Diabetic Retinopathy Detection and Grading using Deep learning

Mohamed A. Berbar
Dept. of Computer Engineering and sciences
Faculty of Electronic Engineering,
Menofia University
Menouf 32952, Egypt
m_berbar@el-eng.menofia.edu.eg
0000-0001-5405-159X

***Abstract*— One of the complications of diabetes disease is diabetic retinopathy (DR). Diabetic patients may suffer from total loss of sight. That's if it is not detected and medicated early enough. The early detection of DR is very important during funds screening on a regular basis. Detection and grading of DR are difficult because most fundus images suffer from undersaturation and noise. This paper proposes a new enhancement process as a solution to the poor quality of fundus images. It also proposes two architectures for convolutional neural network (CNN) models. The first one is the binary classifier of DR images into normal and abnormal. The second CNN architecture to classify the severity grades of DR. In this study, we also utilized different pre-trained convolutional neural network models to show the impact on the performance of the use of transfer learning from pre-trained CNN models vs newly defined architectures. The pre-trained CNN models and the two new proposed CNN models are tested using *Messidor*1, *Messidor*2, and Kaggle EyePACS datasets. The proposed binary classifier model results in F1-scores of 0.9387, 0.9629, and 0.9430 on the Messidor-1, Messidor-2, and EyePACS datasets, respectively. The proposed second model classifies the five grades with an F1-score of 0.9133, 0.9226, and 0.9393 on the Messidor1, Messidor2, and Kaggle EyePACS datasets, respectively. The new proposed CNN model proved its reliability and efficiency in detecting DR and classifying severity grades of DR in fundus images. Preprocessing techniques enhanced the performance by 10.83% of accuracy and 0.13037 in AUC using the binary model.**

Keywords*— *Diabetic Retinopathy; Convolutional Neural Network; Fundus images; Deep learning.

I. INTRODUCTION

Diabetes is a chronic disease caused by the low level of insulin production in humans. Insulin is the hormone responsible for maintaining the sugar level in the blood at its normal levels. The increase in the glucose concentration in the blood leads to the damage of retinal vessels [1]. About 8.3% of the population in the USA, including children and adults, has diabetes [1], according to statistics from the American Diabetes Association. In Egypt—Alexandra, a study has been conducted on over 506 diabetes patients and diabetic retinopathy was detected in 34.6% of the studied patients [2]. Unfortunately, this number will keep increasing because of the way we live our modern lives. Research by [3] has been conducted in 91 countries around the world and concludes that the diabetes rate will increase to a rate of 69% of adults in developing countries and 20% of adults in developed countries in the next few years. Researchers in [4] reported that there is an 80% chance of contracting DR in a person with more than 10 years of diabetes. Diabetic retinopathy causes severe vessel damage, which nourishes the retina mainly by transcytosis of nutrients. Diabetic Retinopathy (DR) is named the "silent disease" because, without the full consciousness of the patient, the disease may damage the retinal vessels and lead to severe eye damage and blindness for diabetic people. Unfortunately, at the beginning of the DR disease, diabetes will not feel any symptoms until serious damage happens to the retina, like vision loss. Blurring, shadows, or loss of vision areas, difficulty seeing at night, blood vessels swelling in the eye, vessels may leak fluid. The high level of sugar results in leakage of fluids such as proteins from blood vessels in the retina, which causes soft and then hard exudates on the surface of the retina. This leakage can be in the form of different lesions like hard exudates, microaneurysms, hemorrhages, etc. [5]. Figure 1 and Figure 2 show the difference between a healthy retina and a DR-affected retina.

Figure 1: Messidor2 sample: [20] (a) Normal retinal image (b) diabetic retinopathy retinal image.

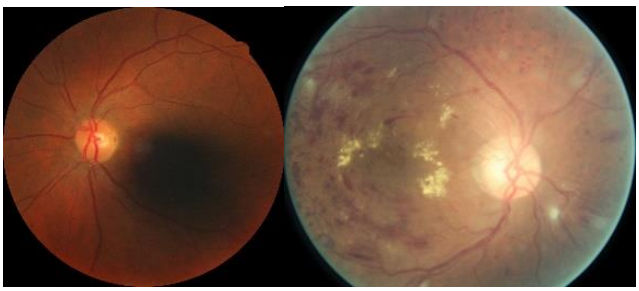

Figure 2: Kaggle EyePACS Dataset sample: [22] (a) Normal retinal image (b) diabetic retinopathy retinal image.

The earliest lesion that can be detected at the start of the DR disease is a microaneurysm (MA). The high sugar level in the blood causes permeability of retinal blood vessels, resulting in the formation of MAs [7]. MA is small and looks like dark red spots and is often temporal to the macula. MA usually has sharp margins that range between 20 μm and 200 μm in diameter [6]. Increases in the number of MAs will lead to retinal ischemia and DR progression. These MAs may herniate, causing hemorrhages [8]. Hemorrhages (HEM) are the second lesion of DR. HEM is primarily caused by the leakage of weak vessels. HEM is also characterized by the form of a red spot with varying density and non-uniform edges. It is larger than 125 m [6]. Hard exudates (HE) is the third lesion of DR that are shiny, yellowish, and irregular in shape. HE is caused by the leakage of proteins out of the retinal vessels [9]. Figure 3 shows an example of these lesions in the human retina images.

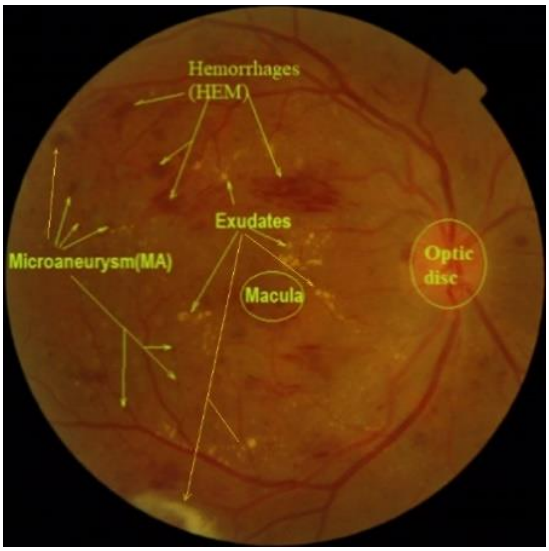

Figure 3: Retina of the eye showing microaneurysm, hemorrhages, and hard exudates.

Hemorrhages and microaneurysms are the first clinically detected lesions indicating the presence of diabetic retinopathy. Most DR detection systems apply the first stage to extract red lesion candidates, then a second stage to decide whether the detected lesions are HEM or MAs. Diabetic retinopathy is classified according to DR severity level into five grades, each labelled with an integer number ranging between zero and four. Many research efforts have been made to detect the early symptoms of this harmful disease using image processing techniques because the early diagnosis of diabetic retinopathy can reduce complications of the disease in 90% of cases. These techniques include the use of image preprocessing techniques such as morphology, enhancement filters, edge detection, histogram manipulation, and the use of neural networks, support vector machines as classification techniques, etc. Deep learning methods such as Convolutional Neural Network (CNN) strongly appear to be the best method for the automated classification of digital medical images [10–13]. DR could be diagnosed directly by detecting abnormalities using CNN. CNNs have become the most famous standard in solving image recognition tasks, especially after the improvement of their performance with supportive tools like activation functions like Rectified Linear Units [16], Dropout [17], regularization, Batch Normalization [18], etc. In this research paper, we propose a new convolutional neural network model to detect diabetic retinopathy directly after applying preprocessing operations to enhance the quality of the fundus image. Investigating the RGB channels by us and by other researchers proved that the blue channel of the fundus image has the lowest contrast and suffers from undersaturation and noise, while the red channel suffers from oversaturation [29]. Furthermore, the green channel has the best contrast [2, 29, 14]. So, the preprocessing stage is essential for DR detection. Histogram matching is utilized on the red and blue images followed by a median filter, and the contrast limited adaptive histogram equalization (CLAHE) method is subsequently utilized on all the RGB channels of the image. At the beginning of the research, we focused on binary classification as normal and abnormal images. We moved to DR grading after the successful detection of abnormalities in images. The proposed system provides a highly accurate diagnosis and grading of DR, a significant improvement over human error. The proposed method allows for both the diagnosis and quantification of the severity of DR. The two proposed models have been trained and tested by standard datasets such as Messidor-1, Messidor-2, and Kaggle EyePACS after applying an enhancement preprocessing stage. The tested resolutions of the training images we have taken are 224×224, 300×300, 380× 380, and 512×512 pixels. The DR grade of a fund image is classified as positive if it belongs to DR Grades 1, 2, 3, or 4, and negative otherwise. During the research, we tested multiple models that have been developed via data-driven approaches to machine learning. We have employed state-of-the-art pre-trained CNN models such as Alex-Net, Google-Net, VG19-Net, and Res50-Net as a direct classifier one time and as a feature extraction method another time. First, as a classifier, we modified the pre-trained CNN model to accept image sizes determined by the user and trained it over publicly available datasets. Secondly, we used the state-pre-trained CNN models to extract features by using the weight values of the last fully connected layer of the pre-trained CNN models as features to be classified using the Support Vector Machine (SVM). We have used accuracy, recall (sensitivity), precision, and F1-score to measure the performance of each model. The confusion matrix is used for performance evaluation (see Figure 4).

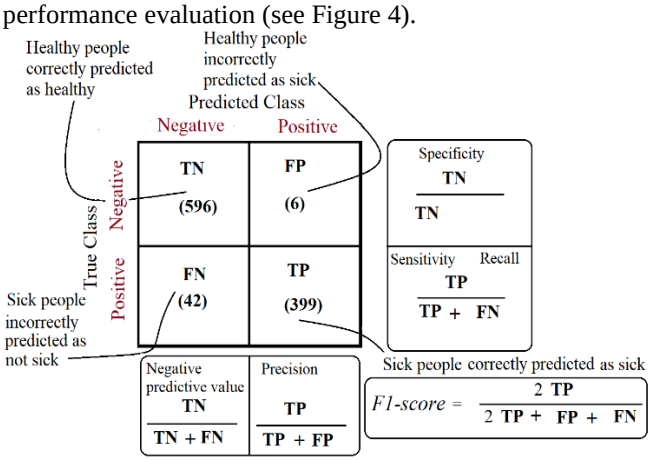

Figure 4: Confusion Matrix

The confusion matrix gives a clear measure of the system, and it does not mislead by only the total accuracy measure, especially when the number of classes is unbalanced.

Precision would determine whether our classifier is reliable or not. Recall (Sensitivity) is the most important evaluation metric in medical diagnosis applications. It tells us how many of the actual DR patients we were able to predict correctly with our classifier. The F1-score captures both sensitivity and precision in a single value and checks the method's performance.

Two proposed models have been proven on fundus images of Messidor-2 (Decencire et al., 2014) [20] and its older version, Messidor-1, and EyePACS datasets for training and validation. Messidor-2 contains 874 examinations (1748 images). Four images were excluded because we didn't have their classification. The study's experiments utilized 1745 fundus images, 1012 of which had no DR lesions and 733 of which had DR lesions graded into four grades according to the severity. The Messidor-2 database was created by a shading video camera. The fundus images are at 2304×1536 ×1488 and 1440×960, 2240pixels. Because of the different image sizes, we resized all images in the database to a fixed size (224 ×224) pixel. The EyePACS database from Kaggle is retinal images provided by EyePACS [37]. It is also used for the evaluation of the proposed system. Kaggle EyePACS fundus images are taken under a variety of imaging conditions. The severity of diabetic retinopathy is graded according to a scale of 0 to 4. Most of the fundus images have noise and contain artifacts. Some are out of focus, overexposed, or underexposed. The images were collected from multiple clinics using a variety of cameras over an extended period. About 3191 images have been selected to be used for the evaluation of our proposed system. Contrast and brightness correction are applied to standardize the brightness and contrast of the image before processing and to highlight its features. During the training of models, a random-split approach was followed instead of k-fold cross-validation. The resolution of the training images we have taken is 224 × 224 pixels. The main findings of this study are (a) insights into using the histogram matching technique in preprocessing steps. (b) testing pre-trained CNN models on funds images using two methods as a classifier and a feature extraction method (c) the efficacy of our proposed models in detecting or grading DR. The situation of our results is superior to leading studies carried out on the same datasets. The developed code for preprocessing, feature extraction, and classification was written using Matlab-19.

II. Related Research Work

Navarro et al. (2016) [19] proposed MA detection methods. In the first phase, they used a polynomial contrast enhancement approach on the green channel of the fundus image. In the second phase, the authors localize MA by applying the Tophet transform technique, and then the Global thresholding model was applied to perform MA segmentation. For the classification of Gaussian data, kernel density-based classifiers and the K Nearest Neighbor Classifier (k-NN) are used. Kumar et al. (2018) [20] applied a preprocessing technique to the green channel, histogram equalization, and morphological processing. They used [43], CLAHE), Principal Component Analysis (PCA), morphological processing, and averaging filtering for MA detection and SVM for classification. They achieved a sensitivity and specificity rate of the DR detection system of around 96% and 92%, respectively, using the DIARETDB1 database, consisting of only 89 color fundus images. Unfortunately, we can't rely on their results because of their small dataset. Hashim et al. [21] are operating region by region to identify the abnormal region in the entire image. Image enhancement and then Gray Level Co-occurrence Matrix (GLCM) are applied to the selected region to perform lesion extraction. In the final step, all extracted lesions were classified by the SVM classifier, and they achieved 82.39% sensitivity. We have reviewed methods to detect and classify lesions in fundus images based on CNN's deep learning. Shaban et al. proposed in [10] a deep convolutional neural network (CNN) to automatically classify DR severity grades into (1-no DR), (2-moderate DR, which is a combination of two grades (mild and moderate DR)), and (3-severe DR, which is a group of severe NPDR and proliferative DR (PDR)) with a validation accuracy of 88%–89%, sensitivity of 87%Fundus images used in this study are publicly available from EyePACS [22]. Abr`amoff et al. [23] used multiple CNNs for the automatic detection of DR lesions in graded fundus images. The system outputs three classes: grade 1 negative (no DR or mild DR), grade 2: referral DR is present, and grade 3: vision-threatening DR is present (vtDR). Abr`amoff et al. obtained results of 96.8% sensitivity and 87% specificity for the detection of referral DR on the Messidor-2 dataset. Gargeya and Leng [24] used a deep CNN model based on deep residual learning. The output of the model was a binary classification (normal or abnormal for DR of any severity level). Preprocessing steps like rotational invariance, and brightness and contrast adjustment were applied to the training set. The dataset Messidor-2 was used for evaluation with a 5-fold cross-validation during training. The model achieved 93% sensitivity and 87% specificity. A random forest-based classifier was presented in [27] by Garima Gupta et al. The classifier achieved 82% sensitivity with 10-fold cross-validations on a small number of images. They used 191 images obtained from 58 diabetic subjects having different degrees of pathological severity. Lam et al. (2018) [31] used pre-trained Google-Net and Alex-Net models to classify DR stages. The Messidor-1 database and the Kaggle EyePACS database were used for training and testing. Their main problem primarily occurs in the misclassification of mild disease as normal. In response, contrast limited adaptive histogram equalization is used to improve their results' accuracies to 74.5% and 68.8% on the Google-Net and Alex-Net models, respectively. Yung-hui et al. [32] used an algorithm based on CNN to extract fundus image features. Two CNNs with a different number of layers are trained to extract features. They replaced the max-pooling layers with fractional max-pooling. The metadata of the image is combined with the extracted features from the two CNNs. An SVM classifier is used for the classification of the fundus images of the EyePACS dataset into DR grades. Their accuracy was 86.17%.

III. Methodology

We aim to classify DR grades in fundus images. At the beginning of the research, we used pre-trained CNN to be

trained on fundus images of the Messidor-2 database. Secondly, we used the pre-trained CNN as a feature extractor and then classified the extracted features using SVM. Finally, we designed our proposed CNN model and used it to solve the problem as a binary classifier (0: no DR, 1: DR is present). Then, the proposed CNN model is modified a little to be used to solve the problem as a five-class classifier for fundus images. The model is used to classify fundus images into (0: no DR, 1: mild DR, 2: moderate, 3: severe DR (NPDR), and 4: PDR). The proposed CNN model architecture comprises an image input layer, followed by three blocks of convolutional, normalization, activation, and maximum pooling layers. The purpose of the pooling layer is summarized by the number of parameters (i.e., weights and biases) by sliding a two-dimensional filter over each channel and reducing the features lying within the region covered by the filter. The fully connected layers are responsible for the classification and usually consist of a set of neurons that are connected with all the activation maps of the neurons of the previous layers. The outputs of convolutional layers are processed using a Rectified Linear Unit (ReLU). Therefore, the number of filters, filter size, and stride are set in advance. Therefore, they do not require training. We treated RGB images as greyscale images represented as two-dimensional arrays, where every value represents pixel intensity. While training the model, we fed the images to CNN in batches. A series of successive transformations have been applied to the input images by the layers of the model. The model produces predictions based on transformations that use the pixel values of the images and the current weights of the model. Then the loss function (binary cross-entropy) compares the predictions with true labels and generates a loss score. This loss score is utilized by the Adam (adaptive moment estimation) optimizer to update the weights (back-propagation) of layers of the model using the training options, including initial learning rate equal to 0.001, L2 regularization factor equal to 0.1, and mini-batch size equal to 128. to update the model parameters that will allow the successful classification of images.

A. Preprocessing

The images within the datasets, as explained in the introduction section, have significant variability in color, illumination, resolution, and quality. Many researchers excluded some bad quality images from processing, and others proposed a CNN model to differentiate between good and bad quality fundus images [11]. For this reason, noise removal and contrast and brightness correction are applied to standardize the brightness and contrast of the image and to highlight features before classification. We apply histogram matching on the red and blue channels to selected and fixed reference images. A reference fundus image should have balanced brightness and color, and thus, careful selection is necessary. The output is an image with a brightness that is comparable to that of the reference image. A median filter and the CLAHE method are subsequently utilized on the three image channels, which will be followed by the unsharp filter as shown in Figure 5.

CLAHE is used for improving the image contrast and was originally given by Zuiderveld [30]. When comparing the original fundus image with the enhanced one, a significant improvement is observed in the information of the enhanced fundus image. Deep learning techniques, including CNN, are susceptible to overfitting. Training the model using a limited dataset causes overfitting and fails when the trained model is applied to a new dataset [33]. The large number of fundus images used in the training reduces errors. We doubled the datasets under processing to decrease overfitting and to provide an unbiased evaluation by applying vertical mirroring before applying the CNN. The increase in the number of layers of networks does not always guarantee higher performance. The deep CNNs extract several low- and high-level features. When the image features get more complicated, it becomes more difficult to interpret [33]. So, the proposed CNN design wasn't too deep.

B. Method 1: Reuse Pretrained CNN

The pre-trained CNN models are trained with the ImageNet database and can classify colour images into over 1000 object categories, such as pencil, clock, cat, and many animals.

1) Method 1A: Pretrained CNN as a classifier

The following diagram shown in Figure 6 illustrates how to reuse pre-trained CNNs such as Alex-Net, Google-Net, and Res50-Net for training from scratch on color fundus images. A new fully connected classification layer with five neurons is inserted at the end of the CNN without removing any layer from the pre-trained CNN. The inserted layer named 'new_fc5Classes' is used for classifying DR grades (0: no DR, 1:4 according to the severity of DR) and is shown in Table 1. The pre-trained CNN requires input color images of size 224×224×3 or 227×227×3, but the color fundus images in the image datastores have different sizes.

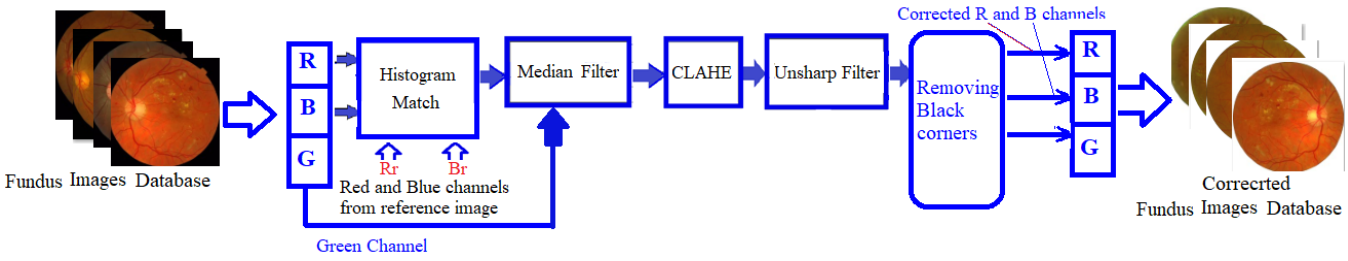

Figure 5: Preprocessing.

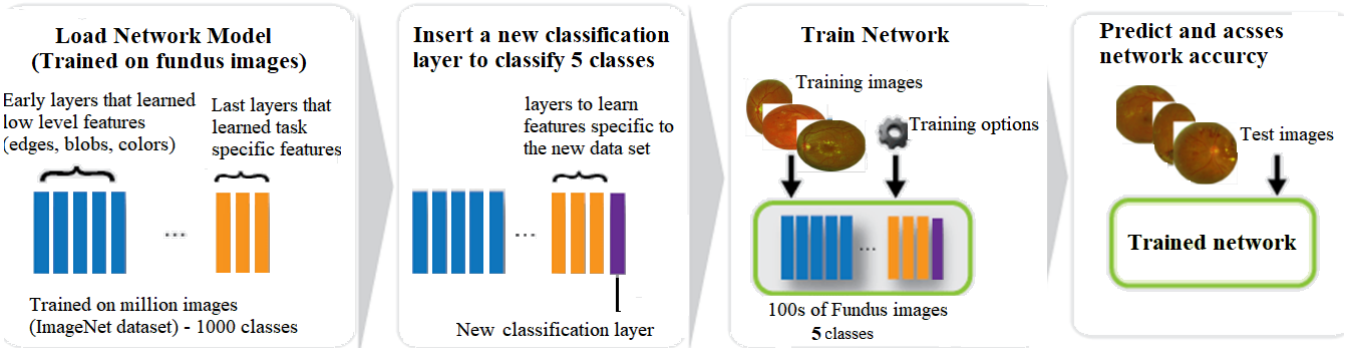

Figure 6: Reuse Pretrained Network to be trained on fundus images.

TABLE:1 THE LAST LAYERS OF GOOGLE-NET AFTER INSERTION OF THE NEW FULLY CONNECTED LAYER

	Name	comments	Type	Activations	Learnable
142	Loss3-classifier 1000 fully connected layer	The layer to be used for features extraction	Fully Connected	1×1×1000	Weights 1000×1024 Bias 1000×1
143	New-fc5Classes 5 fully connected layer	The new inserted layer for classifying 5 classes	Fully Connected	1×1×5	Weights 5×1000 Bias 5×1
144	Prob SoftMax	--	SoftMax	1×1×5	--
145	new_classoutput	--	Classification Output	--	--

We used data augmentation to specify the desired image size. This helps to resize the training images automatically before they are input to the network. The Pretrained CNNs are trained from scratch over fundus images of the Messidor-2 database.

1) Method 1B: Pretrained CNN Models to extract features

In this method, we are not going to use the CNN model for the classification task. We are going to use it to extract features. Each layer produces activation values responding to the input fundus image. Not all layers within a CNN are suitable for image feature extraction; only a few are. The beginning layers of the network are not suitable for feature extraction because they capture basic image features. It is logically preferred to extract features from one of the last layers using the activation method. We think that the activation values of the layer before the last classification layer are more suitable. For example, in Google-Net the layer is named “loss3-classifier”, in Res50-Net, this layer is named 'fc1000’, and in Alex-Net, this layer is named "fc8". The chosen layer in all three CNN models under processing outputs a vector of 1000 features for every input color fundus image (see Figure 7). The multiclass Support Vector Machine is used to be trained on feature vectors extracted from the pre-trained convolutional neural network, which is trained on our fundus images. Feature extraction only requires a single pass through the data.

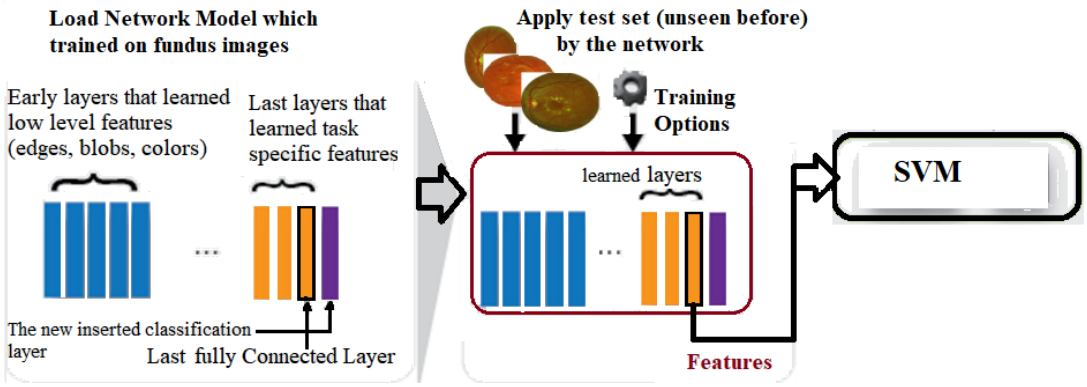

Figure 7: Reuse Network Model to extract features

C. Method 2: Proposed CNN Models (DR2Net and DR5Net)

We used our deep learning CNN model as a binary classifier (DR2Net), which was later modified to (DR5Net) as a five-class classifier, which is shown in Table 2. We used a small filter size of 3 ×3 in the convolution layer to avoid losing the small details and features of the fundus image texture. The layers fc_1 and fc_2 are fully connected layers with 50 neurons for each layer. In the binary classifier DR2Net, Fc_3 is a fully connected layer with 2 neurons.

TABLE 2: OUR DEEP LEARNING CNN MODEL (DR5NET)

	Name	Type	Activations	Learnable
1	Image input	Image	224×224×1	---
2	conv_1 163×3×1convolution with stride [1 1] padding 'same'	Convolution	224×224×16	Weights 3×3×1×16 Bias 1×1×16
3	batchnorm_1 batch normalization with 16 channels	Batch Normalization	224×224×16	Offset 1×1×16 Scale 1×1×16
4	Relu_1	ReLU	224×224×16	---
5	maxpool_1 2×2 max pooling with stride [2 2] and padding [0000]	Max Pooling	112×112×16	---
6	conv_2 32 3×3×16 convolution with stride [1 1] padding 'same'	Convolution	112×112×32	Weights 3×3×16×32 Bias 1×1×32
7	batchnorm_2 batch normalization with 32 channels	Batch Normalization	112×112×32	Offset 1×1×32 Scale 1×1×32
8	Relu_2	ReLU	112×112×32	---
9	maxpool_2 2×2 max pooling with stride [2 2] and padding [0000]	Max Pooling	56×56×32	---
10	conv_3 64 3×3×32 convolution with stride [1 1] padding 'same'	Convolution	56×56×64	Weights 3×3×32×64 Bias 1×1×64
11	batchnorm_3 batch normalization with 64 channels	Batch Normalization	56×56×64	Offset 1×1×64 Scale 1×1×64
12	relu_3	ReLU	56×56×64	---
13	fc_1	Fully connected layer	1×1×50	Weights 50×200704 Bias 50×1
14	fc_2	Fully connected layer	1×1×50	Weights 50×50 Bias 50×1
15	fc_3	Fully connected layer	1×1×5	Weights 5×50 Bias 5×1
16	softmax	Softmax	1×1×5	---
17	class output	Classification Output	---	---

IV. RESULTS AND DISCUSSIONS

A. *Results of Pretrained CNN Models (Method 1)*

The CNN models Alex-Net, Google-Net, and Res50-Net with the inserted layer 'new_fc5Classes' are used for classification. We trained the modified CNN models as shown in Figure 6 from scratch on 70% of the Messidor-2 database after applying preprocessing steps at the beginning as shown in Figure 5. The remaining 30% of the dataset contains fundus images, which have never been seen before by the CNN model and are used for validation. These steps are repeated for Alex-Net, Google-Net, and Res50-Net. The accuracy results range from 72% to 75% (see Table 3). The results of using pertained CNN models as classifiers did not encourage further testing with other datasets. The CNN models Alex-Net, Google-Net, and Res50-Net with the inserted layer 'new_fc5Classes' are used for image feature extraction. We used the layer 'loss3-classifier' in Google-Net, 'fc1000' in Res50-Net, and 'fc8' in Alex-Net to extract features. The chosen layer outputs a vector of 1000 features, responding to each input fundus image. The multiclass linear SVM was trained with these CNN features to grade DR severity in fundus images. The Alex-Net results show a total accuracy of 73.16% and an F1-score of 0.6634. The Google-Net results using SVM had a total accuracy of 73.9% and an F1-score of 0.5682 for five classes (see Table 4).

TABLE 3: CLASSIFICATION RESULTS OF TRAINING PRETRAINED CNN WITH FUNDUS IMAGES.

	Alex-Net	Google-Net	Res50-Net
Accuracy	72.11	74.02	75%

TABLE 4: CLASSIFICATION RESULTS OF USING PRETRAINED CNN MODELS FOR FEATURES EXTRACTION.

	Alex-Net	Google-Net	Res50-Net
Accuracy	73.16%	73.9%	74.2%
F1-score	0.6634	0.5682	0.722

These CNN modes use filter sizes and hyperparameters that are not compatible with the importance of small details in fundus image texture. The filter sizes and hyperparameters in pre-trained models ignore the special nature of the medical image. This may lead to losing important features of small local details.

B. *Results of Proposed CNN Models (Method 2)*

Different input image sizes have been tested to find the smallest possible size without affecting the performance. It has been discovered that 224 is an acceptable size. In this paper, the model is trained with a split ratio of 80/20 and 70/30 of the dataset, which means training the model from scratch with 80% and testing it using the remaining 20% of the dataset, then repeating training from scratch with 70% and testing it using the remaining 30% of the dataset. Three datasets (Messidor-1, Messidor-2, and Kaggle EyePACS datasets) are utilized for training and testing on the binary classifier model DR2Net. Figure 8 shows the classification results of DR2Net on the Messidor1 dataset. The resulted accuracy at a split ratio of 80/20 is 93.33%, sensitivity is 93.5%, and precision is 94.2% with an AUC equal to 0.93427 and an F1-score of 0.9387. The accuracy at a split ratio of 70/30 is 87.916%, the sensitivity is 85.5%, and the precision is 91.8%, with an AUC of 0.8721 and an F1-score of 0.885.

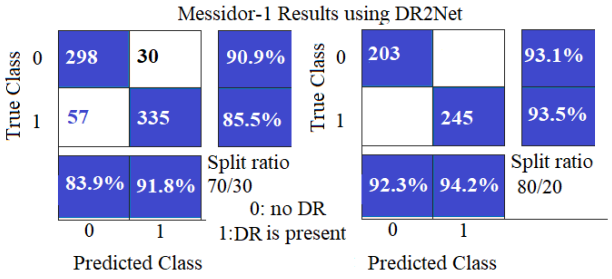

Figure 8: Confusion matrixes of testing DR2Net on Messidor1 Dataset

Training and testing of the DR2Net model by Messidor1 without preprocessing or color correction. The resultant

accuracy was 82.50%, and the AUC was 0.8039. The resulted accuracy of testing the Messidor2 dataset at a split ratio of 80/20 is 96.84%, sensitivity is 97.3%, precision is 95.4%, and F1-score is 0.9629 with an AUC equal to 0.9559. As shown in Figure 9, the accuracy of testing the Messidor2 dataset at a split ratio of 70/30 is 92.8%, the sensitivity is 91.8%, the precision is 91.1%, and the F1-score is 0.9143 with an AUC of 0.932.

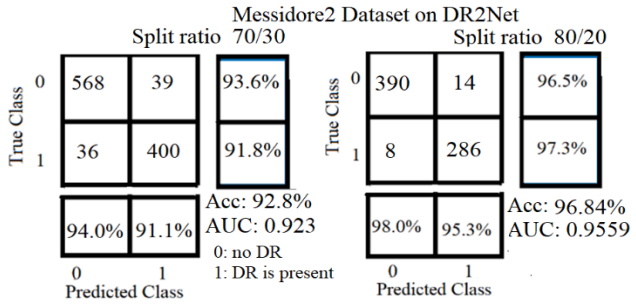

Figure 9: Confusion matrixes of testing Messidor2 Dataset using DR2Net

The accuracy of the testing EyePACS dataset at a split ratio of 80/20 is 95.22%, the sensitivity is 94.0%, and the precision is 94.6%, with an AUC of 0.94749 and an F1-score of 0.9430. As shown in Figure 10, the accuracy of testing the EyePACS dataset at a split ratio of 70/30 is 92.83%, the sensitivity is 90.2%, and the precision is 92.3%, with an AUC of 0.92647 and an F1-score of 0.91217.

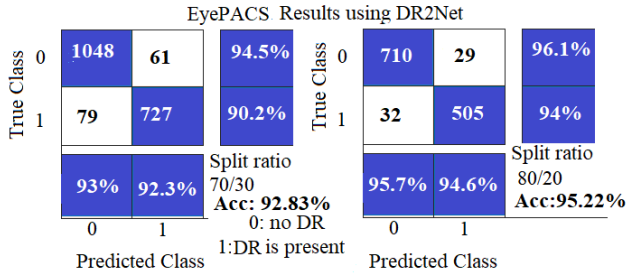

Figure 10: Confusion matrixes of testing Kaggle EyePACS Dataset using DR2Net

A comparison among the results of the three datasets is shown in Table 5. The best performance was with Messidor2. The Messidor1 data set has the worst performance, maybe because it has the smallest number of fundus images. Kaggle EyePACS images are captured using different cameras, and many images have noise and contain artifacts. Some are out of focus, underexposed, or overexposed. Its good results prove the good effect of preprocessing.

TABLE 5: CLASSIFICATION RESULTS OF USING DR2NET MODEL ON THREE DATASETS AT SPLIT RATIOS 70/30 AND 80/20.

Split ratio	Metric	Messidor1	Messidor2	EyePACS Dataset
70/30	Sensitivity	85.5%	91.8%	90.2%
	Accuracy	87.91%	92.8%	92.83
	AUC	0.8721	0.932	0.92647
	Precision	91.8%	91.1%	92.3%
	F1-score	0.9357	0.9142	0.91217
80/20	Sensitivity	93.5%,	97.3%	94.0%
	Accuracy	93.33%	96.85%	95.22%
	AUC	0.93427	0.9559	0.94749
	Precision	94.2%	95.3%	94.6%
	F1-score	0.9387	0.9629	0.9430

In the five-classes model DR5Net, the layers fc_1 and fc_2 are fully connected layers with 50 neurons for each layer, and Fc_3 is a fully connected layer with 5 neurons. The same process that is followed with DR2Net is followed for training DR5Net. The model was applied to the Messidor1 dataset and resulted in the highest accuracy of 90.83% and an F1-score of 0.9133. Details of the results are shown in the following Figure 11. and Table 6.

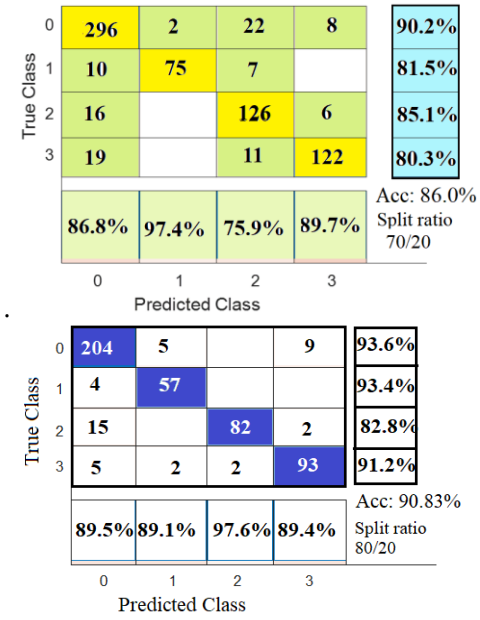

Figure 11: Confusion matrixes of testing the DR5Net model by Messidor1 Dataset at split ratios 70/30 and 80/20.

TABLE 6: SENSITIVITY, PRECISION, AND F1-SCORE OF TESTING THE MODEL DR5NET ON MESSIDOR1 DATASET.

Split ratio	Metric	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
70/30	Sensitivity	90.2%	81.5%	85.1%	80.3%	It hasn't
	Accuracy	86.00%				
	Precision	86.8%	97.4%	75.9%	89.7%	It hasn't
	F1-score	0.8717				
80/20	Sensitivity	93.6%	93.4%	82.8%	91.2%	It hasn't
	Accuracy	90.83%				
	Precision	89.5%	89.1%	97.6%	91.2%	It hasn't
	F1-score	0.9133				

DR5Net was applied to the Messidor2 dataset and resulted in its highest accuracy of 94.11% and an F1-score of 0.9226 at a split ratio of 80/20. The details of the results are shown in the following Figure 12 and Table 7. The Kaggle EyePACS dataset resulted in its highest accuracy of 95.06% and an F1-score of 0.9393 at a split ratio of 80/20 as shown in Figure 13 and Table 8.

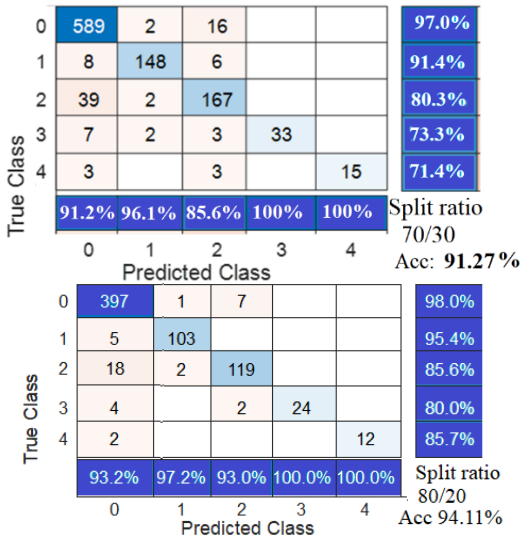

Figure 12: Confusion matrixes of testing the DR5Net model by Messidor2 Dataset at split ratios 70/30 and 80/20.

TABLE 7: SENSITIVITY, PRECISION, AND F1-SCORE OF TESTING THE MODEL DR5NET ON MESSIDOR2 DATASET.

Split ratio	Metric	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
70/30	Sensitivity	97%	91.4 %	80.3 %	73.3%	71.4%
	Accuracy	91.27%				
	Precision	91.2 %	96.1 %	85.6 %	100%	100%
	F1-score	0.8886				
80/20	Sensitivity	98.0 %	95.4 %	85.6 %	80.0%	85.7%
	Accuracy	94.11%				
	Precision	93.2 %	97.2 %	93.0 %	100.0 %	100.0 %
	F1-score	0.9226				

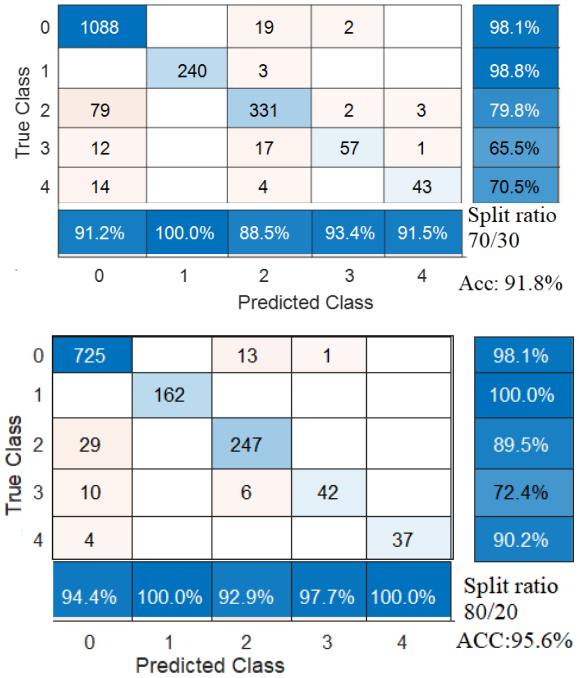

Figure 13: Confusion matrixes of testing the DR5Net model by Kaggle EyePACS Dataset at split ratios 70/30 and 80/20.

TABLE 8: SENSITIVITY AND PRECISION OF DR GRADES RESULTED FROM TESTING THE MODEL DR5NET ON EYEPACS DATASET.

Split ratio	Metric	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
70/30	Sensitivity	98.1 %	98.8%	79.8 %	65.5 %	70.5%
	Accuracy	91.8%				
	Precision	91.2 %	100%	88.5 %	93.4 %	91.5%
	F1-score	0.8958				
80/20	Sensitivity	98.1 %	100.0 %	89.5 %	72.4 %	90.2%
	Accuracy	95.06%				
	Precision	94.4 %	100.0 %	92.9 %	97.7 %	100.0 %
	F1-score	0.9393				

The proposed method applies two CNN models without performing any image segmentation step. The DR detection model (DR2Net) was approved for its excellency by its results compared to others. The grading DR model (DR5Net) grades the DR into five classes and not 3 classes like others [10, 23]. The achieved results outperform the results of comparable ultramodern techniques, as shown in Table 9.

TABLE 9: COMPARISON WITH RELATED WORKS

Cited Papers	Database	Methodology	Result
Shaban et al (2020) [10]	EyePACS (3,648) images	The system outputs only three classes by merging mild and moderate in one class and severe NPDR and PDR in one class then using (CNNs) for classification to Grade DR severity.	Sensitivity:87%-89% Acc: 88%-89% for only 3-classes
Li, 2020 [12]	Messidor-1	Grading DR severity using attention Deep Learning Network based on ResNet50	Acc: 92.6%, Sensitivity: 92.0%,
Chetoui et al.,2020 [13]	Messidor-1	Detecting DR using CNNs (binary classifier) in normal and abnormal.	AUC: 0.963
	EyePACS		AUC: 0.986, Sensitivity: 0.958
Abr`amoff et al. (2016) [23]	Messidor-2	The system outputs only three classes by merging no DR, mild in one class and moderate and severe NPDR in one class, and PDR in one class then using (CNNs) for classification to Grade DR severity.	Sensitivity: 96.8% for only 3-classes
Gargeya and Leng(2017) [24]	Messidor-2	Detecting DR using CNNs (binary classifier) in normal and abnormal.	Sensitivity: 93%

Lam et al., 2018. [31]	Messidor-1	Grading DR severity using GoogLeNet and AlexNet models.	Acc:74.5%
Yung-hui et al. (2019) [32]	EyePACS	(CNN)	Acc: 86.17
Costa, 2018. [34]	Messidor-1	Grade DR severity using Multiple Instance Learning	AUC: 0.9
Dutta, 2018 [35]	EyePACS	Grading DR severity using VGGNet 16	Acc: 86.30%
Kwasigroch, 2018[36]	EyePACS	VGGNet Model	Acc: 81.70% Sensitivity: 89.50%
Chowdhury, 2019. [37]	EyePACS	Inception v3 Model (binary classifier) in normal and abnormal.	Acc: 61.3%
Sayres et al.,2019.[38]	EyePACS 2000 images	Grading DR severity using customized networks CNN	Acc: 88.4%, Sensitivity: 91.5%,
Sengupta, 2019 [39].	EyePACS	Inception-v3 Model	Acc: 90. 4% Sensitivity: 90%
Pao, 2020 [41]	EyePACS	Bi-channel customized CNN	Acc: 87.83% Sensitivity:77.81% Specificity: 93.88% AUC: 0.93
Samanta, 2020, [42].	EyePACS	DenseNet121 based	Acc: 84.1%
Thota, 2020 [43]	EyePACS	VGGNet Model	Acc: 74% Sensitivity: 80.0% Specificity: 65.0% AUC: 0.80
Ludwig, 2020 [40].	Messidor-2	Detect referral-warranted diabetic retinopathy (RDR) using DenseNet201	Acc: 87% Sensitivity: 80%
Proposed Method (Binary classifier)	Messidor-1	Proposed CNN model DR2Net	Sensitivity: 93.5% Acc: 93.33% AUC: 0.93427 Precision: 94.2% F1-score: 0.9387
	Messidor-2 (1745 images)		sensitivity 97.3% Acc: 96.85% AUC: 0. 9559 Precision 95.3% F1-score: 0.9629
	EyePACS (3,190) images		sensitivity :94.0% Acc: 95.22% AUC:0.94749 Precision: 94.6% F1-score: 0.9629
Proposed Method Grading DR	Messidor-1	Proposed CNN model DR5Net	Acc: **90.83%** F1-score 0.9133
	Messidor-2		Acc: 94.11% F1-score: 0.9226
	EyePACS (3548) images		Acc: 95.06% F1-score: 0.9393

V. Conclusion

Two new CNN architecture models are proposed. The first one, DR2Net, is for classifying fundus images into two classes: normal and abnormal. The second one is DR5Net, for classifying fundus images into five classes according to the severity of DR. The two models successfully classified fundus images. The situation of our results stands in a very good position among other leading studies carried out on the same datasets. The two proposed models outperform recent works, outperforming pre-trained models such as Google-Net, Alex-Net, and Res50-Net.Using a pre-trained model was not successful in extracting features from fundus images. Classifying the severity of DR in fundus images needs more research to obtain better performance near 100%. Preprocessing techniques enhanced the accuracy performance by 10.83% and AUC by 0.13037 using the binary model.

VI. Statements and Declarations

The author confirms that no conflict of interest and there are no financial funds.

VII. References

[1] Boles, S.F. and A.A.E. Center, Diabetic Retinopathy: What You Should Know, https://www.nei.nih.gov/sites/default/files/2019-06/Diabetic-Retinopathy-What-You-Should-Know-508.pdf

[2] Khalil SA, Megallaa MH, Rohoma KH, Guindy MA, Zaki A, Hassanein M, Malaty AH, Ismael HM, Kharboush IF, Kafash E. Prevalence of chronic diabetic complications in newly diagnosed versus known type 2 diabetic subjects in a sample of Alexandria population, Egypt. Medicine Current diabetes reviews. 2019;15(1):74–83. DOI:10.2174/1573399814666180125100917

[3] Shaw, J.E., R.A. Sicree, and P.Z. Zimmet, Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes research and clinical practice, 2010. 87(1): p. 4-14.

[4] N. Singh and L. Kaur, "A survey on blood vessel segmentation methods in retinal images," 2015 International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV), 2015, pp. 23-28, doi: 10.1109/EDCAV.2015.7060532. Ng, D.S., et al., Retinal ganglion cell neuronal damage in diabetes and diabetic retinopathy. Clinical & experimental ophthalmology, 2016. 44(4): p. 243-250.

[5] Group, E.T.D.R.S.R., Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie House classification: ETDRS report number 10. Ophthalmology, 1991. 98(5): p. 786-806.

[6] Williams, R., et al., Epidemiology of diabetic retinopathy and macular oedema: a systematic review. Eye, 2004. 18(10): p. 963.

[7] Boelter, M.C., et al., Fatores de risco para retinopatia diabética. Arquivos brasileiros de oftalmologia. São Paulo. Vol. 66, n. 2 (mar./abr. 2003), 239-247, 2003.

[8] Singh, R., et al., Diabetic retinopathy: an update. Indian journal of ophthalmology, 2008. 56(3): p. 179.

[9] Shaban M, Ogur Z, Mahmoud A, Switala A, Shalaby A, Abu Khalifeh H, et al. (2020) A convolutional neural network for the screening and staging of diabetic retinopathy. PLoS ONE 15(6): e0233514. https://doi.org/10.1371/journal. pone.0233514…

[10] K. Pin, J. -Y. Kim, J. H. Chang and Y. Nam, "Quality Evaluation of Fundus Images using Transfer Learning," *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2020, pp. 742-744, doi: 10.1109/CSCI51800.2020.00139.

[11] Li, X., Hu, X., Yu, L., Zhu, L., Fu, C.-W., & Heng, P.-A. (2020). CANet: Cross-Disease Attention Network for Joint Diabetic Retinopathy and Diabetic Macular Edema Grading. IEEE Transactions on Medical Imaging, 39(5), 1483–1493. https://doi.org/10.1109/tmi.2019.2951844

[12] Mohamed Chetoui and Moulay A. Akhloufi "Explainable end-to-end deep learning for diabetic retinopathy detection across multiple datasets," Journal of Medical Imaging 7(4), 044503 (28 August 2020). https://doi.org/10.1117/1.JMI.7.4.044503

[13] W. Luangruangrong, P. Kulkasem, S. Rasmequan, A. Rodtook, and K. Chinnasarn, "Automatic exudates detection in retinal images using efficient integrated approaches," vol. 2014, 2014 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2014.

[14] Krizhevsky Alex, Sutskever Ilya, Hinton Geoffrey E. ImageNet classification with deep convolutional neural networks. Communications of the ACM 2017;60(6): 84–90. https://doi.org/10.1145/3065386.

[15] Nair Vinod, Hinton Geoffrey E. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML); 2010.

[16] Srivastava Nitish, et al. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929–58.

[17] IIoffe Sergey, Szegedy Christian. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. arXiv preprint arXiv: 1502.03167.

[18] Navarro, P.J., D. Alonso, and K. Stathis, Automatic detection of microaneurysms in diabetic retinopathy fundus images using the L* a* b color space. JOSA A, 2016. 33(1): p. 74-83.

[19] Kumar, S. and B. Kumar. Diabetic Retinopathy Detection by Extracting Area and Number of Microaneurysm from Colour Fundus Image. in 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN). 2018. IEEE.

[20] Hashim, M.F. and S.Z.M. Hashim. Diabetic retinopathy lesion detection using region-based approach. in 2014 8th. Malaysian Software Engineering Conference (MySEC). 2014. IEEE.

[21] Asia Pacific Tele-Ophthalmology Society, "APTOS 2019 blindness detection," Kaggle, https://www.kaggle.com/c/aptos2019-blindness-detection/data, 2019, [Dataset]

[22] Michael David Abràmoff, Yiyue Lou, Ali Erginay, Warren Clarida, Ryan Amelon, James C. Folk, Meindert Niemeijer; Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning. *Invest. Ophthalmol. Vis. Sci.* 2016;57(13):5200-5206. doi: https://doi.org/10.1167/iovs.16-19964.

[23] Gargeya Rishab, Leng Theodore. Automated identification of diabetic retinopathy using deep learning. Ophthalmology 2017;124(7):962–9.

[24] Ting Daniel Shu Wei, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. Jama 2017;318(22):2211–23.

[25] Gulshan Varun, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama 2016; 316(22):2402–10.

[26] Krause Jonathan, et al. Grader variability and the importance of reference standardsfor evaluating machine learning models for diabetic retinopathy. Ophthalmology 2018;125(8):1264–72.

[27] Gupta Garima, Ram Keerthi, Kulasekaran S, Joshi Niranjan, Sivaprakasam Mohanasankar, Gandhi Rashmin. Detection of retinal hemorrhages in the presence of blood vessels. Proceedings of the ophthalmic medical image analysis, OMIA 2014, MICCAI 2014, boston, Massachusetts, September 14; 2014. p. 105–12. *doi*: *https*://*doi.org/10.17077/omia.1015*.

[28] M. U. Akram, S. Khalid, and S. A. Khan, "Identification and classification of microaneurysms for early detection of diabetic retinopathy," Pattern Recognition, vol. 46, no. 1, pp. 107–116, 2013. https://doi.org/10.1016/j.patcog.2012.07.002.

[29] K. Zuiderveld. 'Contrast limited adaptive histogram equalization,' Graphics gems IV, 42, no. 1, pp. 474–485, 1994.

[30] Lam C, Yi D, Guo M, Lindsey T. Automated Detection of Diabetic Retinopathy using Deep Learning. AMIA Jt Summits Transl Sci Proc. 2018 May 18;2017:147-155. PMID: 29888061; PMCID: PMC5961805.

[31] *Li, Yung-hui et al. "Computer-Assisted Diagnosis for Diabetic Retinopathy Based on Fundus Images Using Deep Convolutional Neural Network." Mob. Inf. Syst. 2019, 6142839:1-6142839:14.* https://doi.org/10.1155/2019/6142839

[32] Lakshminarayanan V, Kheradfallah H, Sarkar A, Jothi Balaji J. Automated Detection and Diagnosis of Diabetic Retinopathy: A Comprehensive Survey. J Imaging. 2021 Aug 27;7(9):165. doi: 10.3390/jimaging7090165. PMID: 34460801; PMCID: PMC8468161.

[33] P. Costa, A. Galdran, A. Smailagic and A. Campilho, "A Weakly-Supervised Framework for Interpretable Diabetic Retinopathy Detection on Retinal Images," in *IEEE Access*, vol. 6, pp. 18747-18758, 2018, doi: 10.1109/ACCESS.2018.2816003.

[34] Dutta, S.; Manideep, B.C.; Basha, S.M.; Caytiles, R.D.; Iyengar, N.C.S.N. Classification of Diabetic Retinopathy Images by Using Deep Learning Models. Int. J. Grid Distrib. Comput. 2018, 11, 99–106.

[35] A. Kwasigroch, B. Jarzembinski and M. Grochowski, "Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy," *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 2018, pp. 111-116, doi: 10.1109/IIPHDW.2018.8388337.

[36] Chowdhury, M.M.H.; Meem, N.T.A. A Machine Learning Approach to Detect Diabetic Retinopathy Using Convolutional Neural Network; Springer: Singapore, 2019; pp. 255–264.

[37] Sayres, R.; Taly, A.; Rahimy, E.; Blumer, K.; Coz, D.; Hammel, N.; Krause, J.; Narayanaswamy, A.; Rastegar, Z.; Wu, D.; et al. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. Ophthalmology 2019, 126, 552–564.

[38] Sourya Sengupta, Amitojdeep Singh, John Zelek, and Vasudevan Lakshminarayanan "Cross-domain diabetic retinopathy detection using deep learning", Proc. SPIE 11139, Applications of Machine Learning, 111390V (6 September 2019); https://doi.org/10.1117/12.2529450

[39] Cassie A. Ludwig, Chandrashan Perera, David Myung, Margaret A. Greven, Stephen J. Smith, Robert T. Chang, Theodore Leng; Automatic Identification of Referral-Warranted Diabetic Retinopathy Using Deep Learning on Mobile Phone Images. *Trans. Vis. Sci. Tech.* 2020;9(2):60. doi: https://doi.org/10.1167/tvst.9.2.60.

[40] Pao, S.-I.; Lin, H.-Z.; Chien, K.-H.; Tai, M.-C.; Chen, J.-T.; Lin, G.-M. Detection of Diabetic Retinopathy Using Bichannel Convolutional Neural Network. J. Ophthalmol. 2020, 2020, 1–7.

[41] Samanta, A.; Saha, A.; Satapathy, S.C.; Fernandes, S.L.; Zhang, Y.-D. Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset. Pattern Recognit. Lett. 2020, 135, 293–298.

[42] Thota, N.B.; Reddy, D.U. Improving the Accuracy of Diabetic Retinopathy Severity Classification with Transfer Learning. In Proceedings of the 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), Springfield, MA, USA, 9–12 August 2020; pp. 1003–1006.