

# Diabetic Retinopathy Detection From Fundus Images Using Multi-Tasking Model With EfficientNet B5

Yash Bhawarkar<sup>1,\*</sup>, Kaustubh Bhure<sup>1,\*\*</sup>, Vinayak Chaudhary<sup>1,\*\*\*</sup> and Bhavana Alte<sup>1,†</sup>

<sup>1</sup>Dept. of Computer Engineering Ramrao Adik Institute of Technology, Nerul, India.

**Abstract:** Diabetic Retinopathy (DR) is a common eye disease that affects over 3 million people annually. People with diabetes are more prone to suffer from Diabetic Retinopathy. This condition can cause blurring of vision and blindness. Early detection and treatment are the most effective ways to manage Diabetic Retinopathy. Due to the huge number of diabetic patients and the need for more accurate and automatic diagnosis, the development of deep neural networks has been acknowledged. One of the issues with deep learning classification tasks, particularly in the medical field, is a lack of labelled training data. Transfer learning enables the deep learning model to be trained with a minimal Training Dataset. In Deep convolutional networks, transfer learning can be utilized to solve the problem of insufficient training data. Previous studies on deep neural networks have been promising. Through a color fundus photography study, we have developed a model to identify the different stages of Diabetic Retinopathy using Deep learning. Our proposed model has achieved 87% of accuracy with the help of EfficientNet model. The main aim of this work is to develop a robust system for detecting DR automatically.

**Key words:** CNN, Focal Loss, ResNet, EfficientNet

## 1. INTRODUCTION

A diabetic eye is a chronic illness that damages the blood vessels in the eye. It is considered the leading cause of vision impairment in the world. A comprehensive eye examination is performed to diagnose diabetes and other vision-related problems. It involves analyzing the central vision and the retina using various tests and procedures. Some of these include: Vision acuity measurements, visual complexity, and the use of a dilated pupil. Currently, the traditional method of detecting diabetic retinopathy involves manual examinations. Doing so requires a trained specialist to thoroughly examine the photos of the fundus images. Before, manual feature extraction was mainly used for detecting diabetic retinopathy. However, it can be very challenging to extract effective options from large-scale datasets due to the complexity of the task. Also, hand-woven features are sensitive to certain conditions, such as noise and photography equipment. Furthermore, as for earlier classification techniques, their generalization is limited and cannot be utilized for detecting Diabetic Retinopathy.

Our proposed model is based on Convolution Neural Network with multi-tasking model. As CNN can make predictions directly from images at the pixel level, thus they require minimal pre-processing. Using deep learning we will take the human factor of the equation. This makes the predictions reliable and superfast.

Diabetic Retinopathy affects nearly 80% of diabetics who have had the disease for more than 20 years. According to studies, 90% of such cases can be avoided if the right remedy is used and the patients' eyes are tracked. Diagnosis of Diabetic Retinopathy calls for professional understanding and we are able to locate the usage of deep studying strategies however it calls for large dataset which it lacks in healthcare and takes a lot time which we are able to dispose of in transfer learning.

## 2. RELATED WORK

In the past decades, many tasks related to automatic DR diagnosis have been performed. This paper presents a brief summary of these works, which are mainly focused on image classification techniques.

*Supriya Mishra*[1] proposed comparison based result of vgg16 and DenseNet121 models in Diabetic Retinopathy Detection using Deep Learning IEEE 2020. In preprocessing cropping, resizing and removing of black images is done and converted to NumPy. In proposed vgg16 training data is used without ImageNet dataset and in denseNet121 pretrained convolutional neural network

\*e-mail: yas.bha.rt18@rait.ac.in

\*\*e-mail: kau.bhu.rt18@rait.ac.in

\*\*\*e-mail: vin.cha.rt18@rait.ac.in

†e-mail: bhavana.alte@gmail.com

(CNN) using transfer learning. Quadratic kappa score 0.893 generated using matrix of observed score, the matrix of expected scores based on chance agreement and the weight matrix. Accuracy of vgg16 without ImageNet dataset is 73.26 which is less than DenseNet mode with ImageNet dataset. *ebin pm and p. ranjana*[2] proposed comparison based result of vgg16 and Inception V3 models in An Approach Using Transfer Learning to Disclose Diabetic Retinopathy in Early Stage IEEE Explore 2020. Proposed model pretrained on convolutional neural network (CNN) using transfer learning. Data preparation is done by using RGB Channel (Matrices). Four images are captured. First with all RGB channel. Second with only G and B channel. Third with only R and B channel. Fourth with only R and G channel. Final result generated is classified in three classes only (0,1,2). Sensitivity of proposed models for vgg16 is 0.082 and for inception V3 is 0.293. *Md. Ahsan Habib Raj, Md. Al Mamun and Md. Farukuzzaman Faruk*[3] did preprocess on fundus images as conversion to grey scale and then applied filter for noise reduction. Then proposed model masked the image for blood vessel and Micro aneurysm Detection and used VGG19 for classification. The preprocessing of the dataset as well as the feature extraction technique have a big impact on the performance of a classification model. *Nishi Doshi, Urvi Oza and Pankaj Kumar*[4] did preprocessing on fundus images as cropping, Down sampling by 8x used Bilinear and Learned Image Down sampling LID. Then they applied padding, Inception V3 network is trained to classify images. Multi-Channel CNN are used to train multiple networks in parallel. The last layers of the Inception V3 network were concatenated. The proposed model obtained accuracy of 85.2% and it was stated that eep CNN can increase the performance. *Jayant Yadav, Manish Sharma and Vikas Saxena* [5] proposed model focuses to detect DR using Computer vision and neural network with a fully automated approach and implemented using opensource tools. Model highly focuses on classification of dot hemorrhages and exudates. The accuracy achieved by the proposed model was 75%. *Savita Choudhary*[6] discussed a deep learning ConvNet model that is trained on different data sets throughout the experiment. The experiment consisted of various factors such as batch size, epoch, preprocessing and dataset training. The hyper parameters played a key role in the performance of a model and deriving concrete methods to find the optimal hyper-parameters quickly. The observation made under the experiment was the constantly increasing with tweaking of hyper parameter. *Sagar Suresh Karki and Pradnya Kulkarni*[7] firstly did Intense augmentations for better generalization, they also tried different protocols (median blur, various parameters) for augmentation but obtained results here similar. Including methods to segment only retinal blood vessels did not improve performance. Training was done on EyePacs dataset using EfficientNet. In training they freeze the classification head for 1 epoch, then unfreeze model and trained for various EfficientNet architectures using different learning rate and Optimizers. Their worked concluded stating EfficientNet B4 and B6 using their approach overfitted. Proper results can be achieved with

selective training data and proper loss function. *ZHENTAO GAO* proposed system that can give better severity ratings are more practical for clinical application[8]. The model proposed a new labeling scheme for a large dataset. A pre-processing pipeline was proposed to enable to transform fundus images into uniform formats. The testing of this pipeline's efficiency was done against several mainstream CNN models and achieved a accuracy of 88.72%. *Sumit Thorat and Akshay Chavan*[9] proposed convolutional Neural newtwork deep learning model Diabetic Retinopathy Detection by means of Deep Learning IEEE 2020.Imabalnce Dataset of EyePacs consisting of 35126 images was balanced using data augmentation of only images of class 1,2,3,4 i.e. DR leveled images. In preprocessing black images are dropout which result in better precision and recall. Neurons from convolutional neural network consisting of convolution, pooling, fully connected and dropout layer is finally passed to classification layer. Precision and recall in Severe DR i.e., class 3(0.256) and 4(0.447) is less compared to mild and no DR class.

### 3. METHODOLOGY

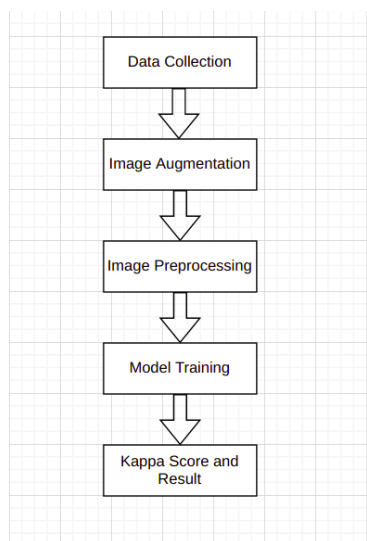
The proposed Diabetic Retinopathy (DR) detection from Fundus Images using Multi-Tasking Model with EfficientNet B5 model focuses on estimating the severity of the disease on the basis of their retinal fundus images. Model will help the medical staff to increase the efficiency and accuracy. The result produced by the Model will be from the five class. According to severity and intensity of diabetic retinopathy, one of the following class scale results is predicted: scale will be in range of 0 to 4.

- No DR (scale 0)
- Mild DR (scale 1)
- Moderate DR (scale 2)
- Severe DR (scale 3)
- Extreme Proliferative DR (scale 4).

Project will be having 4 subtasks:

1. EDA (Exploratory Data Analysis): Data will be collected, analyze in the form of fundus images and appropriate class of DR.
2. Image Processing and Augmentation:
  - 2.1 Image Augmentation: Images will be rotated, flip, padded and added to dataset to increase robustness and strength of dataset.
  - 2.2 Image processing: Images will be passed with Gaussian filter and converted to sharp grey images.
  - 2.3 Images are cropped in circular shape and Vector representation is used for images
3. Training of model: A Multi Task Learning model which will have backbone of feature extraction as ResNet50/ EfficientNetB5 and further will contain two sections, namely Classification, Ordinal regression section followed by a Single neuron with linear activation function.

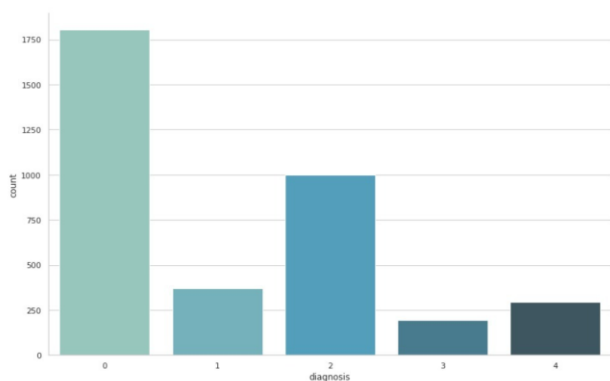
#### 4. Model Evaluation.



**Figure 1:** Workflow of System

#### 3.1 Data Collection

We have used the dataset of Fundus Images. Fundus images of retinal eye part will be used. Reflected light will be used to generate 2D image from 3D representation of semitransparent retinal images of eyes. Specialized fundus camera which contains microscope and flash is used to capture magnified and flashed images. Intensity of image in photo signifies light reflected while capturing image [10]. Collected Data is in form of fundus images and appropriate class of DR of image.



**Figure 2:** Distribution of dataset over classes

#### 3.2 Image Augmentations

Image Augmentation is the method which is used to make dataset stronger and robust. Additional images are added in dataset using previously existed images. This can be achieved by using various techniques like flipping of images, resizing images, cropping images, rotation of images, padding and shift in images [12]. Above techniques combined or alone can be used to generate more images which helps to make dataset stronger. Mirror image is generated using image augmentation.

We have created and applied following image

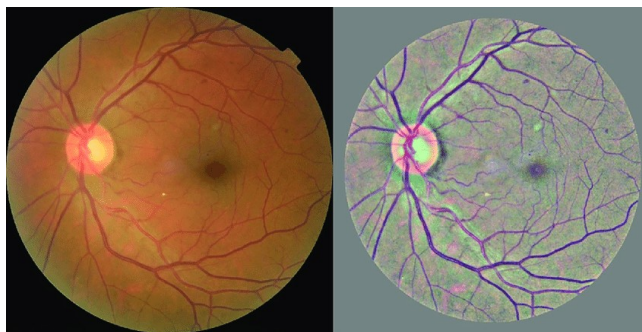
Augmentation techniques using keras image Data Generator class. We have kept height and width of images as 320 and 320 resp.

- 1)Rescale = 1/255,
- 2)Rotation = 360 and 60
- 3)Brightness\_range = [0.5, 1.5]
- 4)Zoom\_range = [1, 1.2]
- 5)Zca\_whitening = True
- 6)Horizontal\_flip = True
- 7)Vertical\_flip = True

#### 3.3 Image Preprocessing

In Diabetic Retinopathy, Higher Hemorrhages and exudates means chances of blindness is more. In our scale range 3-4 is considered as Highly affective DR. So, our main aim is to spot these Hemorrhages and exudates in given dataset of image. Images in dataset are not clear and ready to use in project [14]. Some of the images in dataset are out of focused, under focused, over focused. Some have low brightness while some have high brightness to overcome above problem, we had used low pass filter to avoid unnecessary details. We had used OpenCV library and Gaussian function which is low pass filter used to remove high frequency component from image. This is known as Gaussian Blur technique [13]. To sum up we have converted RGB fundus image to sharp grey image. We have also crop images in circular form as eye shape is circular, so we can extract more accurate features from circular image.

To carry out circular crop, we have set center at height/2 and width/2 and radius is set to min(height,width) with thickness 1.



**Figure 3:** Fundus image on left and preprocessed image on right i.e., after gaussian blur and circular crop

#### 3.4 Metric and Losses

The quadratic weighted kappa (QWK) is agreement or similarity between two predictions. The QWK is between the range of 0 to 1 i.e., from no agreement to complete agreement. There are five steps in calculating QWK

$$\kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}$$

where  $w$  is the weighted matrix,  $O$  is the histogram matrix and  $E$  being the expected matrix. 0.61 to 0.8 QWK score means Substantial agreement and 0.81 to 1 QWK score means almost perfect agreement.

## 4. SYSTEM DESIGN

The EfficientNet is considered as dynamic than the ResNet, DesNet, ResNeXt, Xception and some other CNN models. While creating EfficientNet the developers did scaling uniformly and gradually i.e., they scaled depth, resolution and width. Stem has input layer, rescaling, normalization, zero padding, Conv2D, batch norm., activation and then final layer block. The five modules and stem make all layers of EfficientNet.

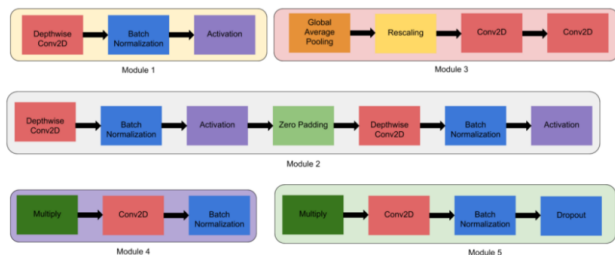


Figure 4: Five Modules of EfficientNet

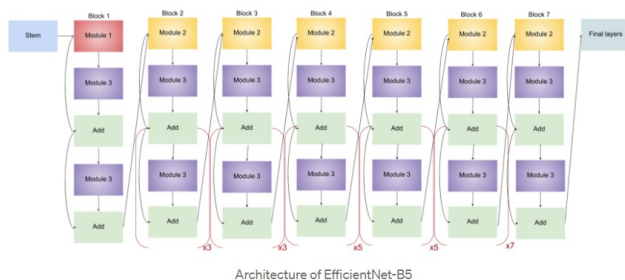


Figure 5: Layers of EfficientNet-B5

We have proposed a multi-task learning model which will have backbone of feature extraction as EfficientNet B5. The pretrained model in which include\_top layer will be set to false which will not load the Fully Connected layer for giving output.

The weights of EfficientNet on ImageNet dataset are been initialized but the last layer weights are not taken since we want to extract only feature and we don't want to predict classes of the ImageNet Dataset.

Then the extracted features are further given to Global Average Pooling layer and Global Max Pooling layer. Later the concatenation of the output is performed and then passed it to dense layer for the reduction in the length of the vector. Then dropout is implemented and the output of dropout is given to two sections namely:

1. Classification section
2. Ordinal - Regression section

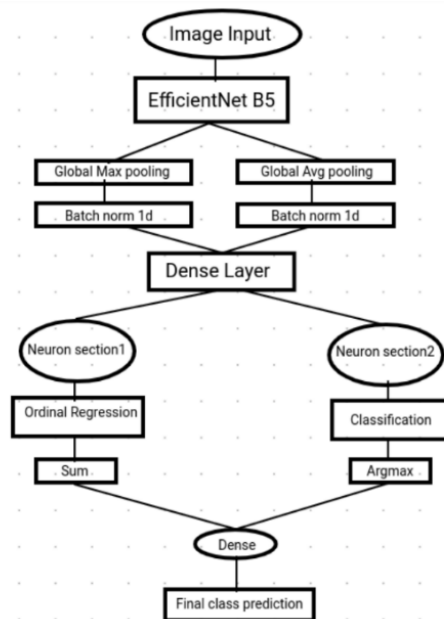


Figure 6: Flowchart of Proposed Model

## 5. MODEL TRAINING

Classification section is dense layer with 5 neurons with activation function of SoftMax, we have 5 neurons because we have to predict severity and severity scale has 0,1,2,3,4 as classes to predict. Ordinal regression section is dense layer with 5 neurons with activation function of SoftMax. Ordinal Regression section is different from Classification section with respect to loss function i.e., classification has categorical cross entropy and ordinal regression has binary cross entropy.

Further, we train the model for 20 epoch which is done for pretraining for model. Later, we change the loss function and freeze the feature extraction layer and train the model to carry out the warm up for main training.

While training the model we have changed the loss function to focal loss from cross entropy. The focal loss deals better with the data imbalance in one staged detector. The data imbalance cause ineffectiveness in training which causes degeneration of model.

The loss function reshapes the cross-entropy\_loss which reduces weights assigned to properly detected/classified examples. The scaling factor reduces to zero as the classification class prediction becomes confident. It focuses on hard negative examples. 'γ' is called the focusing parameter.

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

\*\*α t is a weighted term

\*\*p (model's estimated probability for the class with label y = 1)

Then we unfreeze all layers and train our model for around 40 to 50 epochs. The main training is done after this step. Then we take the output of classification section and apply

argmax function. The output of ordinal regression section is given to sum function.

Then processed output is given for post training to a single neuron with linear activation function.

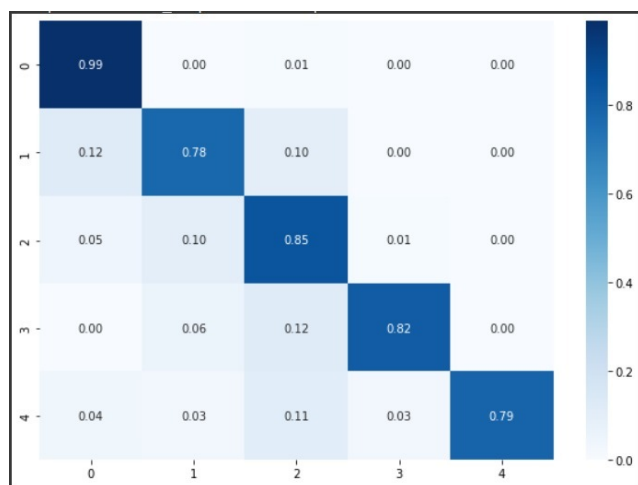
The single neuron will be trained on output for 50-100 epochs for reducing the error in the prediction.

This neuron will be giving us the predicted value and we will set threshold for determining our severity-on-severity scale from 0 to 4.

## 6. RESULTS

We have proposed a model which have image augmentation, dataset size reduction, multi-task learning model with feature extraction backbone as EfficientNet B5 with two sections as Classification and Ordinal Regression section. This is further lead by a single dense neuron with linear activation function which will predict the severity of diabetic retinopathy on severity scale 0 to 4. Hence the problem of Prediction of Diabetic retinopathy severity using Fundus image will be solved by using the proposed model. Using our approach, we got Train Cohen Kappa score as 0.870, Train Accuracy score as 0.878, Test Cohen Kappa score as 0.856 and Test Accuracy score as 0.877.

## 7. DISCUSSION AND ANALYSIS



	ResNet50	EfficientNet B5	EfficientNet B5 (Classification + Ordinal Regression)
Train Cohen Kappa score	0.832	0.841	0.870
Train Accuracy score	0.747	0.806	0.878
Test Cohen Kappa score	0.803	0.822	0.856
Test Accuracy score	0.775	0.821	0.877

We have compared three models on basis of kappa score and accuracy. Using Resnet50 model we got train Cohen kappa score as 0.832 and Train Accuracy as 0.747. changing CNN model to EfficientNet B5 we received

Train Cohen kappa score as 0.841 and Train Accuracy as 0.806 which is slightly greater than ResNet50. By dropping regression section with EfficientNet B5 with only two section (classification, ordinal regression) we got Train Cohen kappa score as 0.870 and Train Accuracy as 0.878.

Test Cohen kappa score for Resnet Model is 0.803, for EfficientNet B5 is 0.822 and for EfficientNet with two section is 0.856. Test Accuracy score for Resnet Model is 0.775, for EfficientNet B5 is 0.821 and for EfficientNet with two section is 0.877.

As above observation, we found more accurate result for EfficientNet B5 with two sections (Classification, Ordinal variable). We tried different protocols such as median blur and changes in the parameter in augmentation and preprocessing but results obtained where very similar. We also tried to segment retinal blood vessel and pass that along with hemorrhages and hard exudates but no improvement in performance was found. We also found multi-tasking models performs better than a uni-tasking model in terms of F1 score, recall and accuracy.

Due to the multiple sections used in our model, the model becomes more robust. This happens when the two sections produce different output results fed to the last neuron to predict the correct cumulative result based on previous prediction losses.

The key benefit of this approach is that it decreases variance by integrating many sections of networks that have been pretrained on a vast dataset and finely tuned on the target dataset.

## 8. CONCLUSION

We have compared ResNet50 with three sections (Classification, Regression and Ordinal Regression), EfficientNet B5 with three sections (Classification, Regression and Ordinal Regression), and EfficientNet with classification and ordinal regression section. Using EfficientNet B5 with classification and ordinal regression section, we obtained Train Cohen Kappa score as 0.870, Train Accuracy score as 0.878, Test Cohen Kappa score as 0.856 and Test Accuracy score as 0.877. We found that the preprocessing of data had a major impact on the results also the model performance is decreased if data imbalance is not considered and taken into account. The results of EfficientNet B5 with two sections also displayed that regression section does not perform quite well with classification and ordinal regression.

The research's future goals include implementing unsupervised learning and training on a unified dataset of all open-source datasets like Missidor and IDRid, as well as making optimal use of computational resources and combining them with supervised learning methods and models. Furthermore, we may conduct trials using pretrained encoders on various tasks related to eye disorders. This technique might also be used to examine meta-learning, although this would need separate additional research.

## 9. REFERENCES

1. P.M. Ebin, P.Ranjana, "An Approach Using Transfer Learning to Disclose Diabetic Retinopathy in Early Stage ", Presented at 2020 International Conference on Futuristic Technologies in Control Systems & Renewable Energy (ICFCR),Malappuram, India, (Sept 2020).
2. S. Mishra, S. Hanchate, Z. Saquib, "Diabetic Retinopathy Detection using Deep Learning", Presented at 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, (Oct 2020).
3. A. Habib Raj, A. Mamun, F. Faruk, "CNN Based Diabetic Retinopathy Status Prediction Using Fundus Images", Presented at IEEE Region 10 Symposium (TENSYP), Dhaka, Bangladesh, (June 2020).
4. N. Doshi, U. Oza and P. Kumar, "Diabetic Retinopathy Classification using Downscaling Algorithms and Deep Learning", Presented at 7th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, (Feb. 2020).
5. J. Yadav, M. Sharma and V. Saxena," Diabetic Retinopathy Detection using feedforward Neural Network", Presented at Tenth International Conference on Contemporary Computing (IC3), 9August 2017).
6. Y. Kanungo, B. Srinivasan and S. Choudhary," Detecting Diabetic Retinopathy using Deep Learning", Presented at IEEE International Conference on Recent Trends in Electronics Information & Communication Technology (RTEICT), (May 2017), India.
7. S. Karki and P. Kulkarni," Diabetic Retinopathy Classification using a Combination of EfficientNets", Presented at International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, (March 2021).
8. Z. Gao, J. Li, J. Guo, Y. Chen, Z. Yi and J. Zhong," Diagnosis of Diabetic Retinopathy Using Deep Neural Networks".
9. S. Thorat, A. Chavan, P. Sawant, S. Kulkarni, N. Sisodiya and A. Kolapkar," Diabetic Retinopathy Detection by means of Deep Learning", Presented at 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)Madurai, India, (May 2021).
10. S. M. Saiful and M. M. Hasan, "Deep Learning based Early Detection and Grading of Diabetic Retinopathy Using Retinal Fundus Images," arXiv preprint arXiv:1812.10595. (Dec 2018).
11. F. DS, A. L, G. TW, K. GL, B. G, C. JD, K. R, and A. D. Association, "Diabetic retinopathy," Diabetes care, vol. 26, no. 1, pp. 226–229, (2003).
12. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot et al., "Mastering the game of go with deep neural networks and tree search," nature, vol. 529, no. 7587, p. 484, (2016).
13. A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE Transactions on Medical Imaging, vol. 19, no. 3, pp. 203– 210, (March 2000).
14. T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, and J.-C. Klein, "Automatic detection of microaneurysms in color fundus images," Medical Image Analysis, vol. 11, no. 6, pp. 555–566, (2007).