An early detection and diagnosis are necessary to stop DR from progressing into severe stages and to stop blindness for which regular screening of eye is mandatory. To do this, several machine learning (ML) models are available. However, when used with bigger datasets, classical ML models either need more training time and have less generalisation in feature extraction and classification than when used with smaller data volumes. As a result, Deep Learning (DL), a newer ML paradigm that can manage a relatively small data volume with aid of effective data processing methods is presented. They do, still, often use bigger data in the deep network structure to improve feature extraction and picture classification performance. This study presents a CNN model for DR classification and compares with other variants of pre-trained DL models for initial recognition of DR through binary and multi-class classification. The attained result of 97% accuracy reveals that pre-trained ResNet model’s efficacy is better in diagnosing DR.

Keywords: Diabetic retinopathy; Deep learning; CNN; ResNet; fundus images

1. Introduction

The most common reason of visual loss among diabetics is diabetic retinopathy (DR). The strength and severity of DR is usually determined by an ophthalmologist doing a physical check of the fundus and examining colour photographs. Early detection and treatment of DR can prevent the loss of vision. So, once or twice a year, diabetic individuals are routinely recommended for retinal screening. Since, early detection is essential to stop more damage from occurring since DR is a rapidly developing condition that often results in blindness. Given the vast number of diabetes patients worldwide this procedure is both expensive and time-consuming [1]. Although, screening of DR exists to identify and treat retinopathy satisfactorily on an individual basis. The availability of ophthalmologists and the required medical infrastructure are key factors in the diabetic eye treatment. Accurate intervention
requires precise diagnosis since retinal bleeding is one of the first signs of diabetic retinopathy. Inspecting numerous photos to physically recognise lesions of various forms and dimensions is one of the main challenges that ophthalmologists have when trying to make a timely and accurate diagnosis. As a result, millions of individuals throughout the globe continue to lose their eyesight owing to a lack of adequate diagnosis and eye treatment. The implementation, administration, accessibility of grading experts, and continuing economic viability of programmes to check such a huge population for DR are all problematic. Therefore, computer-aided diagnostic methods are necessary for screening such a huge population that needs ongoing monitoring for DR and to help significantly lessen the strain on ophthalmologists. To do this, scientists are developing an automated procedure for detecting diabetic retinopathy. Moreover, to address the shortcomings of traditional diagnostic methodologies, automated systems for diagnosing retinal illness using stained fundus images have been developed [2-3]. Using such a device, distributed technologists might assess several sufferers independently and often except depending on doctors, easing the pressure on skilled professionals. Human graders may benefit from computer-aided diagnosis, which will have a big impact on the world's healthcare system in the future.

Numerous automated approaches for the conclusion of DR have been developed due to the growing awareness of the condition. Many studies have concentrated their attention on the regions of lesion related to DR since ophthalmologists will pay greater attention to these lesions when examining fundus photographs, like region around haemorrhage, microaneurysms etc. In recent years, scientists have extensively considered the application of artificial intelligence-based methods and achieved outstanding outcomes in numerous areas, notably in computer vision domain, and were possibly capable of competing against humans in some circumstances [4]. Fundus image detection method based on DL has recently advanced in terms of efficiency and reliability as compared to conventional diabetic retinopathy diagnosis techniques. In order to detect retinal haemorrhages in fundus pictures, a variety of CNN architecture is present. However, the primary investigation interest of DL-based methods is regarding how to construct a potent and effective categorization framework that frequently calls for a sizable volume of computing infrastructure to run and a sizable amount of data to fetch the superior completion of the model. A deep learning-based method that has been trained and tested on a large dataset is already available that can reasonably grade retinal pictures [5]. These kinds of algorithms might be used to help graders distinguish between eyes with and without referable retinopathy as a preliminary screener, allowing graders to focus just on those eyes. In DL based frameworks, preprocessing is often used to enhance the picture quality and expand the data, both of which are crucial for defining the intricate characteristics necessary for the segmentation job. The learning neural network subsequently underwent a remarkable improvement, successfully segmenting the bleeding with high sensitivity, specificity, and accuracy. The involvement of this study is entirely evident to the primary recognition of DR by preprocessing the eye retina photographs having established conclusions. In this work, we developed a CNN for to categorize DR. We further utilized pre-trained DL framework to grade the picture to validate our assumptions since deep learning models can automatically extract essential properties from the input image. To display the result, the image is put into the pre-trained ResNet, VGG19, DenseNet, and MobileNet. The benefit of using such architecture is that it may deliver a quicker diagnosis and grade an individual like experts. A monthly eye screening and primary recognition of DR in diabetics may dramatically lower the risk of vision loss. The accuracies obtained from the model accounts to 76 - 97%. The remainder of the article is arranged as follows: Section 2 discusses the previous research on DL-based DR grading, followed by methodologies of the technique revealed in the research in Section 3. Section 4 entails the findings of the advanced models on the APTOS dataset. Finally, the paper is concluded in Section 5.
2. Related Work

With the introduction of ML, numerous academic researchers started to investigate the application of this technique for the diagnosis of DR, which considerably aided the progress of the disease. Haloi et al. [6] proposed an approach, which first assesses if every pixel in the picture is 'microaneurysm' and subsequently diagnoses DR, uses just five layers of neural networks. When Alban et al. [7] used a pre-trained model to migrate the photos for the purpose of grading and diagnosing diabetic retinopathy, they successfully achieved a decent classification result after preprocessing the images using denoising, cropping, normalisation, and padding. A deep learning multi-task learning strategy was put out by Zhou et al. [8] to concurrently predict the precise result of fundus pictures using classification and regression. Support vector machines rather than SoftMax were employed by Qomariah et al. [9] for image classification and CNN networks rather than SoftMax for feature extraction. A Swin-Transformer network constructed using the Transformer architectural style was suggested by Liu et al. [10] and successfully completed a number of classification, detection, and segmentation tasks. This research creatively offers a combined investigatory method for segregation of eye vessels and DR (MTNet) relying the CSP UNet framework with tight sharing of parameters approach in multi-task learning to improve the grading accuracy in analysis of DR. Experiments show that the method makes it possible for the graded diagnosis of DR and the vascular segmentation job to work in tandem. Bora et al. [11] validated and developed two iterations of a DL approach to predict the development of DR in diabetic persons who received tele-retinal screening for DR in a therapeutic setting. Tang et al. [12] developed a deep learning technique that could distinguish VTDR and RDR in images. The ResNet50 CNNs learnt through popular transfer learning (TL) technique to evaluate gradeability and recognize RDR and VTDR. Hacisoffaoglu et al. [13] used the well-known AlexNet, ResNet, and GoogleNet architectures for TL approach. There are several hybrid ML-DL architectures for DR detection and classification. By collecting feature of DR from Kaggle, Pratt et al. [14] projected CNN architecture for identifying DR. The cross-entropy loss function was utilised for improvement after the colour normalization step, which also used augmented data, was used to update the weight and bias. From being learned by Stochastic Gradient Descent (SGD), the suggested CNN has a 95% specificity, 75% accuracy, and sensitivity of 30%. In order to identify DR using the Kaggle dataset, Xu et al. [15] suggested a Deep CNN model with data augmentation that employs transformations within same grade for picture classification. The two classifiers utilized in the suggested technique integrates all the feature retrieved using Gradient Boosting Machine and the extreme Gradient Boosting algorithms, while remaining classifiers employ feature obtained through CNN, either in addition to and without data augmentation. Backpropagation and SGD are used to optimize the whole network. The suggested technique has recognized MAs, RBVs, EXs, and red lesions. The accuracy rates for the suggested approach were above 90%. Comparing CNN-extracted features to traditional feature extraction techniques through data augmentation; it is shown that DL models accomplish improved results.

3. Materials and Methods

In this study, we use fundus photography and developed a CNN along with four variants of the most current and most effective CNN architecture. Fundus photography is a tool that trained medical professionals may use to examine and assess a patient's condition. The APTOS dataset [16] and the four designs that were employed in our experimental research are briefly reviewed in this section. The 3691 photos (including images from data
augmentation) in the training set had established conclusions, and the remaining eye images collected via data augmentation were used to complement the model's training. An eye picture from the collection was used for the overall confirmation.

### 3.1 The APTOS dataset

The database comprises fundus photographs that professional at Aravind Eye Hospital in India took from residents of rural areas. When this dataset was gathered, it was analysed and categorised by experts before being posted to the Kaggle competition website as APTOS Blindness detection, about sickness identification from these images, which was conducted in 2019. Also, this knowledge was presented alongside eye doctors in a series symposium. Most of the images of the fundus has been used. These images are broken down into five categories: Normal eye or No DR (0), Mild DR (1), Moderate DR (2), Severe DR (3), and Proliferative DR (4). Traditional data augmentation methods, such as zooming, flipping horizontally and vertically, are used to prepare the final dataset for training of the models.

For binary classification (2-class), No_DR is treated as one class while rest would be ‘DR’ class. As can be observed in Fig. 1, the dataset is balanced in terms of No_DR and DR class. But in multiclass classification, it would be imbalanced class problem which needs to be taken into consideration.

![Distribution of dataset images](image)

**Fig.1. Distribution of dataset images**

### 3.2 The proposed CNN and Pre-trained model

We present the design of the proposed convolutional neural network (CNN) model in this section [17-19]. One perceptron neural network-inspired learning network may be regarded as a neural network. Three layers make up this dynamic network: 1 input layer, 1 output layer, and a layer that is barely visible. The algorithm is then applied to the picture or data of the issue. The weights remaining hidden in the output layer would then emerge in various ways. When the output comprises numerous numerical values, like a binary number (for example, picture classification, where normal = 0 and abnormal = 1), a classification or recognition approach is utilised. After training with multiple photographs, the grades are weighted. The method detects the photos when fresh images that are distinct from the training images is added to the algorithm. The method identifies the grades of DR on the basis of attained weights. The convolutional sublayer serves as a pillar of the CNN, while a three-dimensional neuron matrix serves as its output matrix. Pre-trained neural networks are taken into consideration for a better understanding. A frequent strategy in conventional design is to
sandwich a pooling layer in between multiple succeeding layer. By overfitting the display, this layer aims to decrease the number of parameters and observations in the matrix, leading to a more compact vector (input) (width and height). Every depth prune in the input matrix is dealt with autonomy by the pooling layer. The position size may be changed using the max-pooling parameter. Artificial neural networks use an activation mechanism to choose an output node or "neuron" depending on an input or set of inputs. This output is referred to be the input of the node after it. This happens before a problem-solving approach is found. The outcomes are converted to a target interval, like 0 to 1 or 0 to 4 (which depends on the method of activation). For instance, the activation function in logistic regression converts all inputs to genuine interval between 0 and 1. The details of the proposed CNN is shown in Fig. 2

Fig. 2. The proposed CNN model.

The network model's specific position is determined by categorizing it through a multi-class system based on a 5-point rating scale. The output nodes range from 0 to 4 and signify different degrees of DR severity. A score of 0 indicates that there is no DR, while grades 1, 2, and 3 indicate mild, moderate, and severe DR, respectively. A score of 4 represents proliferated DR. To calculate the final model score for detecting referable DR, a weighted sum of scores is used.

### 3.3 Model Training

The model was trained using retinal fundus photographs from numerous diabetes patients in rural India whose images are accessible in the APTOS dataset along with some images obtained through traditional augmentation methods. The training set comprises of 80% fundus picture files assessed with the ICDRSS scale. The model must be tested on different files of test data which is 20% of total data. In this study, we trained our proposed CNN along with 4 (four) other variants of CNN architectures, VGG19, DenseNet169, MobileNetV3, and ResNet50V2, using pre-trained weights and biases and used the highest performing model of the three. The pretrained ResNet50V2 network from ImageNet is used in this study's architecture, with atrous convolutions serving as its primary feature extractor. In order to standardise input data, quicken the learning procedure, and minimize system sensitivities across convolutional layer and non-linearity, a batch normalisation layer is added into the framework. The FC layer needs a dropout layer which is widely used to produce an
abnormal picture enhancement. At the conclusion of the hidden layer, we used completely connected layers to demonstrate how to distinguish between images. The deep learning layer produces a fully linked layer as its output, which directs the categorization decision.

4. Results and Discussion

We next carried out an analysis and an interpretation of the data after training the networks using the dataset. The example photos divided into four categories: grade 0-4 and accompanying findings are shown in subsequent section. This provides a basic assessment of the classification quality; however, it is still limited to certain circumstances. We use the established metrics to report, evaluate, and interpret numerical findings. This enables us to draw conclusions about the effectiveness, advantages, and disadvantages of the techniques and to provide recommendations for more research and development in the future. The parameter for the proposed CNN is configured as follows: There are 32 batch sizes, 30 epochs in case of the proposed CNN while 10 epochs in pre-trained models, and patience=3 as 'Early Stopping' condition. Accuracy and Loss per epoch graph is shown in Fig. 3.

![Graph representing (a) Accuracy per epoch (b) Loss per epoch while DR grading into 5 class](image)

Fig. 3. Graph representing (a) Accuracy per epoch (b) Loss per epoch while DR grading into 5 class

Intending to assess the performance of the proposed model, we compared with four variants of CNNs with pre-trained weights as discussed earlier. Table 1 displays the model's overall DR classification results analysis for a single run.

<table>
<thead>
<tr>
<th>Sl</th>
<th>Model name</th>
<th>Train accuracy</th>
<th>Validation accuracy</th>
<th>Train time (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VGG19</td>
<td>0.6571</td>
<td>0.6809</td>
<td>236.52</td>
</tr>
<tr>
<td>2</td>
<td>DenseNet169</td>
<td>0.7114</td>
<td>0.7386</td>
<td>167.30</td>
</tr>
<tr>
<td>3</td>
<td>MobileNetV3</td>
<td>0.6207</td>
<td>0.6596</td>
<td>45.85</td>
</tr>
<tr>
<td>4</td>
<td>ResNet50V2</td>
<td>0.7303</td>
<td>0.7173</td>
<td>95.67</td>
</tr>
</tbody>
</table>

Table 1. Comparison of initial Classification descriptions
From the start, ResNet50V2 CNN model performed best on the training set. As a result, ResNet is selected to categorise DR into 5 and 2 classes. Table 2 shows the DR classification performance analysis for 5-class classification.

Table 2. Multi-class Classification descriptions

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>f1-score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>No_DR</td>
<td>0.94</td>
<td>0.97</td>
<td>0.96</td>
<td>172</td>
</tr>
<tr>
<td>Mild</td>
<td>0.65</td>
<td>0.52</td>
<td>0.58</td>
<td>50</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.59</td>
<td>0.78</td>
<td>0.68</td>
<td>95</td>
</tr>
<tr>
<td>Severe</td>
<td>0.43</td>
<td>0.15</td>
<td>0.22</td>
<td>20</td>
</tr>
<tr>
<td>Proliferative_DR</td>
<td>0.67</td>
<td>0.07</td>
<td>0.12</td>
<td>30</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Macro avg</td>
<td>0.65</td>
<td>0.51</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>Wtd. avg</td>
<td>0.76</td>
<td>0.76</td>
<td>0.73</td>
<td></td>
</tr>
</tbody>
</table>

The model performed best in recognising normal eyes ('No DR') with a precision of 0.94, followed by 'Proliferative' and 'Mild' with a precision of 0.67 and 0.65 respectively. Although, the accuracy in identifying 'Severe' was insufficient. The accuracy improves as the duration of the run increases. The loss function on the test run findings highlighted the training approach, which used backpropagation to lessen the disparity between the model's actual and predicted results. The confusion matrix [20] may be used to demonstrate the effectiveness of a classification model. The metrics in the confusion matrix are generated using the predictions and the actual grades, as shown in Fig. 4.

Fig. 4. Confusion matrix of 5-class classification

To deal with imbalance class problem where the majority of images are of healthy eyes, binary classification would be more accurate and effective. This will simply classify the images into healthy and diseased eyes, where the diseased eyes are identified as having any stage of DR. Table 2 displays an examination of the model's average DR classification performance in 2-class classification.

Table 3. Binary Classification report

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>f1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>No_DR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mild</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall accuracy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macro avg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wtd. avg</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 5 depicts the model's accuracy breakdown graph on the test data set. The results showed that the model's accuracy improved as the number of epochs increased. As compared to val_accuracy, the model has improved validation accuracy after epoch 3.

![Model Accuracy](image)

![Model Loss](image)

Fig. 5. Graph representing (a) Accuracy per epoch (b) Loss per epoch in binary grading of DR

The confusion matrix in binary classification displays the model's ability to identify the existence of DR. As illustrated in Fig. 6, the normalised confusion matrix is created using the forecasts and the ground truth and achieves an accuracy of 97%.

![Normalized Confusion Matrix](image)

Fig. 6. Confusion matrix of binary classification

### 4.1 Analytical Metrics

Precision, recall, f1-score and support are some of the measures used to analyse the outcomes that are most often seen in CNN toolboxes for the model in the identification of DR [21]. One of the most used measures for assessing classification is the accuracy. Each of these measures provides a valuable interpretation of the observed result, making them all relevant in evaluating classification results in practise. Importantly, we analyse the findings
independently for each class of DR as well as the overall outcomes (accuracy, macro average, weighted average). For us to draw pertinent conclusions about the strengths and limits, the usage of these indicators was essential. We next did a binary and multiclass classification on the risk factors for human grader and DL model findings in order to detect referable DR. To show the efficacy of the classification model, the confusion matrix is utilized. The parameters in the confusion matrix are constructed using the predicted and the actual values. The studies were done using the Python environment on the Windows platform.

5. Conclusion and Future Work

DR is a serious medical issue that causes blindness, and DL approaches may be more useful than conventional procedures for diagnosing and initial identification of the condition. However, one of the most difficult issues for ophthalmologists is detecting DR and classifying its severity levels. This study contributes by developing a model that aids in uncovering and classification of DR at several phases. As the instances available in various DR datasets is very restricted, the appropriate technique shown in this study is to apply model pre-trained with weights and biases to develop a framework being capable to categorize the DR occurrences with existing data. In this study, the diabetic retinopathy detection is accomplished using CNN-based pre-trained deep learning system. The findings demonstrate that the approach is suitable for tracking and managing DR screening at cheaper cost with optimal accuracy and lower time constraint. The best performing Xception model, achieved 96% accuracy, 94% sensitivity, and 98% specificity, which is comparable to cutting-edge models. We demonstrated that CNN models identify lesions in retina fundus images without being instructed to do so. Although previous research has focused on model performance without taking crucial aspects into consideration, our findings show that DL models are capable of detecting and displaying the desired traits.

The benefits and drawbacks of deep learning-based models developed only for diagnosis against multi-tasking objectives have yet to be established. In conclusion, the importance of efficiently utilising the benefits of medical images for DL activities, investigate the prospects of multi-tasking analytic framework for DL based eye image diagnostics, and provide a technological solution for fundus inspection cannot be overstated.

6. Acknowledgements

This research study received funding from AICTE under RPS-NER scheme. 

References

4. [E3S Web of Conferences 430, 01292 (2023)](https://doi.org/10.1051/e3sconf/202343001292)
5. [ICMPC 2023](https://doi.org/10.1051/e3sconf/202343001292)


