Classification of Diabetic Retinopathy Disease with Transfer Learning using Deep Convolutional Neural Networks

Krishnamoorthy SOMASUNDARAM[1], Paulraj SIVAKUMAR[1], Durairaj SURESH[2]
[1]*Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu, India*
[2]*Department of Computer Science and Engineering, PSNA College of Engineering and Technology, Dindigul, Tamil Nadu, India*
sks.it@psgtech.ac.in

***Abstract*—Diabetic Retinopathy (DR) stays a main source of vision deterioration around world and it is getting exacerbated day by day. Almost no warning signs for detecting DR which will be greater challenge with us today. So, it is extremely preferred that DR has to be discovered on time. Adversely, the existing result involves an ophthalmologist to manually check and identify DR by positioning the exudates related with vascular irregularity due to diabetes from fundus image. In this work, we are able to classify images based on different severity levels through an automatic DR classification system. To extract specific features of image without any loss in spatial information, a Convolutional Neural Network (CNN) models which possesses an image with a distinct weight matrix is used. In the beginning, we estimate various CNN models to conclude the best performing CNN for DR classification with an objective to obtain much better accuracy. In the classification of DR disease with transfer learning using deep CNN models, 97.72% of accuracy is provided by the proposed CNN model for Kaggle dataset. The proposed CNN model provides a classification accuracy of 97.58% for MESSIDOR dataset. The proposed technique provides better results than other state-of-art methods.**

***Index Terms*—computer aided diagnosis, image classification, learning, neural networks, retinopathy.**

I. INTRODUCTION

The Diabetic Retinopathy (DR) is a persistent disease in which initial signs are difficult to notice and predict. Besides, if DR disease does not get identified and cured properly in time, it may cause permanent blindness. Identifying retinal fundus diseases in advance helps ophthalmologists in applying appropriate treatments and cures the diseases or reduces its severity and therefore protects patients from acute blindness. Hence, it is important to identify the retinal features in the fundus image of the eye for DR disease diagnosis and early detection. Recently, most of the research works have been developed for diabetic retinopathy disease diagnosis. But, the effectiveness of diabetic retinopathy disease diagnosis is not sufficient for early detection.

Basically, DR affects blood vessels in the retina which is light sensitive portion of the eye. Nowadays, it becomes the foremost root of visualization mutilation and blindness for working group people in the world. Almost no warning signs for detecting DR which will be greater challenge with us today. So, it is extremely preferred that DR has to be discovered on time. Adversely, the existing DR detection result in practice is almost insufficient to manage this necessity. Specifically, the existing result involves an ophthalmologist to manually check and identify DR by positioning the exudates related with vascular irregularity due to diabetes from the digital fundus retinal image. Even though this existing manual method is accurate, it takes a lot of time and finding reliable ophthalmologists is again a great challenge. The experimental classification process consists of detection of other valid features, such as micro aneurysms, hemorrhages, lesions and exudates on retinal images of the eye as in Fig. 1. In general, using deep learning for classification has improved accuracies in detecting the DR disease, but combined-phase classification results are less attractive, especially for early detection of the disease. To resolve this concern, an automatic diagnosis of DR disease is the order of the day.

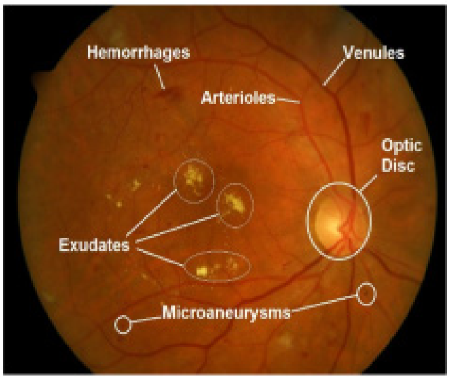


Figure 1. Few salient features in the retinal fundus image [15]

In this work, we are able to categorize images based on four different severity levels through an automatic DR classification system. To extract specific features of image without any loss in the spatial information, a Convolutional Neural Network (CNN) which possesses an image with a distinct weight matrix is used. In the beginning, we estimate various CNN models to conclude the best performing CNN for the image classification and disease diagnosis with an objective to obtain much better accuracy percentage.

The rest of the article is organized as follows: related work is presented in the section 2, followed by the dataset and its preprocessing methods in the section 3, while methodology followed and the experimental results are discussed in the section 4 and 5 respectively. Finally, we draw our conclusion and future plan in the section 6.

Digital Object Identifier 10.4316/AECE.2021.03006

II. Related Works

Monzurul Islam et al. [1] developed an algorithm to identify diabetic retinopathy in the retinal eye image using a combined approach of machine learning and image processing. This computer aided technique categorizes diabetic retinopathy effectively from the retinal images which are taken from the dataset such as, Diabetic Retinopathy Image Dataset and Structured Analysis of the Retina. Diah Wahyu Safitri et al. [2] developed a combined approach called fractal analysis and K Nearest Neighbor (KNN) algorithm to identify the diabetic retinopathy disease in the retinal eye image.

Omer Deperlioglu et al. [3] implemented a modified histogram equalization technique in order to enhance the underwater images. In this approach underwater image from RGB color space is converted into HSV space, in which modified histogram equalization is applied to the V channel of the HSV space. Then again modified HSV space is converted back to RGB color space to get good quality images. Ritesh Beohar et al. [4] proposed an underwater image enhancement algorithm using Contrast Limited Adaptive Histogram Equalization (CLAHE) and 2D median filter. This method is tested on different underwater images and it is evaluated using the measured such as, Peak Signal to Noise Ratio, Root Mean Square Error, and Mean Brightness Error.

Nishanthan Ramachandran et al. [5] implemented a deep CNN approach to identify and detect the diabetic retinopathy in the retinal eye image. In this method receiver operating characteristic curve is presented to demonstrate the capability of a deep CNN network to classify diabetic retinopathy in the retinal eye image. Gen-Min et al. [6] proposed a computer aided technique based on deep learning to detect and classify diabetic retinopathy disease. This method compares the performance of detecting diabetic retinopathy between input retinal fundus image and the entropy output by deep learning.

Aujih AB, et al. [7] implemented based on the segmentation of retinal vessel and severity classification of diabetic retinopathy (DR) from retinal funduscopic image and analysis of retinal vessel segmentation with deep learning and its effect on diabetic retinopathy classification. Sreejini et al. [8] implemented a diabetic retinopathy disease detection technique based on the combined approach of particle swarm optimization and Fuzzy C-Means (FCM) Clustering algorithm. Rajendra acharya et al. [9] proposed a technique to identify stages of diabetic retinopathy disease in the retinal eye image. Four stages of diabetic retinopathy namely normal retina, non-proliferative diabetic retinopathy, proliferative diabetic retinopathy, and macular edema can be identified using this method.

Becherer et. al. [10] suggested that the initialization gives an advantage however the initialization isn't higher that complete training at the target. Next, in the have a look at of various source tasks, they determined that pre-training on a dataset with an extensive sort of finely grained images gives a higher source task than a smaller sort of greater fashionable classes. Furthermore, pre-training on a dataset and not using a overlap with the target dataset gives much less of an advantage than a greater visually comparable source dataset. The quantity of training data in the goal task appears to have little correlation with the advantage furnished with the aid of using parameter fine-tuning. Lastly, they have a look at the impact of fine-tuning ensembles of networks. As expected, an ensemble of fine-tuned networks outperforms an ensemble of randomly initialized networks. However, it additionally appears to be the case that a fine-tuned network can outperform a whole ensemble of randomly initialized networks. The results display that parameter fine-tuning nearly usually results in a development in image class accuracy. Given those results, they see no purpose now no longer to make use of parameter fine-tuning. The best viable drawback is the multiplied training time required for pre-training. However, if the use of a current network structure with existing training models, there can be no boom in training time.

R. J. Borgli et al. [11] suggested the benefit of application of deep learning in Computer-Aided Diagnosis (CAD) structures. These structures help physicians in the analysis of disorders and anomalies in the usage of visual data. However, hyper-parameter optimization is commonly executed manually, taking a long term and with a threat of now no longer locating the satisfactory parameters for classification accuracy. Using transfer learning, they purpose for better accuracy in anomaly detection, and they described a machine for computerized hyper-parameter optimization of Convolutional Neural Network (CNN) models. The presented methodology makes use of Bayesian optimization and is used to give experiments with three optimization techniques mechanically optimizing hyper-parameters for CNN models on datasets. With those experiments, they tested that computerized hyper-parameter optimization is a powerful method for improving overall performance in transfer learning.

III. Dataset and Preprocessing

For DR classification, the dataset used were from public Kaggle website (https://www.kaggle.com/c/diabetic-retinopathy-detection), which involved in developing DR detection model and non-publicly available MESSIDOR dataset (https://www.adcis.net/en/third-party/messidor/), which include retinal images for diagnosis and classification.

A. Public Kaggle Dataset

The Kaggle dataset contains 35,126 high resolution colored retinal fundus image with resolution of about 3500 X 3500 in various imaging environments including poor contrast, low resolution, etc. The images in the dataset are graded by ophthalmologist into 5 groups from group-0 to group-4, which is accordingly to the Fig. 2.

The presence of DR in the image is categorized as in Table I. The distribution of the image groups is highly unfair as the image in group-2 is almost 8 times as that of image in group-4 and is taken care by utilizing the technique of Light Gradient Boosting Decision Tree (Light GBM) [12]. Light GBM heaps incessant feature values into isolated bins to speed up the training process. It supplants incessant values using isolated bins to lessen memory norm. It can construct abundant complex trees by ensuing a leaf shrewd fragmented method, resulting in achieving higher accuracy. It also provisions both feature parallel and data parallel. The method can handle large scale data with corporate support.

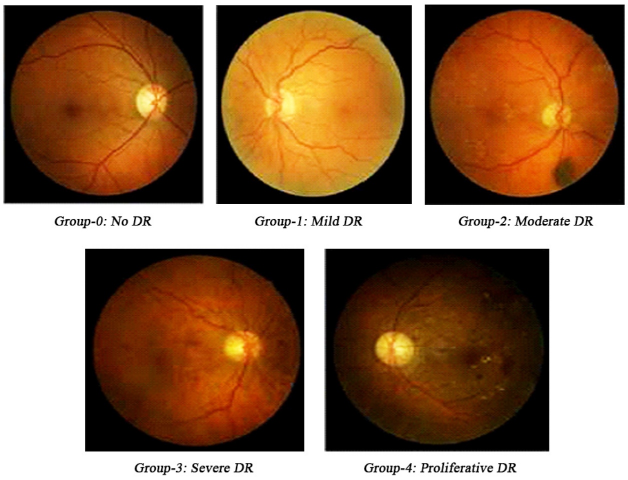


Figure 2. Groups of images in Kaggle dataset

TABLE I. STAGES OF IMAGES IN KAGGLE DATASET

Group Name	Stage of DR	Number of Images
Group-0	Normal / No DR	25810
Group-1	Mild DR	2443
Group-2	Moderate DR	5292
Group-3	Severe DR	873
Group-4	Proliferative DR	708

B. *Non-public MESSIDOR Dataset*

The MESSIDOR dataset is not available publicly and made available based on the request. The dataset possesses a total of 1200 colored retinal fundus images. The color fundus images of a posterior pole taken using the color video 3CCD camera over a 45-degree field of view. The images are taken with 8-bits color plane with 1440 X 960, 2240 X 1488 or 2304 X 1536 pixels respectively. Out of 1200 images, 400 images are acquired without pupil dilation and remaining 800 images are acquired with proper pupil dilation. Our examination just uses the retinopathy grade as a source of perspective, a portrayal of which is furnished in Table II along with the numbers of images for every class. To assess the proposed procedure, leave-one-out cross-validation is performed [13]. At every emphasis, 1199 images are utilized to prepare the model and the single outstanding image is utilized for testing. This progression is rehashed until all the 1200 images have been utilized for testing.

TABLE II. MESSIDOR DATASET - RETINOPATHY GRADES

Grade	Description	N_b Images
R_0 (No DR)	$(N_{MA} = 0)$ AND $(N_{HE} = 0)$	546
R_1 (Mild)	$(0 < N_{MA} \leq 5)$ AND $(N_{HE} = 0)$	153
R_2 (Moderate)	$(5 < N_{MA} < 15)$ AND $(0 < N_{HE} < 5)$ AND $(N_{NV} = 0)$	247
R_3 (Severe)	$(N_{MA} \geq 15)$ OR $(N_{HE} \geq 5)$ OR $(N_{NV} > 0)$	254

where, N_{MA}, N_{HE}, N_{NV}: number of microaneurysms (MA), haemorrhages (HE) and neovessels (NV), respectively.

C. *Preprocessing*

The preprocessing is an essential operation to improve the quality of images, because low quality images will pull down the accuracy by producing inappropriate results. The preprocessed image will be used as an input for the training and for categorizing the images into their 5 groups (group-0 to group-4). The preprocessing also involves data augmentation to increase the number of training data samples and normalization to overcome noise from the image for better classification of fundus images.

To improve the focusing of the model, the instant data augmentation was preferred all over the training. At every single time each image was haphazardly modified with: random rotation 10-80 degrees, random horizontal and vertical flips and random horizontal and vertical shifts.

In the proposed work, the following phases were used: First the modified image is constructed, where the sizes of the target images are decreased to 256 X 256 pixels, to lessen the measure of required PC memory during deep learning process. At the subsequent phase, the retinal fundus images are isolated into three image parts of R (red), G (green), and B (blue). Next, Contrast Limited Adaptive Histogram Equalization (CLAHE) [4] is used for fundus image enhancement. These image enhancements are considered because they are very fast and easy to perform under any method.

The general histogram equalization formula is represented in eqn. 1.

$$h(v) = round\left(\frac{cdf(v) - cdf_{\min}}{(M \times N) - cdf_{\min}} \times (L - 1)\right) \tag{1}$$

where, cdf_{min} is the least non-zero rate of the cumulative distribution function, $M \times N$ gives the number of pixels in the image, v gives the different grey levels and L is the number of grey levels used. The cumulative distributive function (cdf) is a cumulative sum of the considerable number of probabilities lying in its space and characterized in eqn. 2:

$$cdf(x) = \sum_{k=-\infty}^{x} P(k) \tag{2}$$

where, P(k) represents the probability of an occurrence of a pixel of level k in the image.

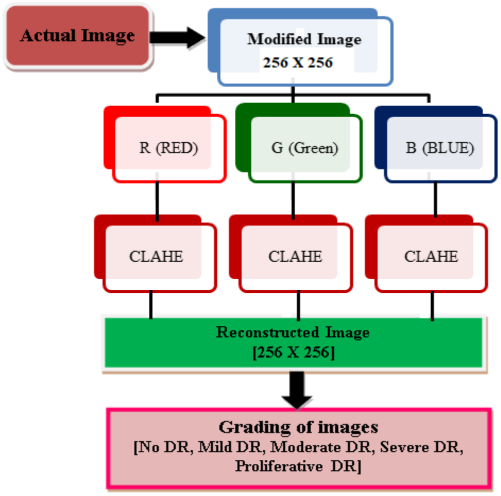


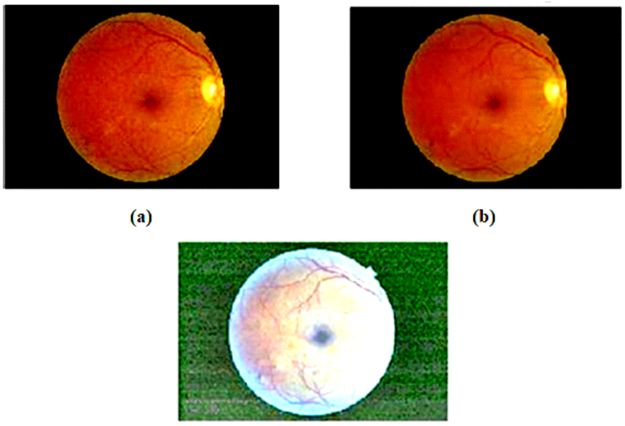
Figure 3. Stages within the flow of the proposed approach

In this work, the following stages were followed as shown in Fig. 3.

Stage 1: To reduce the amount of required memory during the learning process, the actual images are reduced to 256 X 256 pixels, say reconstructed image.

Stage 2: The image is segregated into three image components of R(Red), G(Green), and B(Blue).
Stage 3: To each R, G, and B components, the fundus image enhancement techniques CLAHE were applied.
Stage 4: R, G, and B components are merged to produce the colored image, which will be in turn used by the Convolutional Neural Network.

The sample images for Original image (a), Reconstructed image (b), CLAHE (c) were given in Fig. 4 for the 20051019_38557_0100_PP.tif image file from the MESSIDOR DB1.



(a) (b)

(c)

Figure 4. The sample images from MESSIDOR DB1: a) Original image; b) Reconstructed image; c) CLAHE

IV. Materials and Methods

The Classification of images can be successfully and proficiently done utilizing Convolutional Neural Networks (CNNs) on account of their vast success and their great execution. The compelling thought behind CNNs is that a nearby comprehension of an image is adequate. The useful advantage is that having less parameter incredibly improves the time it takes to learn and also diminishes the measure of data required to prepare the model. Rather than a fully connected network of weights from every pixel, a CNN has quite recently enough loads to take a gander at a little fix of the image. It resembles perusing a book by utilizing an amplifying glass; in the end, you read the entire page, yet you take a gander at just a little fix of the page at some random time.

When, considering a 256 x 256 image, CNN can effectively check it piece by lump e.g., a 5 × 5 window. The 5 × 5 window moves along the image (generally left to right, and start to finish), as demonstrated as follows. The rapidness of its moves is known as its stride length. For instance, a stride length of 2 represents the 5 × 5 sliding window slides by 2 pixels one after another until it traverses the whole image.

A convolution is a weighted aggregate of the pixel estimations of an image, as the window moves over the entire image. Turns out, this convolution procedure all through an image with a weight framework creates another image of a similar size, contingent upon the show. Convolving is the way towards applying a convolution.

The sliding-window illusion occurs in the convolution layer of the neural system. A characteristic CNN has various convolution layers. Each convolutional layer regularly produces many exchange convolutions, so the weight grid is a tensor of 5 × 5 × n, where n is the quantity of convolutions. The significance of the CNN is that the quantity of metrics is free of the size of the actual image. You can run the equivalent CNN on a 300 × 300 image, and the quantity of metrics won't change in the convolution layer.

The latest Deep CNNs such as AlexNet, ZFNet, GoogLeNet, VGGNet-16, ResNet-152 were used to grade the DR images with an addition of transfer learning and hyper-metrics modification. Comparative study for DR diagnosis based on the performance of models is presented.

A. *AlexNet*

AlexNet was designed by the Supervision group, consisting of Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever. The network had a very similar architecture as LeNet by Yann LeCun et. al. [18] but was deeper, with more filters per layer, and with stacked convolution layers. It consisted 3x3, 5x5, 11x11 convolutions, max pooling, dropout, data augmentation, ReLU activations, SGD with momentum. It attached ReLU activations after every convolutional and fully-connected layer. AlexNet was trained for 6 days at the same time on two Nvidia Geforce GTX580 GPUs this is why their network is split into two pipelines.

B. *ZFNet*

ZFNet [18] was mostly an achievement by modifying the hyper-metrics of AlexNet while following the same structure with additional deep learning elements.

C. *GoogLeNet*

The GoogLeNet (a.k.a. Inception V1) from Google [19] used a CNN inspired by LeNet but implemented a novel element which is dubbed an inception module. It used batch normalization, image distortions and RMSprop. This module is based on several very small convolutions in order to drastically reduce the number of parameters. Their architecture consisted of a 22-layer deep CNN but reduced the number of parameters from 60 million (AlexNet) to 4 million. For better understanding, the GoogLeNet architecture's inception model (naïve version) and inception model with dimension reductions.

D. *VGGNet*

VGG16 was developed by Simonyan and Zisserman [21]. VGG16 consists of 16 convolutional layers and is very appealing because of its very uniform architecture. Similar to AlexNet, only 3x3 convolutions, but lots of filters, trained on 4 GPUs for 2–3 weeks. It is currently the most preferred choice in the community for extracting features from images. The weight configuration of the VGG16 is publicly available and has been used in many other applications and challenges as a baseline feature extractor. However, VGG16 consists of 138 million parameters, which can be a bit challenging to handle.

E. *ResNet*

A Residual Neural Network (ResNet-152) by Kaiming He et. al. [22] introduced a novel architecture with residual connections and features large batch normalization. Such skip connections are also known as gated units or gated recurrent units and have a strong similarity to recent successful elements applied in Recurrent Neural Networks (RNNs). Using RNNs technique, they were able to train a neural network with 152 layers while still having lower complexity than VGGNet.

F. *Proposed CNN Architecture*

The Proposed CNN architecture is shown in Fig. 5. The factors of the proposed structure have been adjusted with the usage of the pre-trained weights and biases of the real ResNet-152 structure, which were formerly trained and tested successfully. Furthermore, the preliminary factors of the extra proposed convolutional and last fully connected layers have been assigned to the identity. The entire architecture was further trained and validated for DR classification, and categorization, in which the factors of the complete version have been fine-tuned and updated at the Kaggle dataset and the MESSIDOR dataset.

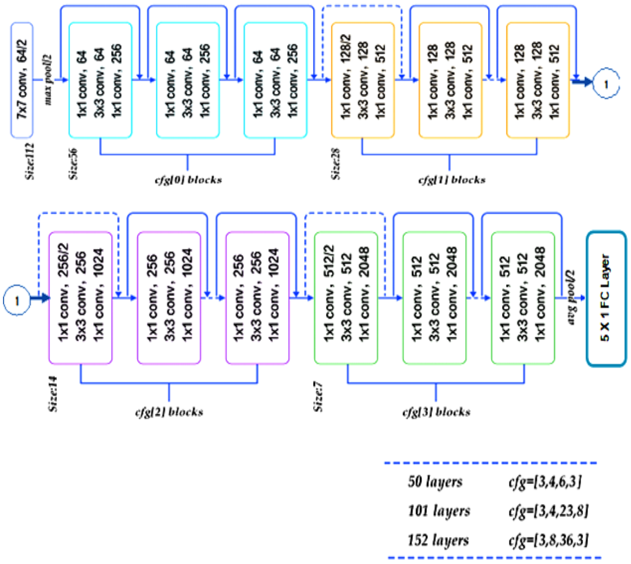


Figure 5. The Proposed CNN Architecture

The ResNet structure is tight that may substantially boost up the training of very-deep neural networks and enhance the accuracy of the model. The ResNet belief stems from what the intensity of CNN will increase, a degradation downside arises. This could be now no longer depend on over-fitting, as an end result of the error will increase not solely inside the test examples however additionally in the training examples itself. If a fairly shallow network that meets the accuracy of saturation and contains many of the congruent mapping layers, then to the minimum, the error won't upload up, that is, the deeper the network shouldn't result in additional training examples error. The concept to skip the preceding output on to consecutive layer by exploitation congruent mapping evokes ResNet.

Assuming that the input to an explicit CNN is x and consequently the expected output is H(x), then our learning intention is F (x) = H (x) - x once directly switch the input x to the output as the preliminary result, represented in Fig. 6.

In transfer learning method, the former fully connected layer of preceding trained CNNs is removed and observed as a feature extorter. We train a classifier on the new dataset until all features of the all retinal fundus images were successfully extracted. The network itself will not initialize the considerations of hyper-metrics modification method. So, it is essential to adjust and enhance these parameters conferring to the results of training the retinal fundus image in modifying the performance. Transfer learning established techniques have been done with pre-trained AlexNet and GoogLeNet architectures from ImageNet. The final fully connected layer became removed, and then a transfer learning scenario [16] was accompanied by handling the remaining network additives as a set characteristic extractor for the fresh dataset. The transfer learning keeps preliminary pre-trained version weights and extracts image features through a final network layer.

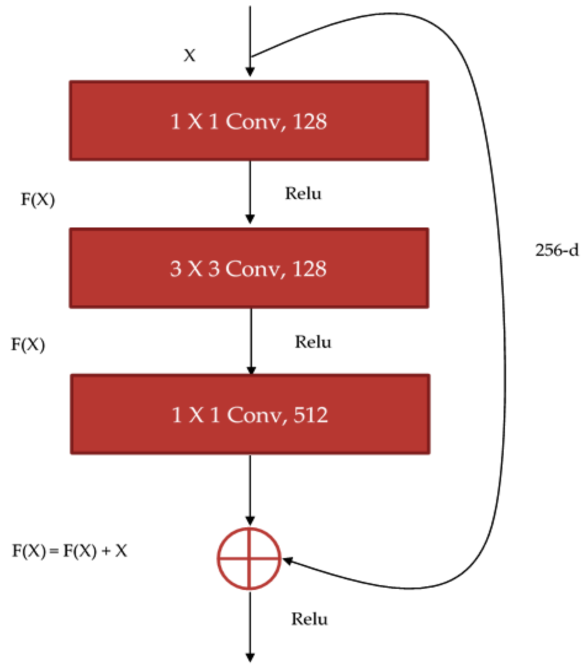


Figure 6. The proposed bottleneck design

The methodology of the proposed DR detection using CNNs can be shown in Fig. 7.

V. RESULTS

In this proposed work, MATLAB r2020a and its related tool were utilized for all image processing and classification stages. A PC framework with Intel Core i7-8700K processor, 3.7GHz CPU, 16 GB RAM, running 64 bit windows-10 Operating system was utilized.

A. *Performance Metrics*

The metrics used to assess the performance of classification algorithms are Sensitivity, Specificity, Accuracy, Precision and G_{mean} which is given as follows:

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{3}$$

$$Specificity = \frac{TN}{(TP + FP)} \tag{4}$$

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

$$Precision = \frac{TP}{(TP + FP)} \tag{6}$$

$$G_{mean} = \sqrt{Sensitivity \times Specificity} \tag{7}$$

In Eqns. 3 - 7, TP denotes the number of True Positive diagnoses, FP represents the number of False Positive diagnoses, TN refers to the number of True Negative diagnoses, and FN represents the number of False Negative diagnoses.

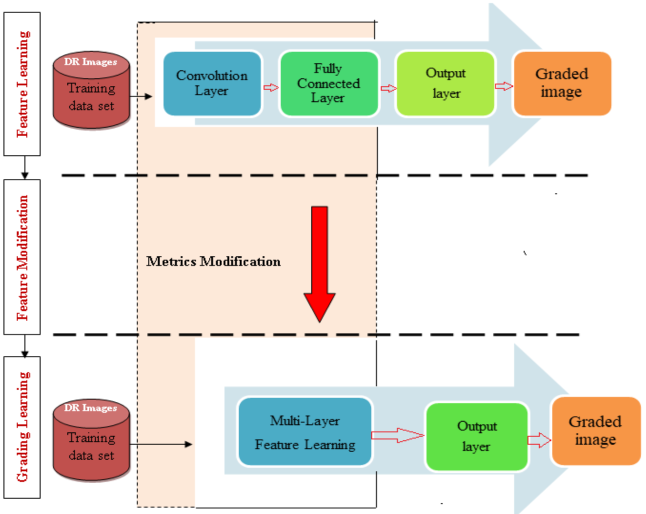


Figure 7. The proposed method using CNNs

B. *Evaluation Results with Kaggle Dataset*

The Table III depicts the classification results of CNN models with randomly initialized parameters. Different CNNs models have different classification performance and the overall classification performances are not satisfactory. To overcome the over-fitting phenomenon existed during the training process; we prefer using transfer learning and hyper-metrics modification techniques to grade the retinal fundus images more accurately.

TABLE III. CLASSIFICATION RESULTS OF DIFFERENT CNN MODELS [KAGGLE DATASET]

Metrics	Alex Net	ZF Net	VGG Net	Goog LeNet	Res Net 152	Proposed CNN model
Sensitivity (%)	84.48	89.88	91.27	92.89	94.8	**94.35**
Specificity (%)	79.21	63.43	72.15	69.38	81.57	**85.77**
Accuracy (%)	83.4	78.77	85.71	80.53	86.12	**88.68**
Precision (%)	81.5	88.56	90.21	92.78	93.21	**94.00**
G_{mean}	0.87	0.92	0.95	0.91	0.92	**0.936**

The following experimental settings were used for transfer learning: the retinal fundus images data was amplified to 30 times of the original, with 50 training iterations, the learning degree is direct difference between [0.001-0.1], and to update the weigh values, the stochastic gradient descent with momentum optimizer is utilized.

The Stochastic Gradient Descent optimizer (SGD) lets in for a quicker training method on the grounds that at every epochs, it simplest ruminates a subclass of the training set.

TABLE IV. CLASSIFICATION RESULTS WITH HYPERMETRICS-MODIFICATION OF CNN MODELS [KAGGLE DATASET]

Metrics	Alex Net	ZF Net	VGG Net	Goog LeNet	Res Net 152	Proposed CNN model
Sensitivity (%)	94.18	95.34	93.23	92.54	93.24	**95.65**
Specificity (%)	82.72	85.74	89.87	93.23	91.75	**94.86**
Accuracy (%)	87.57	91.42	92.55	95.63	96.22	**97.72**
Precision (%)	83.25	90.12	91.56	92.45	92.10	**92.88**
G_{mean}	0.89	0.93	0.94	0.92	0.93	**0.95**

SGD with Momentum (SGDM) is used rather than SGD, via way of means of which includes the momentum and as a result figuring out the following replace of the weights primarily based totally on a linear mixture of the gradient and the preceding replace, it assessments the education method to reveal dithery behavior. This will bring about quicker and correct convergence.

The k-fold cross-validation (where k = 8) is performed to compute the results. The modification of optimal hyper-metrics is performed utilizing the gradient based enhancement which computes the gradient as for hyper-metrics and upgrade them utilizing the algorithm for gradient descent calculation. The accuracy of proposed CNN model classification in the experiment is 97.72% as in Table IV. The misclassification rate is commendably low because of application of hyper parameter training and transfer learning. Few misclassifications occurs due to the low quality of the images and images with more noises tend to miss classify which is reduced to certain extent by our proposed CNN method.

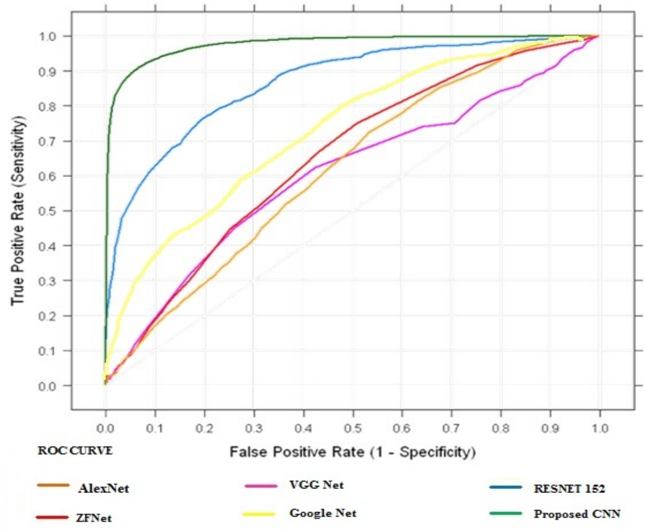


Figure 8. ROC of different CNN models [Kaggle Dataset]

The other CNN models have comparatively less classification accuracy, due to their large complex structure and a greater number of training parameters than the proposed CNN model. The Table V depicts the confusion matrix for retinopathy grade (using 5-fold cross-validation of the proposed CNN model) between the reference standard and the proposed system. The Fig. 8 represents Receiver Operating Characteristic (ROC) of different CNN models**.**

TABLE V. THE CONFUSION MATRIX FOR CLASSIFICATION (USING 5-FOLD CROSS-VALIDATION OF THE PROPOSED CNN MODEL) BETWEEN THE REFERENCE STANDARD AND THE PROPOSED CNN METHOD [KAGGLE DATASET]

	Proposed method					Total
	Group-0	Group-1	Group-2	Group-3	Group4	
Group-0	25489	103	124	52	42	25810

Group-1	55	2358	11	8	11	2443
Group-2	138	52	5060	19	23	5292
Group-3	35	10	5	817	6	873
Group-4	30	6	30	41	601	708
Total	25747	2529	5230	937	683	35126

The proposed model had a classification efficiency of 97.72% when compared to other CNN algorithms.

C. Evaluation Results with MESSIDOR Dataset

In the classification process, images from MESSIDOR DB1 data were selected and augmented to 12,000 images as discussed in the preprocessing section. Now, 11,600 images were chosen for learning, while the other residual 400 images were selected for testing and evaluation. To oversee the aids of the image processing techniques, classification process were performed at every phase before thorough diagnosis. At each phase a total of thirty independent runs were done. Exactly, a hyper-metrics modification technique was followed by using stochastic gradient descent with momentum at the learning degree which is direct difference between [0.001-0.1] and a restriction of 90 iterations maximum. Averagely after 1 hour and 10 minutes, each independent run was completed at the maximum iteration rate. Table VI briefly provides the classification results of CNNs model.

TABLE VI. CLASSIFICATION RESULTS OF DIFFERENT CNNS MODEL [MESSIDOR DATASET]

Metrics	Alex Net	ZF Net	VGG Net	Goog LeNet	Res Net 152	Proposed CNN model
Sensitivity (%)	86.87	85.72	89.08	91.44	92.43	**96.23**
Specificity (%)	89.65	90.27	93.96	95.56	96.10	**96.47**
Accuracy (%)	93.07	93.19	93.74	94.46	96.54	**97.65**
Precision (%)	91.44	92.28	93.03	94.39	94.88	**95.35**
G_{mean}	0.87	0.91	0.92	0.94	0.94	**0.947**

The following experimental settings were used for Transfer learning: the retinal fundus images data was amplified to 30 times of the original, with 50 training iterations, the learning degree is direct difference between [0.001-0.1], and to update the weigh values the stochastic gradient descent with momentum optimizer is utilized.

TABLE VII. CLASSIFICATION RESULTS WITH HYPERMETRICS-MODIFICATION OF CNNS MODEL [MESSIDOR DATASET]

Metrics	Alex Net	ZF Net	VGG Net	Goog LeNet	Res Net 152	Proposed CNN model
Sensitivity (%)	89.38	90.57	91.75	94.35	95.23	**96.73**
Specificity (%)	90.47	92.42	95.27	96.05	96.98	**97.47**
Accuracy (%)	94.43	95.08	95.89	96.62	97.05	**97.58**
Precision (%)	92.42	93.31	94.15	95.62	95.33	**97.45**
G_{mean}	0.91	0.92	0.93	0.96	0.96	**0.978**

Eight times the cross-validation is to compute the results. The modification of optimal hyper-metrics is performed utilizing the gradient based enhancement which computes the gradient as for hyper-metrics and upgrade them utilizing the algorithm for gradient descent calculation. The accuracy of proposed CNN model classification in the experiment is 97.58% and the other existing CNN models classification results were listed in Table VII.

VI. DISCUSSION

The existing CNN models have comparatively less classification accuracy, due to their huge and additional vague structure and a greater number of training parameters than proposed CNN model. The Fig. 9 represents Receiver Operating Characteristic (ROC) of different CNN models.

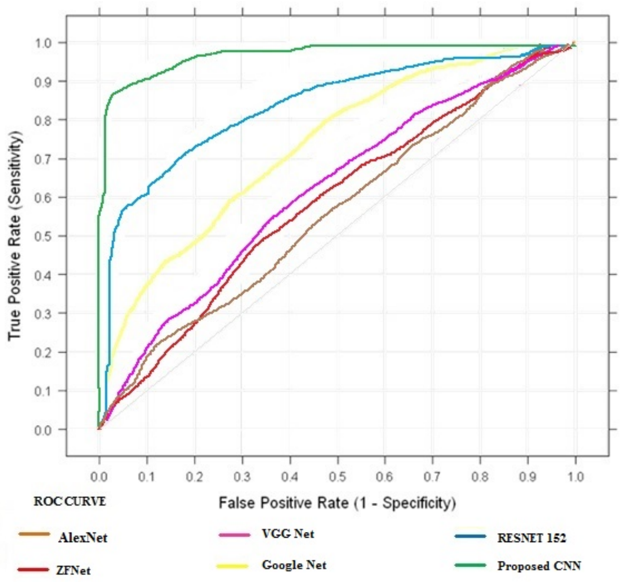


Figure 9. ROC of different CNN models [MESSIDOR Dataset]

TABLE VIII. THE CONFUSION MATRIX FOR CLASSIFICATION (USING 5-FOLD CROSS-VALIDATION OF THE PROPOSED CNN MODEL) BETWEEN THE REFERENCE STANDARD AND THE PROPOSED METHOD [MESSIDOR DATASET]

	Proposed method				Total
	R0	**R1**	**R2**	**R3**	
R0	5390	20	30	20	5460
R1	30	1460	20	20	1530
R2	50	20	2390	10	2470
R3	30	20	20	2470	2540
Total	5500	1520	2460	2520	12000

From the Table VII, the following were the outcomes of results of the proposed method: the classification of diabetic retinopathy disease with transfer learning using deep convolutional neural network models in which the proposed CNN model provides a classification accuracy of 97.58%. Furthermore, the proposed technique had sensitivity of 96.73% and specificity of 97.47%. From these results that the proposed technique provides better results than other methods discussed in the literature. Also, all results show that use of CNN models (indirectly deep learning) had improved results in terms of diagnosis and classification. The Table VIII depicts the confusion matrix for retinopathy classification (using 5-fold cross-validation of the proposed CNN model) between the reference standard and the proposed system.

The proposed model had a classification efficiency of 97.58% when compared to other CNN algorithms. This can be because of the reality that the alternative architectures have large and greater complicate structure and greater training parameters than proposed CNN model. More tracing parameters and much less training information might produce over-fitting phenomenon, which may also yield the much less erroneous classification performance.

VII. Conclusion

The proposed CNN model for the Kaggle dataset provides accuracy of 97.72%, sensitivity of 95.65%, and specificity of 94.86% in the classification of diabetic retinopathy disease with transfer learning using deep convolutional neural network models. For the MESSIDOR dataset, the proposed CNN model has a classification accuracy of 97.58 %, sensitivity of 96.73 %, and specificity of 97.47 %. From these results, the proposed technique provides better results than other state-of-art methods. Also, all results show that use of CNN models had improved results in terms of diagnosis and classification. The results from our CNN models result are impressive from a traditional network topology. In contrasting with the existing methods, no specific features such as optic disc, blood vessels, exudates etc. of retinal fundus images have been considered.

In future, the more understated classification features such as optic disc, blood vessels, exudates etc. can also be considered for classification the retinal fundus image. Furthermore, attention will be given to practice of these techniques in eye diagnosis centres for diagnosing different eye diseases which can be very useful for ophthalmologists for their general diagnosis processes. We can also apply our proposed model as a mobile phone application, to make DR detection simpler and time-saving.

Acknowledgment

The authors would like to thank their friends and colleagues for their constant help and support throughout the study and to obtain the results.

References

[1] I. Monzurul, A. V. Dinh, and K. A. Wahid, “Automated diabetic retinopathy detection using bag of words approach,” Journal of Biomedical Science and Engineering, pp. 86-96, 2017. doi:10.4236/jbise.2017.105B010

[2] S. D. Wahyu, and D. Juniati, “Classification of diabetic retinopathy using fractal dimension analysis of eye fundus image,” In AIP conference proceedings, August 2017. doi:10.1063/1.4994414

[3] D. Omer, U. Kose, and G. E. Guraksin, “Underwater image enhancement with HSV and histogram equalization,” 7th International Conference on Advanced Technologies 2018, pp. 1 – 6. E-ISBN: 978-605-68537-1-5

[4] B. Ritesh, P. Sahu, “Performance analysis of underwater image enhancement with CLAHE 2D median filtering technique on the basis of SNR, RMS error, mean brightness,” International Journal of Engineering and Innovative Technology. pp. 525-528, 2013

[5] N. Ramachandran, S. C. Hong, M. J. Sime, G. A. Wilson, “Diabetic retinopathy screening using deep neural network,” Clinical & experimental ophthalmology, pp. 412-416, 2018. doi:10.1111/ceo.13056

[6] G.-M. Lin, M.-J. Chen, C.-H. Yeh, Y.-Y. Lin, H.-Y. Kuo, M.-H. Lin, M.-C. Chen, D. L. Shinfeng, Y. Gao, A. Ran, Y. C. Cheung, “Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy,” Journal of Ophthalmology, 2018. doi:10.1155/2018/2159702

[7] A. B. Aujih, L. I. Izhar, F. Meriaudeau, M. I. Shapiai, “Analysis of retinal vessel segmentation with deep learning and its effect on diabetic retinopathy classification,” International conference on intelligent and advanced system, 2018 pp. 1-6. doi:10.1109/ICIAS.2018.8540642

[8] K. S. Sreejini, V. K. Govindan, “Severity classification of DME from retina images: a combination of PSO and FCM with bayes classifier,” International Journal of Computer Applications, pp.11-17, 2013. doi:10.5120/14206-2430

[9] U. R. Acharya, E. Y. K. Ng, J. H. Tan, S. V. Sree, K. H. Ng, “An integrated index for the identification of diabetic retinopathy stages using texture parameters,” Journal of medical systems, pp. 2011-2020, 2012. doi:10.1007/s10916-011-9663-8

[10] N. Becherer, J. Pecarina, S. Nykl, K. Hopkinson, “Improving optimization of convolutional neural networks through parameter fine-tuning,” Neural Computing and Applications, pp.3469–3479, 2019. doi:10.1007/s00521-017-3285-0

[11] R. J. Borgli, H. Kvale Stensland, M. A. Riegler, P. Halvorsen, “Automatic hyperparameter optimization for transfer learning on medical image datasets using bayesian optimization,” 13th International Symposium on Medical Information and Communication Technology, 2019, pp. 1-6. doi:10.1109/ismict.2019.8743779

[12] Y. Wang, G. A. Wang, W. Fan, J. Li, “A deep learning based pipeline for image classification of diabetic retinopathy,” International Conference on Smart Health, 2018, pp. 240– 248. doi:10.1007/978-3-030-03649-2_24

[13] F. Ren, P. Cao, D. Zhao, C. Wan, “Diabetic macular edema classification in retinal images using vector quantization and semi-supervised learning,” Technology and Health Care, pp. 389-397, 2018. doi:10.3233/THC-174704

[14] M. B. Rodrigues, R. V. M. Da No´brega, S. S. A. Alves, P. P. Reboucas Filho, J. B. F. Duarte, A. K. Sangaiah, V. H. C. De Albuquerque, “Health of things algorithms for malignancy level classification of lung nodules,” IEEE Access, pp. 18592-18601, 2018. doi:10.1109/ACCESS.2018.2817614

[15] M. Abdullah, M. M. Fraz, S. A. Barman, “Localization and segmentation of optic disc in retinal images using circular Hough transform and grow-cut algorithm,” PeerJ, 2016. doi:10.7717/peerj.2003

[16] M. Oquab, L. Bottou, I. Laptev, J. Sivic, “Learning and transferring mid-level image representations using convolutional neural networks,” Proc. of IEEE conference on computer vision and pattern recognition. 2014, pp. 1717–1724. doi:10.1109/CVPR.2014.222

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Communications of the ACM, pp. 84-90, 2012. doi:10.1145/3065386

[18] L. Y. Cun, L. Bottou, Y. Bengio, P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, 1998 pp. 2278-2324. doi:10.1109/5.726791

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, “Going deeper with convolutions,” Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9, 2015. doi:10.1109/cvpr.2015.7298594

[20] W. Wang, Y. Yang, X. Wang, W. Wang, J. Li, “Development of Convolutional Neural Network and its application in Image classification: A Survey,” Optical Engineering, pp. 1-19 2019. doi:10.1117/1.OE.58.4.040901

[21] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv, 2015. https://arxiv.org/abs/1409.1556v6

[22] K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition, ” Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 171-180. doi:10.1109/cvpr.2016.90

[23] Asia Pacific Tele-Ophthalmology Society, APTOS 2019 blindness detection Database: Kaggle [Internet]. Available from: https://www.kaggle.com/c/aptos2019-blindness-detection/data

[24] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, B. Charton, “Feedback on a publicly distributed image database: the Messidor database”, Image Analysis & Stereology,” pp. 231-234, 2014. doi:10.5566/ias.1155

[25] W. N. J. H. W. Yussof, M. S. Hitam, E. A. Awalludin , Z. Bachok, “Performing contrast limited adaptive histogram equalization technique on combined color models for underwater image enhancement,” International Journal of Interactive Digital Media, vol. 1, no. 1, pp. 1-6, 2013. doi:10.1016/b978-0-12-336156-1.50061-6