IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 1, 2021, accepted April 14, 2021, date of publication April 20, 2021, date of current version April 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3074422

Diabetic Retinopathy Detection Using VGG-NIN a Deep Learning Architecture

ZUBAIR KHAN[1], FIAZ GUL KHAN[1], AHMAD KHAN[1], ZIA UR REHMAN[1], SAJID SHAH[1], SEHRISH QUMMAR[1], FARMAN ALI[2], AND SANGHEON PACK[3], (Senior Member, IEEE)

[1]Department of Computer Science, COMSATS University Islamabad at Abbottabad, Abbottabad 22060, Pakistan
[2]Department of Software, Sejong University, Seoul 05006, South Korea
[3]School of Electrical Engineering, Korea University, Seoul 02841, South Korea

Corresponding authors: Fiaz Gul Khan (fiazkhan@cuiatd.edu.pk) and Sangheon Pack (shpack@korea.ac.kr)

This work was supported by the BK-21 Four Program through the National Research Foundation of Korea (NRF) under the Ministry of Education.

ABSTRACT Diabetic retinopathy (DR) is a disease that damages retinal blood vessels and leads to blindness. Usually, colored fundus shots are used to diagnose this irreversible disease. The manual analysis (by clinicians) of the mentioned images is monotonous and error-prone. Hence, various computer vision hands-on engineering techniques are applied to predict the occurrences of the DR and its stages automatically. However, these methods are computationally expensive and lack to extract highly nonlinear features and, hence, fail to classify DR's different stages effectively. This paper focuses on classifying the DR's different stages with the lowest possible learnable parameters to speed up the training and model convergence. The VGG16, spatial pyramid pooling layer (SPP) and network-in-network (NiN) are stacked to make a highly nonlinear scale-invariant deep model called the VGG-NiN model. The proposed VGG-NiN model can process a DR image at any scale due to the SPP layer's virtue. Moreover, the stacking of NiN adds extra nonlinearity to the model and tends to better classification. The experimental results show that the proposed model performs better in terms of accuracy, computational resource utilization compared to state-of-the-art methods.

INDEX TERMS CNN, colored fundus images, diabetic retinopathy, deep learning.

I. INTRODUCTION

Diabetes is one of the fastest-growing diseases in recent times. Recently, about 382 million people worldwide have diabetes mellitus (DM), and the future projected value of diseases is 592 million by 2025 [1]. Based on the causes and symptoms produced, there are two types of DM called type-I and type-II. Moreover, both types of DM affect vital body organs in humans, including the eye. A significant eye illness that has been reported due to DM is known as diabetic retinopathy (DR) [2]–[4]. Symptoms of DR showed that it produces mutilation of blood vessels in the retina. Among those 382 million of the population of the world, 34.6% are reported to be affected by DR. Apart from DR, proliferative diabetic retinopathy (PDR) and diabetic macular edema (DME) are reported in 7.0% and 6.8% population respectively [5]. By inferring the global situation from these figures, it is estimated that the number of DR cases will rise from 126.6 million to 191.1 million by 2030 [5]. However, reports have shown that blindness caused by DR in the Khyber Pakhtunkhwa (KP) province of Pakistan is 4% of the total population. Moreover, the common reason for DR in the province reported to be the BD type-I DR: Risk Factors Awareness and Presentation, Pakistan, 2017). In KP, 30% of the population reported having DM, with 1.6% of the total patients having type-II DM. Type-II diabetes is about 1.6% of the total population of KP, and around 30% of diabetic patients were reported as DR patients. 2% among these patients have reported having almost developed complete blindness [1]. The primary cause of DR development and its consequences are avoiding the precautionary measure for blood sugar control and a healthy lifestyle. Generally, DR in the early stages is hard to detect, and patients themselves feel asymptomatic. However, patients feel blurred vision, floaters, distortions, and visual acuity loss in advanced stages. Early detection of DR is essential, although it is of utmost necessity to eliminate the worst effects in the coming stages. The two stages of DR are called non-proliferative diabetic

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

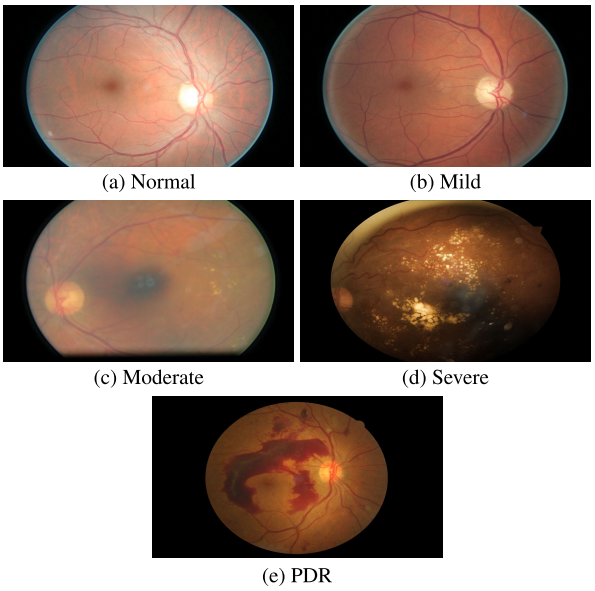

FIGURE 1. The different stages of DR [9].

retinopathy (NPDR) and PDR [6]–[8]. However, NPDR is the early stage of DR, and it is further distinguished into three sub-categories. The effect of different DR stages on the human retina is depicted in Figure 1. The distinguishing factor between these sub-categories is the presence of microaneurysms number and intensity (MA). However, a microaneurysm is nothing but a round shape localized capillary dilatations in the eye. Names of NDPR sub-categories are Mild, Moderate, and Severe. There is a small round red spot at the end of the blood capillary in the mild stage of DR. Moreover, in moderate more than five MA occurs with flame-shaped hemorrhages. In the last stage of NPDR, there are more than 20 intraretinal hemorrhages. In response to these damages, new blood vessels are formed, and the phenomenon is called neovascularization, which covers the entire inner surface of the retina.

The early detection of DR is helpful to reduce the chances of progression from NPDR to PDR. A sample of 130 patients was observed for DR symptoms. Out of 130 patients, 23.85% of the patients were reported DR positive. Among these DR positive patients, 25.8% were with PDR stage [6] which is the last stage of DR.

In this paper, our focus is to detect all five stages of DR from a given set of fundus images using minimum learning parameters. Where the fundus is the inner surface of the eye, which is nearly opposite to the lens and includes the macula, retina, fovea, optic disc, and posterior pole. A recent investigation in the matter [10], [11] has shown that CNN is widely used and preferred over the other techniques. Such CNN architecture preferences over the conventional method are because it has shown impressive results for object detection, classification, and segmentation problems [10]. However, the proposed CNN architectures [10], [11] show poor classification results on the given problem. On the contrary, this paper has discussed the use of transfer learning and improved the existing results. The proposed architecture of this paper is the Vgg16 model, which has been introduced with minor modifications. The output of Vgg16 is fed to a new version of Network in Network (NiN) architecture. Moreover, the paper concluded the effects of the Spatial Pyramid Pooling layer on the images provided.

The rest of the paper is organized as follows: Section II provide the current state-of-the-art in DR and its stages detection using machine learning. The proposed methodology and the input data's preprocessing steps are discussed in Section III. Experimental setup, performance parameters used, and results and discussion are presented in Section IV. Finally, the current work is summarized in Section V.

II. RELATED WORK

Different research work has performed on the classification of DR stages. In the given section, recent related work and their proposed methodologies explained. There are five stages of DR based on the condition of disease that gradually increases the risk of eye-sight loss [12], [13]. Researchers have proposed different models for classifying different DR stages; this section discusses the most prominent techniques. The classification mechanism is single, binary, or multiclass. On the single and binary classification, many authors have worked and got good results. Gondal *et al.* [14] proposed a CNN Referable Diabetic Retinopathy (RDR). They evaluated the network performance on two different data sets for binary classification. They did binary classification where they considered stage-0 and stage-1 as one group, and the rest of the 2,3 and 4 are grouped. Stage-1 features and lesions are challenging to detect because they may appear different or has fewer sample images. As we know, DR has five stages based on a disease that gradually increases the risk of eye-sight loss. It is always important to detect any disease at its early stage to cure it on time and never led to the most dangerous and non-curable stage, stage 4 (PDR). Yang *et al.* [15] conducted classification on the two stages of DR (NDPR and normal). He has proposed a DCNN with two networks, global and local. The local network highlights the lesions and sends them to the global network for further grading. The evaluation parameter of the kappa score was used in their study. The significant limitation of their study was that it has not considered the entire dataset of five stages. Some other studies [9], [11], [16] has shown interest in improving the classification accuracy of their proposed models. Garcia *et al.* [16] has performed on the existing CNN networks like Alexnet and VGGnet16. They have focused on the preprocessing of contract improvement. Their experiments' best-reported result was on VGG16 with 0.54 sensitivity, 0.93 specificity, and 0.83 accuracy. However, they have not explicitly mentioned the DR stages. Dutta *et al.* [11] proposed the three neural network models (NN), i.e. Feed Forward NN, Deep NN, and Convolutional NN. The preprocessing steps include the calculation of mean, median, Standard deviation, etc. Their work achieved an accuracy of 0.89 on the training dataset.

TABLE 1. Dataset: Number of instances of different classes in the dataset.

	Class - 0 (Normal)	Class - 1 (Mild)	Class - 2 (Moderate)	Class - 3 (Severe)	Class - 4 (PDR)	Total
Total Images	25,810	2,443	5,292	873	708	35,126
Training Set	16,536	1,563	3,359	558	453	22,469
Validation set	5,155	489	1,088	175	142	7,049
Testset	4,119	391	845	140	113	5,608

The five different DR stages differ in severity or progression of DR, and the severity increases from stage 0 to stage 4. The authors in [10], [17]–[22] proposed different methods for the detection of DR stages. In [23], the author achieved an accuracy of 94% on the DRIVE dataset using a deep CNN architecture to detect the DR. Drive dataset have a small number of records and have used spatial feature analysis. In [17], the author proposed an AI-based disease-staging, which takes decision based on retinal area and recommend treatments. For grading the DR, they have not used the modified Davis staging. The model's output has a false-negative rate (FNR) lower than the false positive rate (FPR). In [10], for the detection of DR, the author implemented the CNN model and the MA and HE detection in a retina. In this work, the author achieved the kappa accuracy score of 0.74, but his model could not properly classify the early stages like stage 1 and stage 2. However, it classifies these stages as class 0 means, negative class.

Furthermore, the study of Pratt *et al.* [20] for the classification of five stages of DR has proposed CNN, which did not perform well due to model's inability to classify the mild stage of DR accurately. On the other hand, they have used the skewed dataset of Kaggle, which led to high specificity and low sensitivity. RCNN Resnet is used in [24] to detect all the stages of DR. In [25], they have used SVM and deep learning to classify DR images. In their dataset, they have 170 color fundus images. In [26], the author claimed that using deep learning and backpropagation classifies all the stages of DR by using a Kaggle dataset. According to them, they speed up the training concerning the time-consuming SVM model for DR detection. In [22], the author proposed the CNN model, which only works for some stages of DR. He used ICA [27] at the last layer of CNN to localize the region of lesions. In [28], the author developed a novel Deep CNN to classify all the five stages of DR by using retinal fundus images.

Similarly, the most recent and significant work was proposed by Qummar *et al.* [9] on the Kaggle dataset with five pretrained models Resnet50, Dense169, Inceptionv3, Dense121, Xception. They achieved an accuracy of 80.8% for multi-class classification. They used a specifically stacking technique, where the stacking is created from a diverse group of strong classifiers. Their work's limitation is the ensemble model gets too complex and loss transparency in reaching a model that adjusts for its errors. They also did not provide the results for one single model or discussed which model performs well on which stage of DR. Literature shows that the researchers implemented many methods for the classification of DR. In single-stage detection, some of the lesions are not detected by the existing models, wherein the binary class classification mild stage remains unnoticed. In multi-class classification, the classes were not correctly detected by the model. Neural network models have their limitation on possibly unknown or expertly ignored learning features when the network is fed into an image and its classification, without defining diagnostically essential features and their numbers, which are essential for DR. To the best of our knowledge, multi-class classification has been done. However, all the stages are not correctly classified, especially the early stages. Our model can detect all the DR stages with minimum learn-able parameters and performs better than the current state of the art. Moreover, to the best of our knowledge, no one has discussed the model's total trainable parameters in the related work. If the trainable parameters are less than our model and are not complex, which is discussed in detail in section III-C, then the time and space complexity would also be less.

III. PROPOSED METHOD

A. DATASET DESCRIPTION

In our experimental setup, we used the Kaggle[1] dataset, which is organized by EyePacs and to the best of our knowledge, it is the largest dataset of fundus images for diabetic retinopathy. In the EyePACS dataset, there are 88,702 images; out of the 35126 are labelled images, and the remaining 53,576 images are not labelled. Our task is to classify different stages of diabetic retinopathy, which is a supervised learning problem; hence, we used only the labelled images from this dataset. In future, we can use some semi-supervised learning method where we can use the whole dataset. Dataset is distributed into five distinct classes based on the nature of the severity of DR. The distribution of different classed of DR in this dataset is shown in table 1.

B. PREPROCESSING

For pre-processing the data before providing it to the model first, we resize the images keeping in view the aspect ratio to 1349 × 1024. It helps us to avoid features loss from images. Then, images are randomly cropped to a fixed size of 1024 × 1024. Pre-processing steps are shown in figure 2. The dataset is distributed into training, validation, and test sets with a ratio of 64%, 20%, and 16%, respectively. The validation dataset is used to validate the model's improvement

[1] https://www.kaggle.com/c/diabetic-retinopathy-detection/data

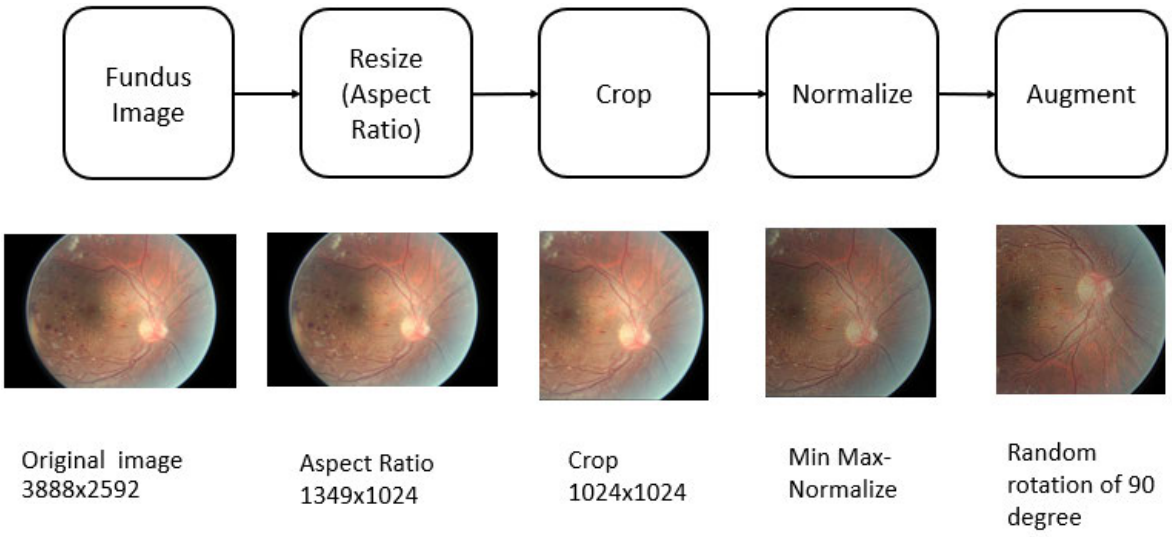

FIGURE 2. Preprocessing steps.

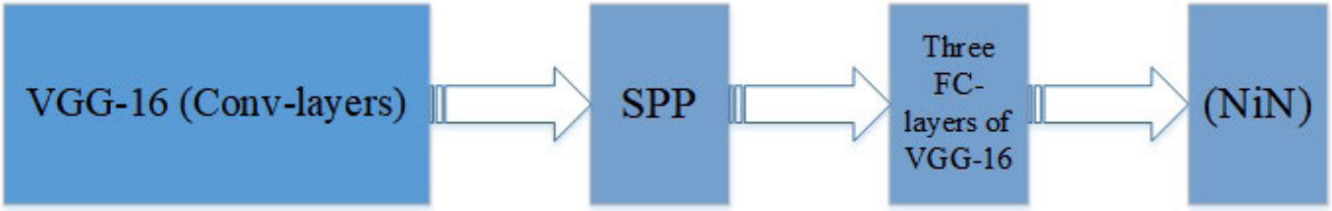

FIGURE 3. Proposed network architecture.

on each epoch. The learning rate is adoptive from 0.01 to 0.0001, seeing the improvement of validation loss to avoid over-fitting. Image augmentation is performed by using the Keras Image Data Generator with a re-scale value set to 1/225, shear, and zoom range is set to 0.2 with horizontal and vertical flip given true. The data generator automatically augments the data on run time.

C. THE VGG-NiN MODEL

The proposed model enjoys the stacking of the VGG16 [29], SPP [30], and NiN [31]. Figure 3 and Table 2 depict the block diagram and detailed architecture of the proposed model, respectively. The VGG16 [29] takes RGB image of size 224×244 as an input. The image is passed through a series of convolutional (*conv*) layers, with filters of 3×3 receptive fields, followed by a block of three fully connected layers. The *conv* layers can process inputs of varying sizes. It slides the input with a stack of kernels and produces an output feature map, $V \in \Re^{a \times a \times d}$,

$$V = f_{VGG16}(I) \tag{1}$$

where $f_{VGG16}(\cdot)$ is the VGG16 [29] network, which performs a series of convolutions and poolings to compute the features map. The feature map maintains the receptive fields responses spatially. However, the VGG16 fully connected layers need a fixed-length vector and limit the model to a fixed size input. Due to the use of fully connected layers, the VGG16 needs a fixed size input in both training and testing phases.

In our case, the dataset contains images of size 1024×1024, which unmatch the required input size. To match the VGG16 input size, we need to crop or reduce the resolution from 1024×1024 to 224×224, which leads to content loss and affects the recognition accuracy. Thus, we embed a Spatial Pyramid Pooling layer (*SPP*) [30] between the last *conv* layer and the first fully connected layer. The SPP layer pools the features and produces a fixed-size output vector compatible with the adjacent fully connected layer's requirement. In a nutshell, the *SPP* layer does information aggregation and avoids problems that arise with cropping or resolution reduction.

Let, $V \in \Re^{a \times a \times d}$ be the feature map generated by the *VGG*16 last convolutional layer. Thus, the SPP converts the varying size feature map V to a fixed size dB-dimensional vector. The B represents the number of bins, and d represents the depth (number of filters in the last convolutional layer). The SPP max-pools the input features map with arbitrary sized windows and strides. The pyramid level $p \times p$ bins are treated as a pooling level with window size, W_s and stride, S,

$$W_s = \lceil \frac{a}{p} \rceil; S = \lfloor \frac{a}{p} \rfloor \tag{2}$$

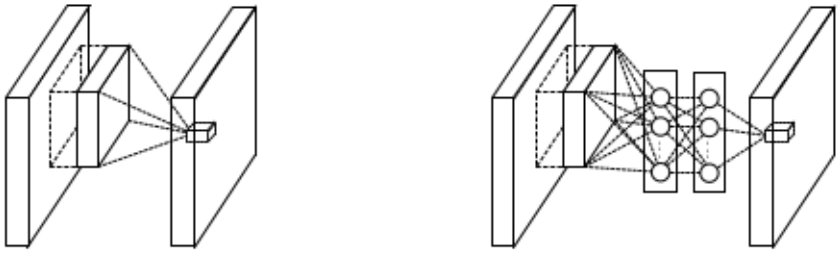

FIGURE 4. Comparison of linear convolution layer (at left) and mlpconv layer (at right) [31]. The mlpconv layer contains a micro network of multilayer perceptron.

TABLE 2. Proposed VGG-NiN model architecture.

Layer	Filters Dimensions
Con-1	$3 \times 3 \times 64$
Con-2	$3 \times 3 \times 64$
MaxPool	$2 \times 2 \times 64$
Con-3	$3 \times 3 \times 128$
Con-4	$3 \times 3 \times 128$
MaxPool	$2 \times 2 \times 128$
Conv - 5	$3 \times 3 \times 256$
Conv - 6	$3 \times 3 \times 256$
Conv - 7	$3 \times 3 \times 256$
MaxPool	$2 \times 2 \times 256$
Conv - 8	$3 \times 3 \times 512$
Conv - 9	$3 \times 3 \times 512$
Conv - 10	$3 \times 3 \times 512$
MaxPool	$2 \times 2 \times 512$
Conv - 11	$3 \times 3 \times 512$
Conv - 12	$3 \times 3 \times 512$
Conv - 13	$3 \times 3 \times 512$
MaxPool	$2 \times 2 \times 512$
SPP	(1,2,4)
FC - 1	512
FC - 2	4096
Conv - 1	$1 \times 1 \times 11$
FC - 1	400
FC - 2	400
Conv - 2	$1 \times 1 \times 11$
FC - 3	100
FC - 4	100
Conv - 3	$1 \times 1 \times 11$
FC - 5	4096
Output	5

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ denote the ceiling and floor operations respectively. The pyramids of all levels are concatenated to form a fixed-sized vector $u \in \Re^{1024}$. The vector u is further converted to a map $x \in \Re^{64 \times 64}$ to meet the input requirements of the NiN network.

The CNN assumes the latent space as linearly separable [31]. Generally, the high dimensional latent spaces contain highly non-linear manifolds. To learn the data's non-linear behaviour, we add the NiN [31] on the top of the SPP layer. The NiN is the collection of micro-networks called *mlpconv* layer, which works on the principle of multilayer perceptron. Figure 4 presents a comparison of linear *conv* layer and *mlpconv*. The NiN formed by stacking multiple *mlpconv* layers. A stack of multilayer perceptrons encode the non-linear features and lead to a high level of abstraction. We have used the parametric relu (*PRelu*) function to compute the activation of *"mlpconv"* sub-layers:

$$f_{i,j,1} = max(w_1^T x_{i,j} + b_1, \alpha(w_1^T x_{i,j} + b_1)) \quad (3)$$

$$f_{i,j,n} = max(w_n^T f_{i,j,n-1} + b_n, \alpha(w_n^T f_{i,j,n-1} + b_n)) \quad (4)$$

where n represents the number of layers in micro net *"mlpconv"* while $x_{i,j}$, w's and b's represent the feature patch centered at (i, j), the weights and biases of different sub-layers, respectively. The *PRelu* learns the rectifier parameters adaptively, and reduces the risk of model overfitting. In the last layer, softmax is adopted as an activation for classification. The categorical cross-entropy is used as a loss function. The training is performed by exploiting the Stochastic Gradient Decent (SGD) optimizer.

D. INITIALIZATION AND HYPER-PARAMETERS SETTING

The proposed model is formed by stacking the VGG network, SPP layer, and NiN model. The convolutional layers of the trained VGG network are frozen using transfer learning. However, the fully connected layers of the VGG are fine-tuned. The NiN part of the model is initialized by the Xavier method.

Table 3 presents the different hyper-parameters and their values. Initially, the learning rate is set to 0.01. Note that the learning is kept adaptive to speed up the learning process and avoid over-fitting. It is decreased by a factor of 0.1 if the validation loss doesn't improve for five successive iterations.

TABLE 3. Hyper-parameters.

Batch size	8
Initial learning rate	0.01
Momentum	0.9
Minimum learning rate	0.000,1
Number of epoch	50

IV. EXPERIMENTAL DETAIL

A. HARDWARE AND SOFTWARE

The newer version of the architecture is trained on a GPU (NVIDIA Tesla k40). This GPU is composed of 2800 CUDA core. The deep learning package Keras (http://keras.io/) is used with the TensorFlow at the back end.

TABLE 4. Comparison of the model performance measures with Sehrish *et.al.* [9].

	Recall		Precision		Specificity		F1-Score	
Classes	Our	Sehrish et. al[9]	Our	Sehrish et. al[9]	Our	Sehrish et. al [9]	Our	Sehrish et. al[9]
Class - 0	**98.0**	97.0	**90.0**	84.0	**64.0**	40.0	**94.0**	90.0
Class - 1	**31.0**	80.0	**55.0**	51.0	**97.0**	99.0	**39.0**	15.0
Class - 2	**65.0**	41.0	**76.0**	65.0	**96.0**	95.0	**70.0**	50.0
Class - 3	**33.0**	51.0	**60.0**	48.0	**99.0**	98.0	**42.0**	49.0
Class - 4	**51.0**	56.0	**54.0**	69.0	**99.0**	98.0	**53.0**	62.0
Average	**55.6**	65	**67**	63	**91**	86	**59.6**	53.2

TABLE 5. Comparison of learn-able parameters with Sehrish *et.al.* [9].

Models	Our Model Parameters	Sehrish et. al Parameters [9]
Resnet50	**0**	25,636,712
Inception V3	**0**	23,851,784
Xception	**0**	22,910,480
Dense121	**0**	8,062,504
Dense169	**0**	14,307,880
Vgg-NiN	**45,486,280**	0
Total Learn-able Parameters	**45,486,280**	94,769,360

B. PERFORMANCE PARAMETERS

To evaluate our proposed model qualitatively, we use accuracy, Area Under the Curve (AUC) [32], and Receiver Operating Curve (ROC) [33] as performance metrics.

Accuracy: Accuracy is the percentage of total accurate predictions which is based on the positive and negative classes calculations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP, TN, FP and FN represent true positive, true negative, false positive and false negative, respectively.

True positive is a number of correctly identified instances of DR images. However, True Negatives (TN) is the negative classes against the given TP instance that are correctly identified as negative. False positives (FP) are the negative classes that are predicted as positive and False Negative (FN) are positive classes that are predicted as negative.

Recall/Sensitivity: Sensitivity is also called TPR (True Positive rate), which is calculated as follows:

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative} \tag{6}$$

Specificity: It is known as TNR (True Negative Rate) It can be used as:

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive} \tag{7}$$

Precision: It is the ratio of a truly classified number of samples and the given sum of True positive and False positive:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{8}$$

F1-Score: It is the harmonic mean of recall and precision:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

TABLE 6. Comparison of AUC comparison with Sehrish *et.al.* [9].

Classes	Our Model AUC	Sehrish et. al AUC [9]
Class - 0	**89.0**	85.0
Class - 1	**82.0**	71.0
Class - 2	**70.0**	85.0
Class - 3	**90.0**	96.0
Class - 4	**88.0**	97.0
Average	**83.8**	86.8
Micro-AUC	**95.0**	95.0
Macro-AUC	**84.0**	87.0

Its values are between 0 and 1 where 1 means best score and 0 means worst.

When TPR is plotted against FPR it is called **ROC [33]:**, However, the FPR can be seen as:

$$FPR = 1 - Specificity \tag{10}$$

On other hand **AUC [32]:** is the degree of separability between classes. Higher the AUC score better the model has learned and better for evaluation and vice versa.

C. RESULTS AND DISCUSSION

This section explains the performance measures (PM) to evaluate the proposed Vgg-NiN model and compare it with state-of-art work. Apart from the PM discussion, we have already provided the hyper-parameters settings in table 3. Moreover, a detailed layer-wise explanation of the Vgg-NiN is shown in table 2.

As we know that the provided data set is imbalanced, so the class-wise values of evaluation parameters are shown in table 6 and table 4. Overall performance of state-of-art Qummar *et al.* [9] is much comparable with our proposed model. However, we can see the comparison of learnable

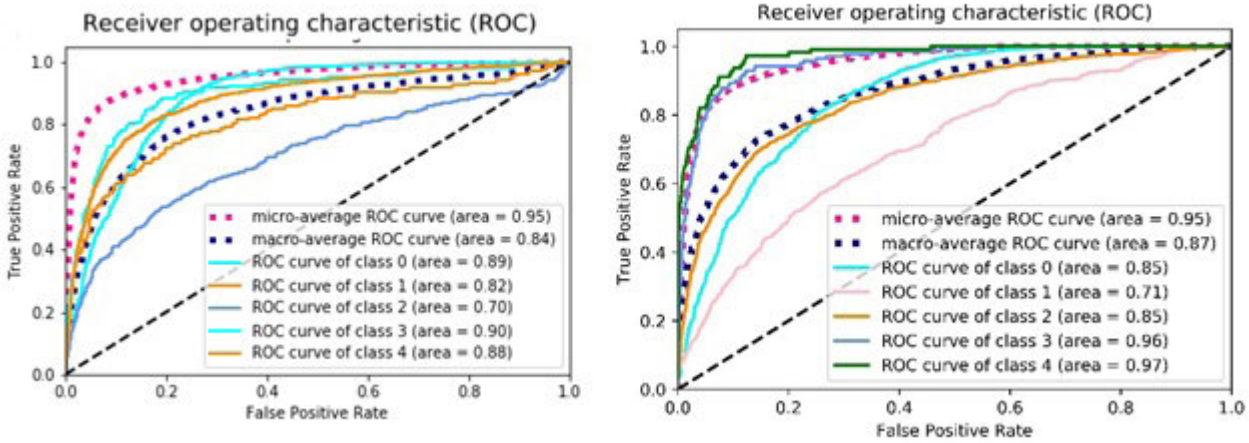

FIGURE 5. Comparison of ROC-curve of VGG-NIN model (at left) with Sehrish's model (at right).

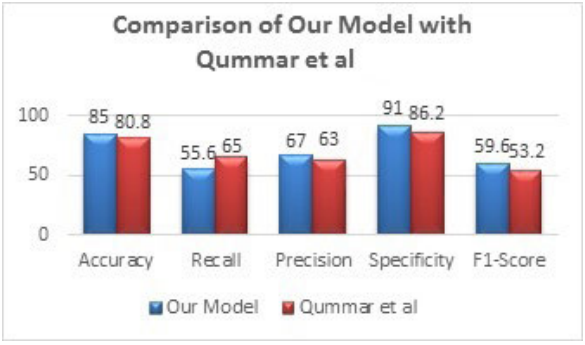

FIGURE 6. Comparison of our model with Qummar *et al.*

parameters in table 5. It shows that our learnable parameters are half of the existing model proposed by Qummar *et al.* On the other hand, all other classes have higher average values of different performance parameters (Precision, Accuracy, Specificity and F1-Score). The precision difference between Qumar *et al.* and Vgg-NiN is 5%. The overall performance is compared in table 4 and Figure 6.

Further, we have calculated the ROC curve to evaluate the model. It is considered one of the best PM for classification models. The ROC curve explains the ability of a model to distinguish among classes. Moreover, the micro-average and macro-average are essential to explain the overall performance of the model. In the case of imbalanced datasets, as in our case, importance is given to the micro-average ROC shown in figure 5. The results are also compared with the competitive methods. The micro-average ROC sum up the individual false and true positive and false negative. Whereas on another side, the macro-average depends on the average precision and recall. As discussed earlier, the micro-average ROC is taken into consideration when the dataset is imbalanced. Results clearly show that our model with 52% lower parameters gives the same micro-AUC of 0.95. Consequently, the proposed model utilizes lower computational resources that provided similar results.

V. CONCLUSION

This paper is an extension of our work [9] in which we proposed the deep learning-based ensemble approach for diabetic retinopathy detection. The major drawback of the ensemble model is the number of learnable parameters. In this paper, we brought architectural changes in existing CNN to enhance the efficiency and accuracy of classification of the DR's stages in color fundus images and reduce the number of learnable parameters. We used imbalanced versions of the Kaggle dataset to validate the performance measures of the proposed model. The results depict that the proposed model is low in computation and better than other state-of-the-art ensemble and non-ensemble methods. In the future, we plan to bring some other productive changes in the existing model's architecture and some preprocessing techniques and discuss how these changes affect the working of a model on the classification of DR's stages, especially the early ones.

ACKNOWLEDGMENT

The authors would like to thank Nvidia Corporation for providing a support by donating them a Telsa K-40 GPU. *(Zubair Khan and Farman Ali are co-first authors.)*

REFERENCES

[1] S. Jan, I. Ahmad, S. Karim, Z. Hussain, M. Rehman, and A. A. Shah, "Status of diabetic retinopathy and its presentation patterns in diabetics at ophthalomogy clinics," *J. Postgraduate Med. Inst. (Peshawar-Pakistan)*, vol. 32, no. 1, 2018.

[2] L. Math and R. Fatima, "Adaptive machine learning classification for diabetic retinopathy," *Multimedia Tools Appl.*, vol. 80, pp. 5173–5186, Oct. 2020.

[3] A. He, T. Li, N. Li, K. Wang, and H. Fu, "CABNet: Category attention block for imbalanced diabetic retinopathy grading," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 143–153, Jan. 2021.

[4] L. Andersen and P. Andersson, "Deep learning approach for diabetic retinopathy grading with transfer learning," Tech. Rep., 2020.

[5] N. Congdon, Y. Zheng, and M. He, "The worldwide epidemic of diabetic retinopathy," *Indian J. Ophthalmol.*, vol. 60, no. 5, p. 428, 2012.

[6] W. R. Memon, B. Lal, and A. A. Sahto, "Diabetic retinopathy," *Prof. Med. J.*, vol. 24, no. 2, pp. 234–238, 2017.

[7] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automatic detection of diabetic eye disease through deep learning using fundus images: A survey," *IEEE Access*, vol. 8, pp. 151133–151149, 2020.
[8] R. E. Putra, H. Tjandrasa, and N. Suciati, "Severity classification of non-proliferative diabetic retinopathy using convolutional support vector machine," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 4, pp. 156–170, Aug. 2020.
[9] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. Ahmed Khan, and W. Jadoon, "A deep learning ensemble approach for diabetic retinopathy detection," *IEEE Access*, vol. 7, pp. 150530–150539, 2019.
[10] R. Ghosh, K. Ghosh, and S. Maitra, "Automatic detection and classification of diabetic retinopathy stages using CNN," in *Proc. 4th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2017, pp. 550–554.
[11] S. Dutta, B. C. Manideep, S. M. Basha, R. D. Caytiles, and N. C. S. N. Iyengar, "Classification of diabetic retinopathy images by using deep learning models," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 1, pp. 89–106, Jan. 2018.
[12] X. Zeng, H. Chen, Y. Luo, and W. Ye, "Automated diabetic retinopathy detection based on binocular siamese-like convolutional neural network," *IEEE Access*, vol. 7, pp. 30744–30753, 2019.
[13] T. Nazir, A. Irtaza, Z. Shabbir, A. Javed, U. Akram, and M. T. Mahmood, "Diabetic retinopathy detection through novel tetragonal local octa patterns and extreme learning machines," *Artif. Intell. Med.*, vol. 99, Aug. 2019, Art. no. 101695.
[14] W. M. Gondal, J. M. Kohler, R. Grzeszick, G. A. Fink, and M. Hirsch, "Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2069–2073.
[15] Y. Yang, T. Li, W. Li, H. Wu, W. Fan, and W. Zhang, "Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2017, pp. 533–540.
[16] G. García, J. Gallardo, A. Mauricio, J. López, and A. D. Carpio, "Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images," in *Proc. Int. Conf. Artif. Neural Netw.* Springer, 2017, pp. 635–642.
[17] H. Takahashi, H. Tampo, Y. Arai, Y. Inoue, and H. Kawashima, "Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy," *PLoS ONE*, vol. 12, no. 6, Jun. 2017, Art. no. e0179790.
[18] Z. Wang, Y. Yin, J. Shi, W. Fang, H. Li, and X. Wang, "Zoom-in-net: Deep mining lesions for diabetic retinopathy detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2017, pp. 267–275.
[19] Q. Abbas, I. Fondon, A. Sarmiento, S. Jiménez, and P. Alemany, "Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features," *Med. Biol. Eng. Comput.*, vol. 55, no. 11, pp. 1959–1974, Nov. 2017.
[20] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Comput. Sci.*, vol. 90, pp. 200–205, 2016.
[21] N. B. Prakash, D. Selvathi, and G. R. Hemalakshmi, "Development of algorithm for dual stage classification to estimate severity level of diabetic retinopathy in retinal images using soft computing techniques," *Int. J. Elect. Eng. Inform.*, vol. 6, no. 4, 2014.
[22] J. de la Torre, A. Valls, D. Puig, and P. Romero-Aroca, "Identification and visualization of the underlying independent causes of the diagnostic of diabetic retinopathy made by a deep learning classifier," 2018, *arXiv:1809.08567*. [Online]. Available: http://arxiv.org/abs/1809.08567
[23] C. T. R. Kathirvel, "Classifying diabetic retinopathy using deep learning architecture," *Int. J. Eng. Res.*, vol. V5, no. 6, pp. 19–24, May 2016.
[24] A. Kind and G. Azzopardi, "An explainable AI-based computer aided detection system for diabetic retinopathy using retinal fundus images," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Springer, 2019, pp. 457–468.
[25] E. Bhatti and P. Kaur, "DRAODM: Diabetic retinopathy analysis through optimized deep learning with multi support vector machine for classification," in *Proc. Int. Conf. Recent Trends Image Process. Pattern Recognit.* Springer, 2018, pp. 174–188.
[26] P. N. S. Kumar, R. U. Deepak, A. Sathar, V. Sahasranamam, and R. R. Kumar, "Automated detection system for diabetic retinopathy using two field fundus photography," *Procedia Comput. Sci.*, vol. 93, pp. 486–494, 2016.
[27] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, 2000.
[28] S. Muhammad Saiful Islam, M. Mahedi Hasan, and S. Abdullah, "Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images," 2018, *arXiv:1812.10595*. [Online]. Available: http://arxiv.org/abs/1812.10595
[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[30] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[31] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, *arXiv:1312.4400*. [Online]. Available: http://arxiv.org/abs/1312.4400
[32] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997.
[33] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

ZUBAIR KHAN is currently pursuing the Ph.D. degree with COMSATS University Islamabad at Abbottabad under the Indigenous Scholarship of Higher Education Commission of Pakistan. He is also doing research-based activities with the Balochistan Forest and Wildlife Department for plants nursery automation and robotics. His research interests include plants detection and identification from images, medical image diagnosis using deep learning, machine learning with focus on deep learning, image processing, and data mining.

FIAZ GUL KHAN received the master's and Ph.D. degrees (Hons.) from Politecnico di Torino, Italy, in 2013. He is currently serving as an Assistant Professor with the Computer Science Department, COMSATS University Islamabad at Abbottabad, Pakistan. His research interests include concurrent computing, machine learning, artificial intelligence, and GPU computing.

AHMAD KHAN received the Ph.D. degree from the National University of Computer and Emerging Sciences (FAST- NU), Islamabad, Pakistan, in 2015. He is currently working as an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad (CUI) at Abbottabad. His research interests include computer vision, machine learning, and evolutionary algorithms.

ZIA UR REHMAN received the Ph.D. degree in information systems from Curtin University, Perth, Australia, in 2015. He is currently an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad (CUI) at Abbottabad, Pakistan. His research interests include cloud computing, machine learning, and computer vision.

SAJID SHAH received the M.S. and Ph.D. degrees from Politecnico di Torino, Italy. He is currently working as an Assistant Professor with COMSATS University Islamabad at Abbottabad. His research interests include data mining, text mining, machine learning, bioinformatics, and image processing.

SEHRISH QUMMAR is currently pursuing the M.S. degree with COMSATS University Islamabad at Abbottabad. She is also doing research with Qingdao Huanghai University, Zhumadian, China. Her research interests include medical image diagnosis using deep learning, image processing, machine learning, and data mining.

FARMAN ALI received the B.S. degree in computer science from the University of Peshawar, Pakistan, in 2011, the M.S. degree in computer science from Gyeongsang National University, South Korea, in 2015, and the Ph.D. degree in information and communication engineering from Inha University, South Korea, in 2018. From September 2018 to August 2019, he worked as a Postdoctoral Fellow with the UWB Wireless Communications Research Center, Inha University. He is currently an Assistant Professor with the Department of Software, Sejong University, South Korea. He has published more than 50 research papers in peer-reviewed international journals and conferences and registered over four patents. His current research interests include sentiment analysis/opinion mining, information extraction, information retrieval, feature fusion, artificial intelligence in text mining, ontology-based recommendation systems, healthcare monitoring systems, deep learning-based data mining, fuzzy ontology, fuzzy logic, and type-2 fuzzy logic. He has been awarded with the Outstanding Research Award (Excellence of Journal Publications-2017) and the President Choice of the Best Researcher Award during graduate program at Inha University.

SANGHEON PACK (Senior Member, IEEE) received the B.S. and Ph.D. degrees in computer engineering from Seoul National University, Seoul, South Korea, in 2000 and 2005, respectively. From 2005 to 2006, he was a Postdoctoral Fellow with the Broadband Communications Research Group, University of Waterloo, Waterloo, ON, Canada. In 2007, he joined the faculty of Korea University, where he is currently a Full Professor with the School of Electrical Engineering. His research interests include future Internet, softwarized networking (SDN/NFV), information centric networking/delay tolerant networking, and vehicular networks. He was a recipient of the IEEE/Institute of Electronics and Information Engineers Joint Award for IT Young Engineers Award, in 2017, the Korean Institute of Information Scientists and Engineers Young Information Scientist Award, in 2017, the Korean Institute of Communications and Information Sciences Haedong Young Scholar Award, in 2013, the LG Yonam Foundation Overseas Research Professor Program, in 2012, and the IEEE ComSoc APB Outstanding Young Researcher Award, in 2009.