Pokhara Engineering College Journal (PECJ)

Applied Sciences and Engineering Insights

ISSN: 3021-9795 (Print) / 3059-9628 (Online)
(Volume-2, Issue I)

Diabetic Retinopathy Detection through Multiclass Classification of Fundus Image Using Convolutional Neural Network

Sandep Gupta[1*], Ashim Khadka [2]

[1] *Department of Computer Engineering, Pokhara Engineering College, Pokhara, Nepal*
[2] *Department of Electronics and Communication Engineering, NCIT, Kathmandu, Nepal*

Corresponding email: kasaudhansandeep48@gmail.com

Received: 20-November-2024; Revised: 25-January-2025; Accepted: 03-February-2025

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. Diabetic retinopathy (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) is more widely used as a deep learning method in medical image analysis. Multiclass image classification of images into non referable DR and referable DR has been done using proposed Convolutional neural network (CNN) model. 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. Bayesian optimization has been performed for tuning learning rate and gives optimal learning rate 0.000358 through the optimization process. The customized CNN is then trained using 0.000358 learning rate and then tested on test data The images in dataset have 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 in DR image.

Keywords: *DR, CNN, CLAHE, Deep Learning, Bayesian Optimization*

1. Introduction

Diabetes is a disease that increases the amount of glucose in the blood caused by a lack of insulin (Alyoubi, Shalash and Abulkhair, 2020). 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 (Alyoubi, Shalash and Abulkhair, 2020). Diabetic Retinopathy (DR) is a complication of diabetes that causes the blood vessels of the retina to swell and to leak fluids and blood (Shanthi and Sabeenian, 2019; Torabian, Ghaderi and Kafiyeh, 2018). 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 (Alyoubi, Shalash and Abulkhair, 2020). DR is a silent disease and affects up to 80 of the diabetics around the world. Every eleventh per- son in the world suffers from diabetes mellitus, a disorder of sugar metabolism, whose prevalence is expected reach every tenth person by 2040 (Piresa et al., 2019). It is reported that approximately 1/3(34.6) of people with diabetes have DR to some extent in the U.S, Europe and Asia. It is also noted that 1 in 10 have vision-threatening DR (Li et al., 2019). The world Health Organization (WHO) and professional organization such as American Academy of ophthalmologists recommend eye examination at least once a year for DR. However poor or isolated communities often cannot afford such frequent consultants with ophthalmologists, frustrating only diagnosis treatment. Since the disease is a progressive process, medical experts suggest that diabetic patient need to be detected not less than twice a year in order for DR. However, the early screening of DR remains a challenge for several reasons. First diabetes is 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 (Li et al., 2019).

Automated DR screening through deep learning is an alternative solution to the above problems, with the advantages of high efficiency, low cost and minimal dependencies on clinicians. Deep learning method has been widely used in image analysis process. It can be 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. CNN are one of the important methods of deep learning in which several layers are trained with a powerful method. This method is very efficient and it is one of the most common methods in the various application of computer vision. CNN consist of convolution layers, max or average pooling layers and fully connected layers as main layers.

Figure 1 shows the fundus images for normal retina and DR including all those lesions (Panel and Patre, 2018). 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, which are briefly described in Table 1 No DR, mild DR, moderate DR are categorized in non-referral DR while severe DR and proliferative DR are referral DR.

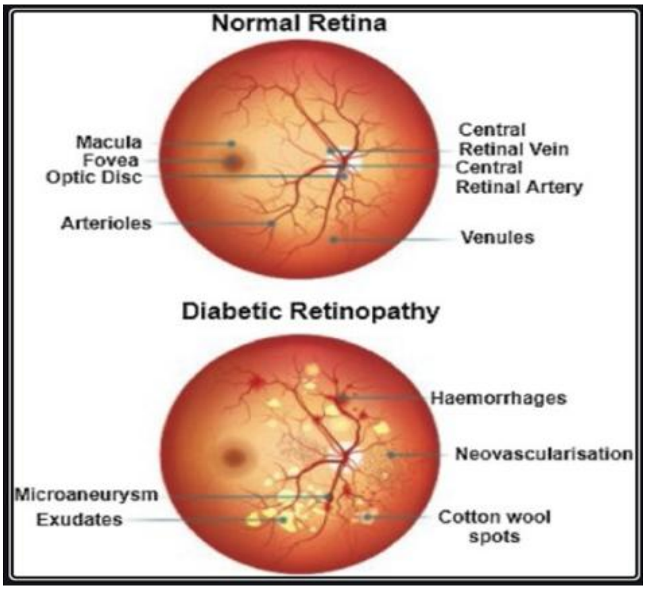

Figure 1: Difference between normal retina and diabetic retinography (Huang et al., 2017).

Table 1: Lists of diabetic retinopathies with its associative lesions (Alyoubi, Shalash and Abulkhair, 2020).

DR severity level	Lesions
No DR	Absent of lesions
Mild non-proliferative DR	Micro-aneurysm only
Moderate non-proliferative DR	more than just microaneurysm but less than severe DR
Severe non-proliferative DR	Any of the following ● more than 20 intra-retinal hemorrhage ● definite venous beeding in 2 quadrants ● Prominent intra-retinal micro-vascular abnormalities in 1 quadrant ● no signs of proliferative DR
Proliferative DR	Any of the following ● Vitreous ● Pre-retinal haemorrhages ● Neo-vascularization

2. Materials and Methods

2.1 Objective and contribution of the research

2.1.1 Objective

To detect the severity of fundus image through multiclass classification.

2.1.2 Contribution of the Research

The contributions of the research are given below

- Bayesian optimization has been performed for tuning learning rate for multi-class classification. Multiclass image classification was 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.
- Dataset were highly pre-processed through a series of multiple tasks to obtain good quality image. The images were cropped in order to remove unnecessary back- ground. 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.
- 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.
- Customized CNN algorithm was used as a deep learning methodology.

2.2 Methodology

The block diagram of the proposed model is shown in Figure 2.

2.2.1 Data Collection

Table 2 shows the number of images taken for the multiclass classification. Class 0 (no DR) includes 4, 676 images, class 1 (mild DR and moderate DR) includes 8, 658 images and class 2 (severe non-proliferative DR and proliferative DR) includes 5, 686 images.

Table 2: Total Number of image data for multiclass classification.

No DR image	Non-Severe DR image	Severe DR image	Total Number of DR image
4,676	8,658	5,686	19,020

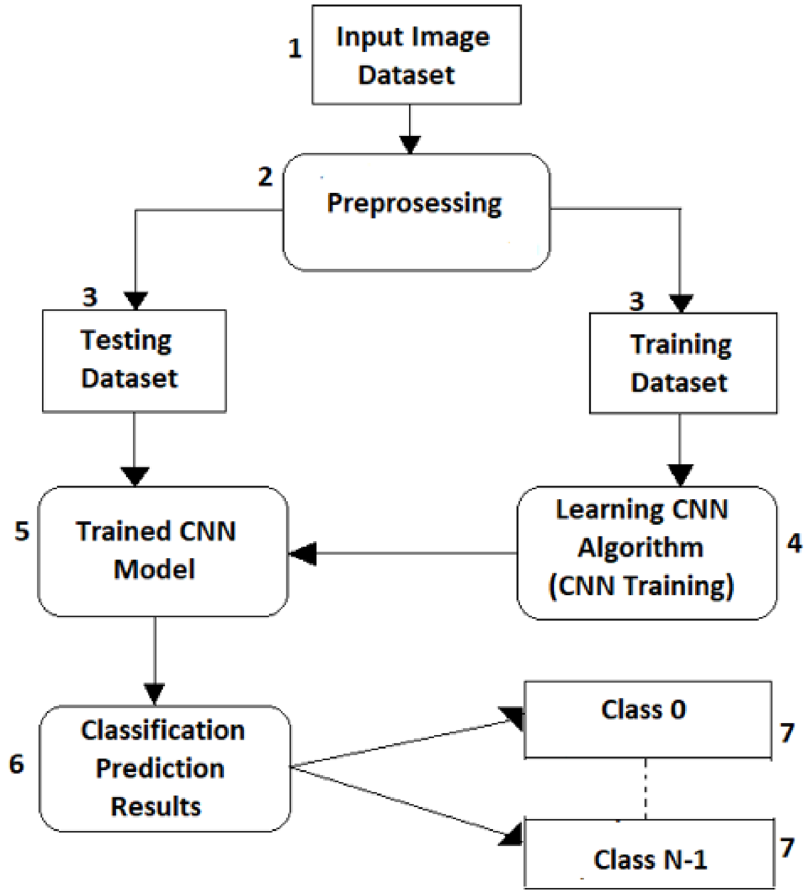

Figure 2: Block diagram of proposed model.

2.2.2 Preprocessing

The aim of preprocessing 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 preprocessing the fundus image dataset to improve the quality of the image. 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, sensing array geometry, and the movement of the eye 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.

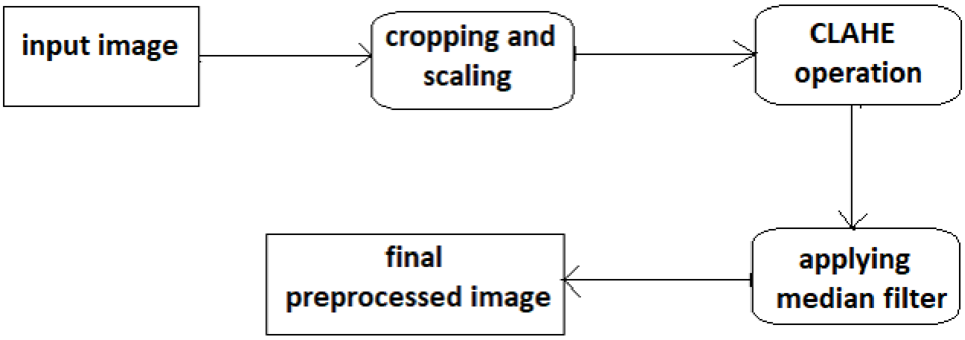

Figure 3: Preprocessing of Fundus Image.

2.2.3 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 have 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 filters 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 nonlinear 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 covariate shift. Batch normalization solves this major problem of internal covariate 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 process 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 overfitting.

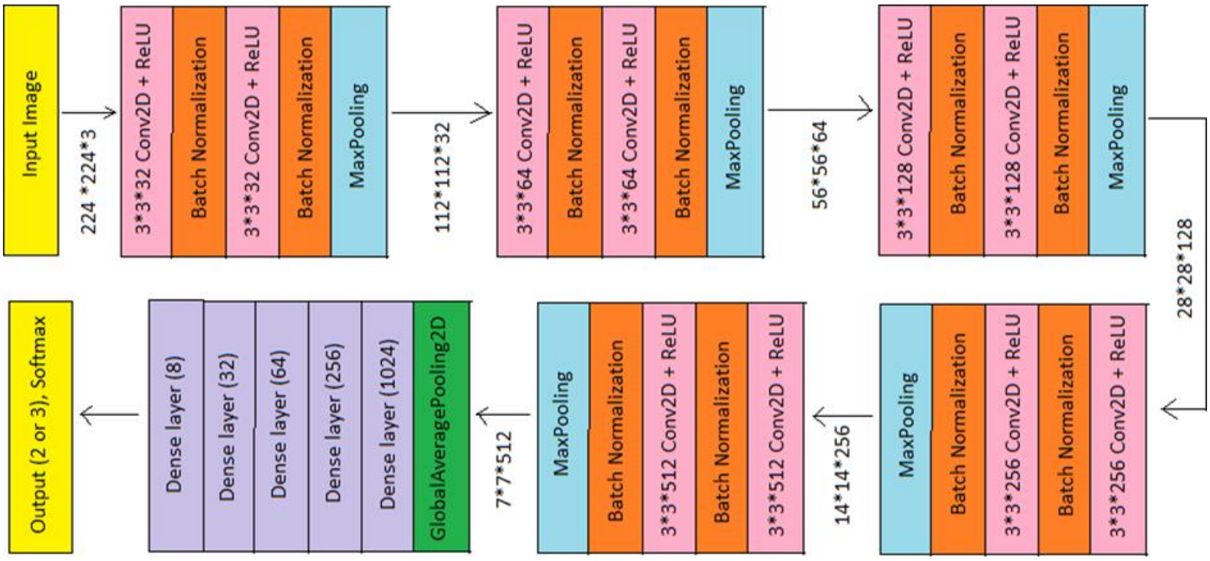

Figure 4: Structure of CNN model.

3. Results and Discussion

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. Table 3 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.

Table 3: Testing accuracy for different learning rate with different split ratio.

Learning rate	80: 20	70: 30	60: 40	50: 50
0.0001	0.6300	0.6000	0.6000	0.5900
0.0002	0.7590	0.7610	0.7600	0.7420
0.0003	0.7630	0.7680	0.7520	0.7270
0.0004	0.7820	**0.7880**	0.7340	0.7130
0.0005	0.7010	0.7090	0.6800	0.6600
0.0006	0.7420	0.7430	0.7030	0.7010
0.0007	0.7710	0.7740	0.7500	0.7320
0.0008	0.7620	0.7640	0.7360	0.7220
0.0009	0.7170	0.7180	0.6980	0.6690
0.001	0.7120	0.7050	0.6940	0.6730

Hence the proposed CNN model is trained with optimal hyperparameter set i.e. learning rate of 0.0004 and 70:30 train-test split ratio. Table 4 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% 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%. Figure 5 shows the confusion metrics and ROC curve for the multiclass classification problems. All the classification curves are near to 1 value of ROC curve. Figure 6 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.

Table 4: Score table for Multiclass classification.

Class level	Sensitivity	Specificity	F2 score
Level 0	0.8175	0.9106	0.8180
Level 1	0.7128	0.7852	0.7001
Level 2	0.7303	0.9339	0.7508

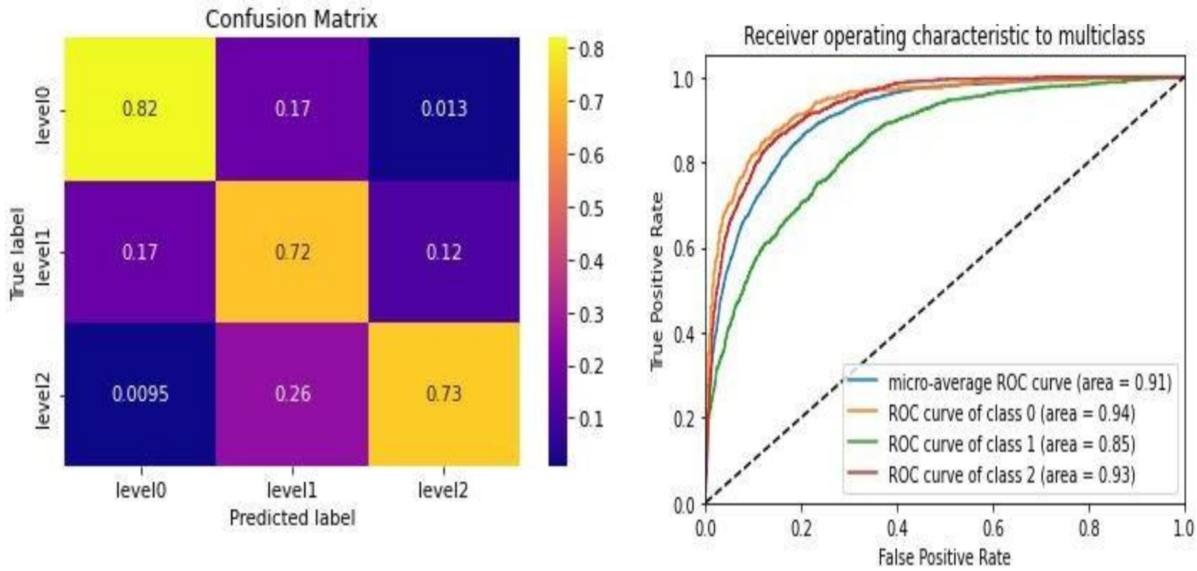

Figure 5: Confusion metrics and ROC curve for learning rate 0.0004 and batch size 32.

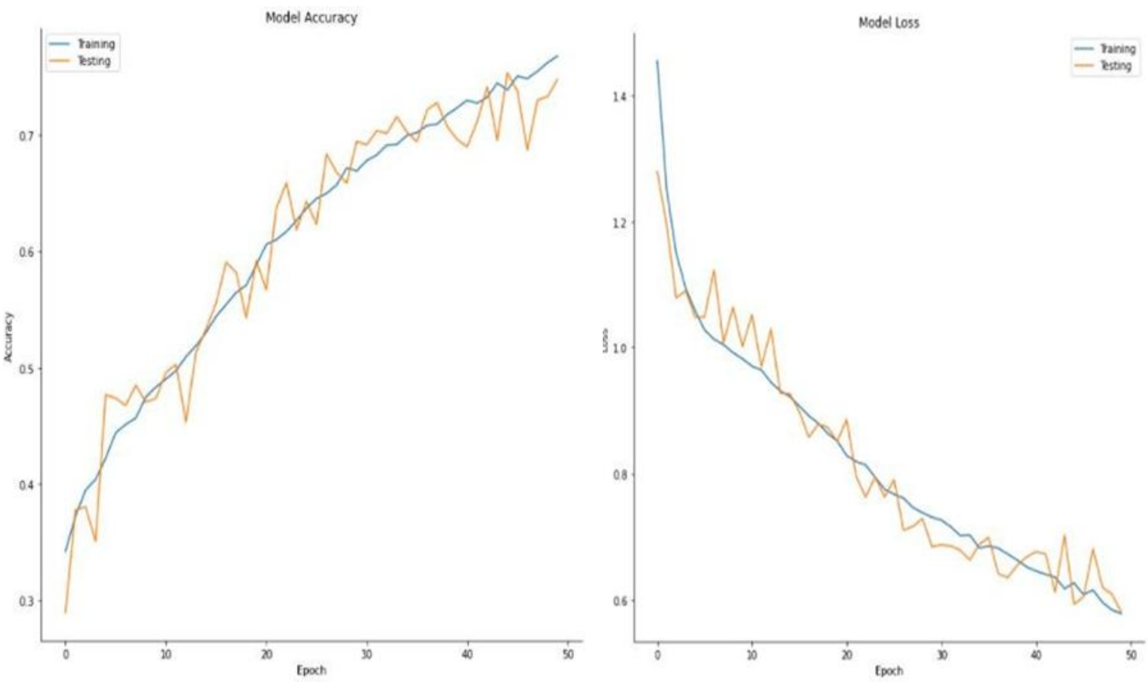

Figure 6: Accuracy and Loss Curve for multiclass classification.

3.1 Bayesian Optimization

Bayesian optimization takes into account past evaluations when choosing the hyperparameter set to evaluate next. By choosing its parameter combination in an inferred way, it enables itself to focus on the area of the parameter space that it behaves will bring the most promising validation scores. Bayesian optimization has been performed for tuning learning rate for multi-class classification and gives optimal learning rate 0.000358 through the optimization process. The customized CNN is then trained using 0.000358 learning rate and then tested on test data. Table below shows the score table for multiclass classification with learning rate tuned through Bayesian optimization. The sensitivity, specificity and F2 score value for class level 0 are 83.66 %, 91.54 % and 83.54 % respectively. For class level 1, sensitivity, specificity and F2 score values are 78 %, 84.93% and 76.77 % respectively. Similarly for class level 2, sensitivity, specificity and F2 score values are 76.07 %, 92.36 % and 77.41 % respectively as shown in table 5. Figure 7 shows the confusion metrics and roc curve for multiclass classification whereas figure 8 below shows the accuracy and loss curve. It can be observed that testing accuracy and loss curve converge and follows the training and loss curve which is better than that of former multiclass classification without using Bayesian optimization algorithm.

Table 5: Score table for multiclass classification, learning rate 0.000358.

Class Level	Sensitivity	Specificity	F2 score
Level 0	0.8366	0.9154	0.8354
Level 1	0.7800	0.8493	0.7677
Level 2	0.7607	0.9236	0.7741

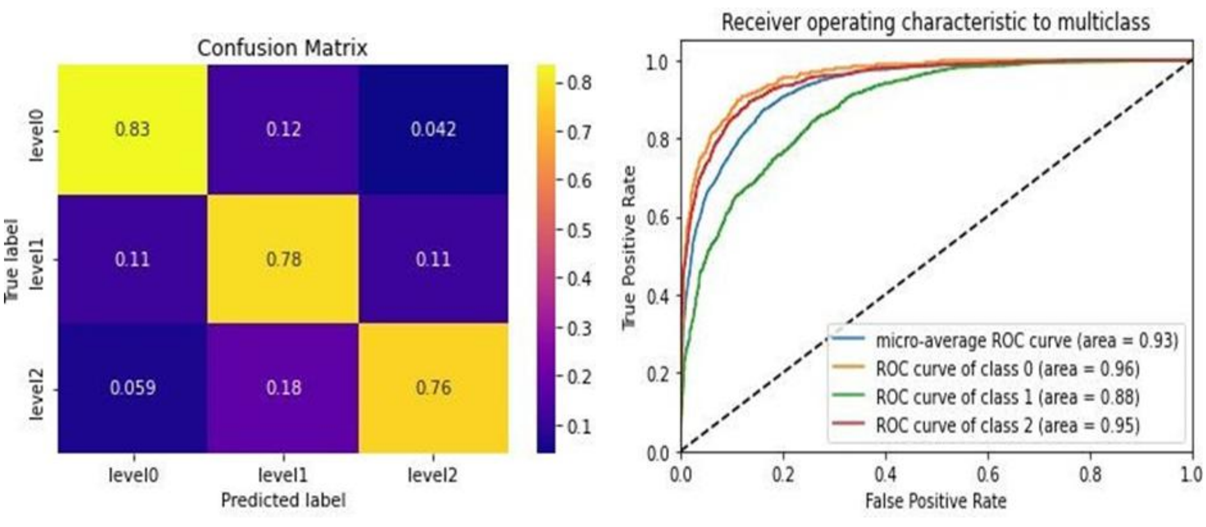

Figure 7: Confusion metrics and ROC curve for learning rate: 0.000358.

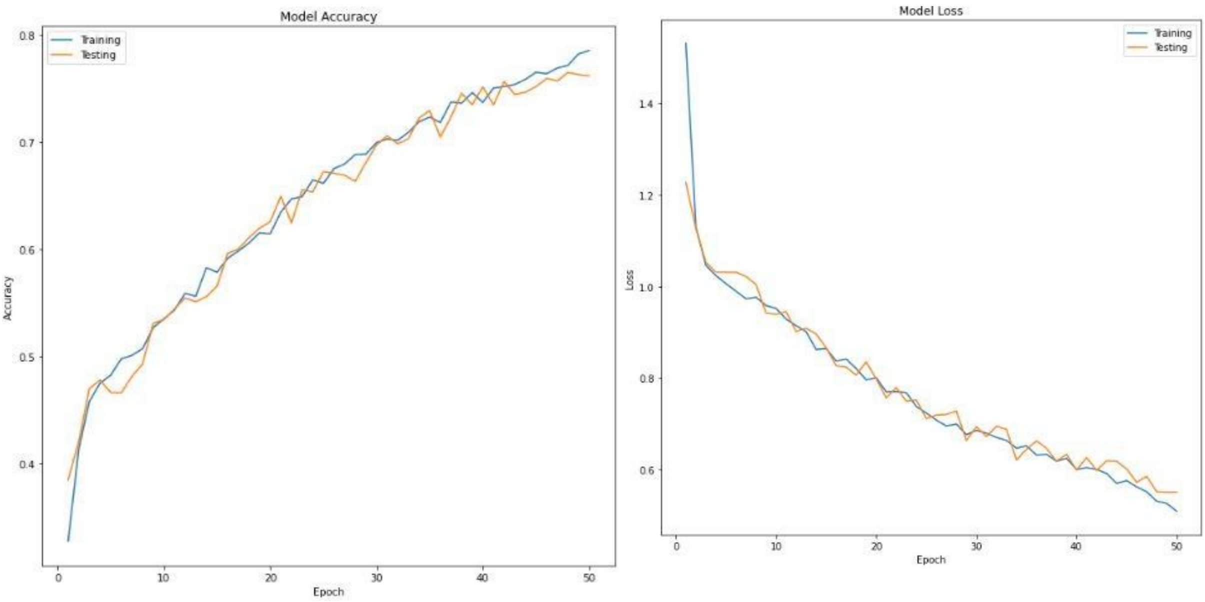

Figure 8: Accuracy and loss curve for learning rate: 0.000358.

4. Conclusions

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%. Bayesian optimization has been performed for tuning learning rate for multi-class classification and gives optimal learning rate 0.000358 through the optimization process.

Acknowledgements

I would like to acknowledge and give my warmest thanks to Ashim Khadka, PhD who made this work possible. His guidance and advice carried me through all the stages of writing my research.

References

Alyoubi, W.L., Shalash, W.M. and Abulkhair, M.F. (2020) 'Diabetic retinopathy detection through deep learning techniques: A review', *Informatics in Medicine Unlocked*, 20, 100377. Available at: https://doi.org/10.1016/j.imu.2020.100377 (Accessed: 13 February 2025).

Huang, G. et al. (2017) 'Densely Connected Convolutional Networks', *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269. Available at: https://doi.org/10.1109/CVPR.2017.243 (Accessed: 13 February 2025).

Li, T. et al. (2019) 'Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening', *Information Sciences*, 501, pp. 511–522. Available at: https://doi.org/10.1016/j.ins.2019.06.001 (Accessed: 13 February 2025).

Panel, R. and Patre, B. (2018) 'Algorithms for red lesion detection in Diabetic Retinopathy: A review', *Biomedicine and Pharmacotherapy*, 107, pp. 681-688. Available at: https://doi.org/10.1016/j.biopha.2018.08.111 (Accessed: 13 February 2025).

Piresa, R. et al. (2019) 'A data-driven approach to referable diabetic retinopathy detection', *Artificial Intelligence In Medicine*, 96, pp. 93–106. Available at: https://doi.org/10.1016/j.artmed.2019.03.005 (Accessed: 13 February 2025).

Shanthi, T. and Sabeenian, R.S. (2019) 'Modified Alexnet architecture for classification of diabetic retinopathy images', *Computers and Electrical Engineering*, 76, pp. 56–64. Available at: https://doi.org/10.1016/j.compeleceng.2019.03.017 (Accessed: 13 February 2025).

Torabian, M., Ghaderi, M. and Kafiyeh, R. (2018) 'Classification of diabetic and normal fundus images using new deep learning method', *Leonardo Electronic Journal of Practices and Technologies*, ISSN 1583-1078.