MONASH University

School Of Information Technology

FIT - 3161

Computer Science Project 1

Semester 1

by

Nabil Ahmed (ID: 25364170)

Kevin Setiawan (ID:25895710)

Literature Review

Classification of Diabetic Retinopathy using Retinal Images (Convolutional Neural Network with MATLAB + Fusion Techniques)

Abstract

Diabetic Retinopathy(DR) is a disease that occurs in the retina of the eye due to long term diabetes. DR is one of the biggest cause of blindness. Early detection of DR can help patients to prevent blindness and save cost. It requires a highly trained clinician and an expert in the field of DR to detect DR from retinal scans. Not only is there a lack of trained clinicians and equipments in areas where DR among general population is high, but also detecting DR is time consuming. The purpose of this literature review is to show that there already exists some methods to detect DR automatically and classify them. In recent years CNN has shown promising results in image classification. By extracting some features like blood vessels, microaneurysms, hemorrhages and exudates from retinal images, DR can be classified into several stages in real time. Automatic DR detection using CNN approach shows encouraging result. This literature reviewed robustly explores eight articles and considers the possibility of effectively and accurately classifying different stages of DR using different CNN approach.

Keywords

CNN, Diabetic Retinopathy, Retinal Images, Image Classification, Fine- Tune, cross validation, Retina, Blood Vessel, Deep Learning, Hemorrhages, Exudates, Microaneurysms, Diabetes.

Introduction

Diabetes is a well-known disease which arises when the body cannot process the insulin secreted by the pancreas or when the pancreas fails to secrete enough of it [Verma, Deep et al. 2011]. The longer a patient has diabetes and the less controlled their blood sugar is, the higher the chances for them for having Diabetic retinopathy(DR). DR affects the patient's eyes. It is caused by damage to the blood vessels of the light-sensitive tissue at the back of the eye (retina) which can lead to blindness in worst case [Doshi, Shenoy et al. 2016].

An example of the patient's vision having DR is shown below.

Fig. 1: Normal Vision[7]

Fig. 2: Vision with diabetic retinopathy[7]

Available : https://en.wikipedia.org/wiki/Diabetic retinopathy

Based on the survey conducted by The World Health Organization(WHO), there are more than 347 million people affected by diabetic retinopathy, this number is estimated to rise to 552 million by 2030. Clinicians identify DR by detecting lesions with vascular abnormalities, however, the clinician must be quite trained and an expert in the given field. The need of specific equipment and trained clinician for detecting DR is a major challenge in many underdeveloped or overpopulated nations like India and China [Ghosh, Ghosh et al. 2017]. For DR classification, numerous features are required to be identified and located which is difficult and time consuming for clinicians. Once trained, computers can obtain much better classification in very less time [Pratt, Coenen et al. 2016]. Thus, automated detection technique for identifying DR is distinguished to tackle the growing number of individuals with diabetes.

Diabetic retinopathy can be categorized into two types:
1. Early DR
2. Advanced DR.

Early DR is known as Non-Proliferative Diabetic Retinopathy (NPDR). In NPDR the blood vessel walls of the retina become weak. Micro aneurysm causes the vessel wall to become smaller due to which it sometimes leaks blood and fluid into the retina. NPDR can progress from mild to severe, depending on how much blood vessels blocked.

In advanced DR, the patient’s condition becomes even worse and can eventually lead to blindness.

Detecting DR can be quite challenging because it requires a highly trained clinician and expensive equipment to effectively and accurately examine DR [Doshi, Shenoy et al. 2016]. While the method that are used by clinician is already effective, but sometimes the demands are too high and the time left is very less. Sometimes early detection of DR can make a huge difference in a patient’s life. Since, classified DR involves a lot of features and is highly time consuming for the clinicians, computers can do classification quickly and accurately once it’s trained.

Our subject for the literature reviews is the use of Convolutional Neural Networks (CNN) which is a branch of deep learning. It has a good track record in the field of image analysis. CNN has a good history of successfully addressing complex image recognition task into many different classes.

In this literature review, we are reviewing different CNN methods for detection of DR by using retinal images.

Literature of articles

Verma, Deep et al. 2011 proposes classification of Diabetic Retinopathy without a CNN approach. They do the classification by using the random forest technique. Convolutional Neural Networks only recently became famous in the world of machine learning and artificial intelligence. But, back in 2011, using CNN were still not orthodox with the limitations in both the hardware and software.

They took a dataset with just 65 images acquired through fundus camera. Their primary aim is to detect blood vessels (example shown in figure 1-4), identify hemorrhages (example shown in figure 5-6) and finally classify the different stages of diabetes into normal, moderate and non-proliferative diabetic retinopathy. The different stages of Diabetic Retinopathy are categorized based on the quantity and discovery of the blood vessels and hemorrhages in the retinal image. The contrast between blood vessels and surrounding background in the image is used for segmenting the retinal vascular (retinal vascular basically includes the blood vessels in the retina). Density analysis and bounding box technique is used to distinguish the hemorrhages. Finally, random forest is used for classifying the different stages of Diabetic Retinopathy. They do this classification based on the location and perimeter of the hemorrhages and blood vessels.

Figure 1. Retinal image after removing the noise.

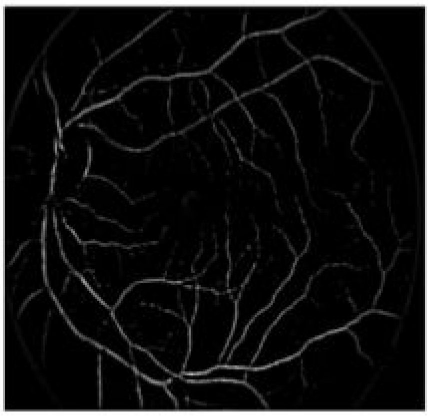


Figure 2. Image obtained after passing through the matched filter.

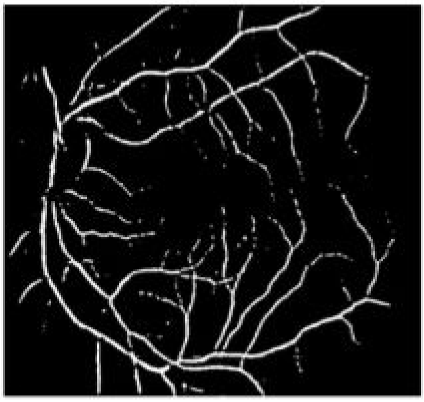


Figure 3. Figure 3: Binary image after thresholding.

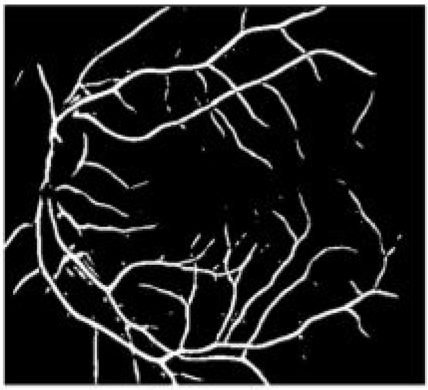


Figure 4. Image after perception based binarisation.

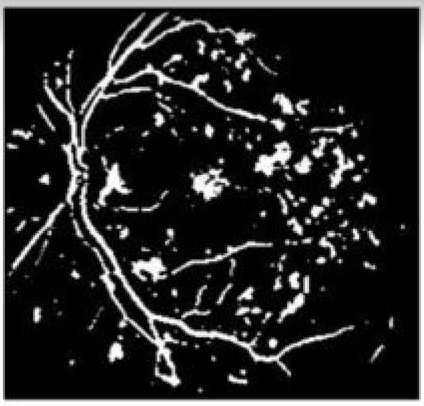


Figure 5. Thresholded image.

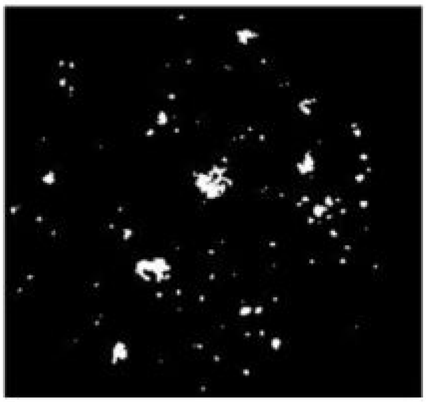


Figure 6. Hemorrhages in the retinal image.

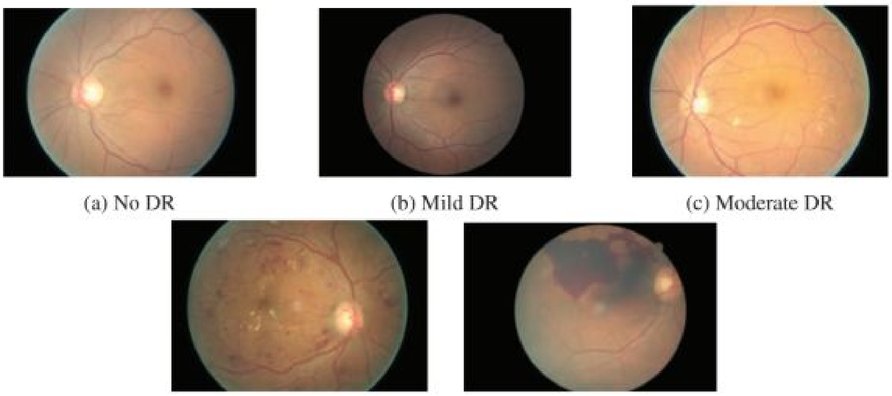
The result of the experiment concluded that severe non-proliferative diabetic retinopathy was classified with an accuracy of 87.5% while normal cases were classified with a 90% accuracy.

This research paper gives us an insight about how to construct an approach for detecting diabetic retinopathy using machine learning. This is used as a base for detecting DR using machine learning for our research. The approaches used in the later articles uses CNN. The researchers did a great job and produced great experimental results especially considering the hardware available at the time and the dataset given which only had 65 images. The scope of the research could be further improved by using a bigger dataset consisting of retinal images with various stages and timeline of the disease. The efficiency could also be improved by extracting more number of features for classification.

Pratt, Coenen et al. 2016 uses a CNN architecture for classifying the severity of Diabetic Retinopathy. They use digital fundus images as input(testing). The features they focus on for the classification purpose are the microaneurysms, hemorrhages and exudate on the retina because these features are observed to be the most crucial when detecting Diabetic Retinopathy [Acharya et al]. When the input image is given the CNN architecture will automatically classify the severity of DR based on the image. The architecture has already been trained by using a high-end GPU (graphics processor unit). The dataset used for training the architecture was taken from the publicly available Kaggle dataset which consisted of 80,000 images. They show remarkable results of 95% sensitivity and 75% accuracy on 5000 validation images.

The proposed CNN classifies the severity of the DR into the five stages:

Harry Pratt et al. / Procedia Computer Science 90 (2016) 200 – 205



(a) No DR (b) Mild DR (c) Moderate DR

(d) Severe DR (e) Proliferative DR

Fig 1: Stages of diabetic retinopathy (DR) with increasing severity

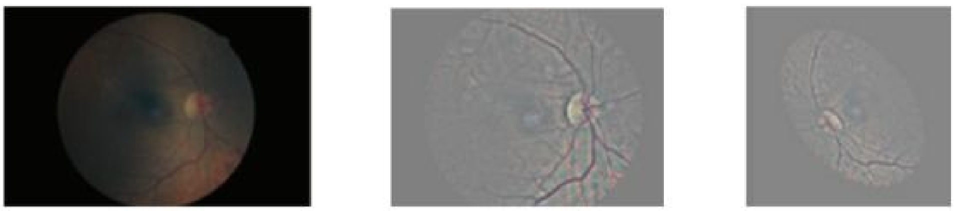
They also claim that they are the first to classify DR into five stages using a CNN approach.

The article is explored in more details compared to other articles as it is completely and directly related to our project literature review.

The Building of the CNN architecture for classifying the different stages of DR

Preprocessing is done on all the images from the Kaggle dataset. The dataset consists of images from patients of varying age groups, people of different races and culture and extremely wide-ranged level of lighting in the fundus photograph. This creates variations not related to classification levels since the pixel intensity is affected. This problem was solved by using color normalization. They also resize the images to one particular size to retain only the intricate features required for identifying while doing classification. All these original preprocessed images are then used for training the network once. Then, real-time data augmentation is used throughout training to improve the localization ability.

Harry Pratt et al. / Procedia Computer Science 90 (2016) 200 – 205



(a) Original image (b) Preprocessed image (c) Augmented image

Fig 3: Illustration of the preprocessing and augmentation processes

The article proves that by using CNN it is possible to classify DR into five different stages. It accurately classifies most of proliferative cases and cases with no DR. They showed that by using a much more general dataset, their method produced comparable results in comparison to other (feature specific detection) methods rather than CNN.
The network achieved this result with only one image as well. Although, the network does not face any problems while detecting the image of a healthy eye, the article does not mention what happens when it uses image of an unhealthy eye. The network also struggled to learn deep enough features of some of the aspects of DR, which is observed from the sensitivity results of mild and moderate classes. This was mainly due to the images in the dataset which was not fully cleaner and required quite a substantial preprocessing step. However, overall the network did a good job in classifying the images into the different categories.

Doshi, Shenoy et al. 2016 proposes the use of deep learning for the automatic detection of DR into its different stages. The proposed model uses a GPU accelerated deep convolutional network to automatically classify into five different stages (no DR, mild, moderate, severe and proliferative). The dataset used to train the model consisted of 35,126 high-resolution color fundus retinal image.

TABLE I: Class Distribution in Original Dataset

Class	Name	No. of images	Percentage
0	No DR	25810	73.48%
1	Mild DR	2443	6.96%
2	Moderate DR	5292	15.07%
3	Severe DR	873	2.48%
4	Proliferative DR	708	2.01%

As shown in the table above a trained clinician has evaluated the existence of DR into the 5 classes for every image in the dataset. The images are taken with different types of cameras so the visual appearance of the images differs. Since, the resolution of all the images does not follow uniformity, they could not be directly utilized for training. They had to be preprocessed to a specific standard for classification. The images were scaled down to a resolution size of 512x512 pixels. Then, they were converted to a single channel(green) since it retains information better than other channels.

3 CNN models are used to measure the accuracy. The results from the three models were evaluated by using a quadratic weighted kappa metric. A best score of 0.3996 was obtained by collaborating the 3 models.

Quality retinal images as input highly affects the result of automatic DR (Tennakoon, Mahapatra et al.2017). They said that using hand-crafted features is not accurate to generalize. Therefore, a new approach by use computational algorithm together with CNN is highly recommended. From digital fundus retinal images can diagnoses DR, age-related macular degeneration and glaucoma. The success of this method depends on quality of input images. Image Quality Classification(IQC) can be divide into 2 categories which is generic and structural quality parameters. Generic image quality parameter like contrast, clarity will include histogram matching and distribution of magnitudes. By using low computational algorithm, this method cannot always get the diversity of condition that affect images. In addition, structural quality parameters such as visibility of optic disc and macula based on retinal landmark. It requires complex segmentation for images. Because of that, the create a new method that uses new algorithm to work on human visual system to identify poor quality images and adapt to new scenarios in database. They create method using CNN with avoid hand crafted features because they believe that combine both structural and geometric information will achieve a high accuracy.

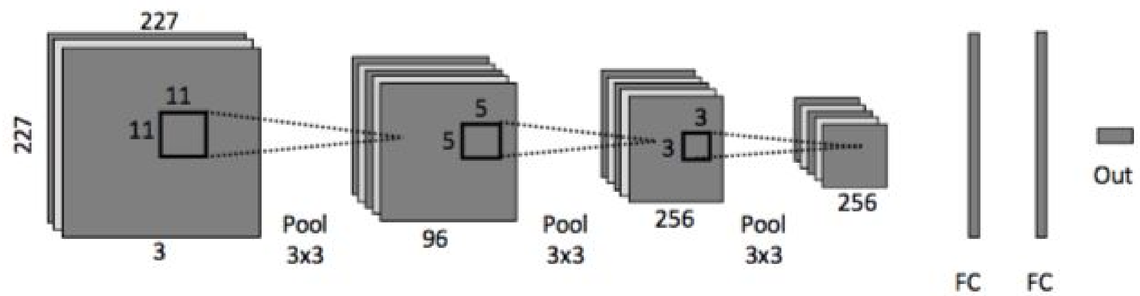
The function that they use in this method stated below:

$$\underset{w}{arg\ max} \bigcup_{i \in \mathcal{D}} p\left(C = c_i \mid I_i, w\right)$$

where D = [Ii , ci]ni=1 is the training dataset and w are the parameters or weights of the CNN model with a specified architecture.

Nowadays,3 main approach that applied on CNN is train from the scratch, used trained CNN to extract many features and unsupervised pre-training to huge dataset. Train CNN from scratch need large size of data which is very difficult to gain for medical application. This experiment will more focus on 2 other CNN architecture. First, shallow CNN which train from scratch using color retinal images and the second method is use transfer learning that use pre-trained filters to extract features.

Shallow net architecture



Use 96, 256,256 convolutional layer of kernel size 11 x 11.5 x 5.3 x 3. Each layer followed by linear activations and max pooling layers. The feature is produced by convolution filter then used in 2 fully connected layer to compute the classification using softmax classification layer. To tackle deep network, they applied batch normalization to compute 2 convolutional layers and dropout regularization to prevent overfitting.

AlexNet architecture is consist of 5 convolutional layers, 3 pooling layers and 2 local response normalization layers and 2 fully connected layers. The approach was trained on natural images from ImageNet. The practice also known as transfer learning and it can tackle of drawback of limited training data. The training images were directly used in the networks. All extracted featured used to train 4 classifiers: single layer Neural Network, linear support vector machine, boosted trees and KNN. Later the hyper-parameter will be used in 5- fold cross validation.

Most CNN method will apply data augmentation to delicate input differences. This method will rotate all training images with fixed angle multiple of 6 from 6 to 210 degree.

The result of this work is highly accurate. All the method achieves more than 96% accuracy. This is table of the result.

Network	Accuracy (%)	Specificity (%)	Sensitivity (%)
Proposed shallowNet	98.27	97.46	99.12
Alexnet-FT	98.27	97.03	99.55
Alexnet-SVM	97.19	95.38	99.12
Alexnet-BT	96.98	99.15	94.71
Alexnet-KNN	96.98	96.19	97.80

According to Ghosh, Ghosh et al. 2017, Convolutional Neural Network (CNN) is one of the approach to automatically detect Diabetic Retinopathy(DR). The researchers use CNN for identify some features such as microaneurysms and hemorrhages on patient's retina. The dataset is provided Kaggle and maintained by EyePacs. It consists color fundus retinal images from different sources. All the images already classified by clinician based on severity of DR. But the dataset that used in this research is highly imbalance that make the model is hard to train. On the pre-processing part, all the images are cropped and resized to squares of 512 pixels for suitable use of CNN. After that, data augmentation is applied in this research to increase the class size also to make the model immune to lighting conditions and several orientations. Then, researches used normalization and denoising.

Furthermore, the model has 6 layers which is convolutional layer, max pooling layer, activation layer, dropout layer, fully connected layer and classification layer. Researches use Xavier initialization to make sure weight are just right when maintain the signal in a realistic range of value over many layers.

$$W = random(fan_{in}, fan_{out})/(sqrt(fan_{in})/2)$$

It will determine the scale of initialization based on input and output.

Next, there is loss function where the softmax layer is used for the prediction.

$$L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right)$$

Where fj is the j-th element of the vector of class scores f.

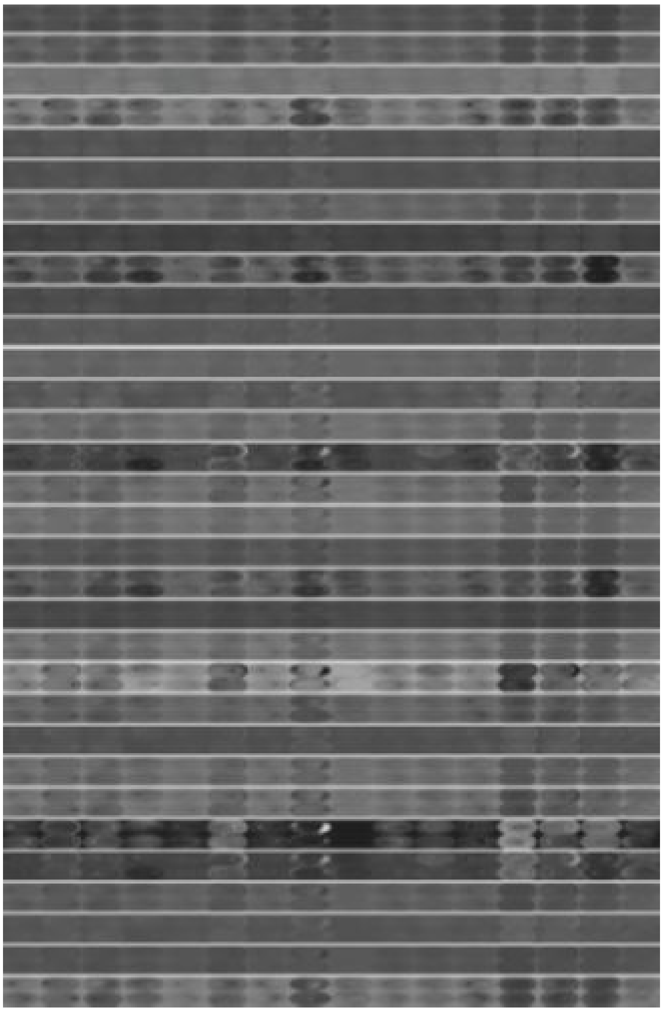
And use dropout technique to effectively do regularization. It performs by keeping the neuron active with some hyper parameter that already set as 0.5 for the max amount regularization. They also do parameter updates by using nesterov momentum because it has been gaining popularity and stronger theoretical convergence for convex functions compare to other version of momentum. Researches use fixed no of 250 epochs is used to train the networks. By training the algorithm on class-balance subset, the problem of high imbalanced dataset can be solved. The dataset also was preprocessed by producing augmentation to create uniformly balanced training set.

In the result, there show the hidden layers and evaluation metrics. 4 metrics are figured for performance evaluation of models. The proportion sample that already classified by trained clinicians gives the accuracy of the model. Because of the dataset is imbalance which cannot be a good measure for the model, the other metrics are recall giving the proportion of positives correctly precision and predicted. Quadratic weighted kappa is computed for performance evaluation

$$w_{i,j} = \frac{(i-j)^2}{(N-1)^2}$$

.

This research take about 10% of images for validation purpose. The final trained achieved around 95% accuracy for 2 class classification and 85% accuracy for 5 class classification. The final model also get 0.74 on kappa score and 0.754 when using ensemble.



Hidden layers

True Label

	0	1	2	3	4
0	**2186**	47	29	31	7
1	84	**79**	34	3	0
2	197	68	**210**	18	7
3	5	2	17	**51**	15
4	6	0	8	37	**49**

Predicted Label

Metrics

Class Label	Precision	Recall
Class0	0.882	0.950
Class1	0.403	0.395
Class2	0.704	0.420
Class3	0.365	0.567
Class4	0.628	0.490

Results

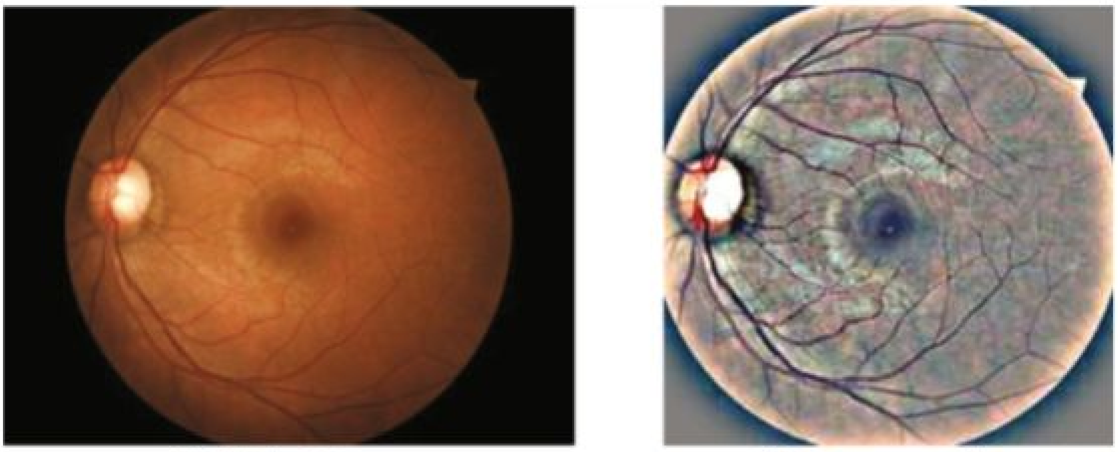
The accuracy of this method can be increased by using other denoising techniques and incorporating experimental errors image capture will make more efficient on normalization methods.

Related to Sun, Wan et al. 2017, the quality classification from retinal images can constructs a good difference in automated DR detection. Also with increasing of application of convenient fundus camera, larger number of retinal images can be taken. With poor quality of training dataset can result low accuracy of detection of DR. The researches propose 4 CNN architectures to classify retinal fundus images. Then, they pick and join the top 2 networks. Four architectures that used in this method is AlexNet, GoggleNet, VGG-16, and ResNet-50.

First, the images will be resized to 256x256 pixels to avoid the disadvantages from lighting conditions on the preprocessing stage. Then, the images will be normalized by using this

$$I(x,y) = \alpha I^o(x,y) + \beta Gaussian(x,y,\omega) * I^o(x,y) + \gamma$$

Also, the images are clipped to 90% of original size to decrease the black space. The effect from preprocessing stages can be seen below.



(a) Original image (b) Preprocessed image

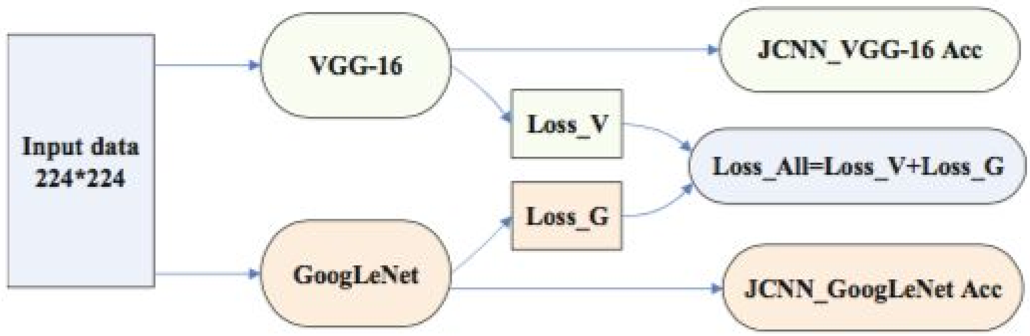
Next method is to do data augmentation which mostly used in training of CNN. The images are rotate randomly or verticals flip with random horizontal or vertical shifts with random horizontal. With this method, the training set increased about 8 times. On the network architecture, deep CNN is pre-train established on huge dataset and weight of train deep CNN will be set earlier.

AlexNet is proposed to get good performance in ImageNet. This CNN include 5 convolutional layer, max pooling layers, dropout layers and 3 fully connected layers.

GoogleNet is the architecture which uses different inception modules to make 22 layers.
VGG-16 only uses 3 x 3 filters in convolutional layers and later combine as a sequence. This create from 13 convolutional layers,5 max pooling layers and 3 fully connected layers.
ResNet-50 is the method with only 3.6% error rate that contains residual block and skip connection bypassing. ResNet has deeper substance which can consists 101 or even 152 layers. This architecture has 6 modules. Conv1 as convolutional layer, conv2 x, conv3 x, conv4 x, conv5 x as residual block and fc as fully connected layer.

Join Fine-tuned CNN: they pick GoogleNet and VGG-16 as 2 top network. The total loss of the network can be defined from this

Loss_All = Loss_V + Loss_G



On the training phase, by removing last fully connected layer and replace with 2 outputs because of in full fine tune, only the last one from layer is touched. The rate of learning last fully connected layer is increased 10 times. For 4CNN architecture, the training data is put in the network together with pre-train weight parameters. Researches fine tune GoogleNet and VGG-16 at the same time with same input for each channel. It will backward circulation training technique to show the classifier slope to all networks
Around 80,000 images are provided by Kaggle website is effectively used in the dataset. They select 2,894 images for sample and 2,170 images for training and test set. The experiment result shown below.

Algorithm	Acc	AUC
AlexNet	96.53%	0.993
GoogLeNet	97.04%	0.994
VGG-16	96.87%	0.995
ResNet-50	96.20%	0.992
JCNN_GoogLeNet	97.00%	0.995
JCNN_VGG-16	**97.12%**	**0.995**

Algorithm	GoogLeNet	GoogLeNet-NP	GoogLeNet-NA
Accuracy	**97.04%**	96.12%	96.49%

As can be seen from table, all method performs a good accuracy on the result. This shown that the knowledge from taken from natural image can be transfer to make image classification effectively. Specially for JCNN between VGG-16 and GoogLeNet, both network already is the best choice network together with they are build based on 2 architectural assumptions that make them cannot miss the classification behavior.

Diabetic Retinopathy became one of the leading cause of blindness (García, Gallardo et al. 2017). Performing manual diagnosis on retinal images and needed of highly trained clinicians to detect small details on retina is time consuming and exhaustive work. The experiment is to create a computer-assisted to help classify retina images to diagnose DR faster and accurately. Using CNN architecture, the computer will identify micro aneurysms, exudates and hemorrhages.

EyePACS as a free platform for retinopathy classification provide the all dataset for this experiment. The dataset already consists retinal images in high resolution quality and several conditions. But the images provided has many differences at aspect ratio, lighting condition and color average. In preprocessing image, all images were scaled into standard size (256x256), subtract the color mean, map in to gray color and separate it into left and right eye. They also perform a binary classification to tackle the class imbalance. The comparison between 2 classes (diseased and healthy) is 1 to 2.74. Another version of data set is created test with ratio 1 to 1.

In the neural networks, many configuration on convolutional network were tested.

Network	Distribution	Layers	Training Mode	Learning rate
$Model_1$	50/50	6	From scratch	0.01
$Model_2$	50/50	9	From scratch	0.01
VGG16	50/50	16	Pre-train	0.0001
$VGG16noFC_1$	50/50	15	Pre-train	0.0001
$VGG16noFC_2$	Original	15	Pre-train	0.0001

All models are motivated from Alex-Net technique. Model 1 and model 2 were train from the scratch so the network must learn simple filtering. But on pre-train mode, network already has many filters and the model is trained in ImageNet database They also just have 1 fully connected layers.

This is the result.

Network	Epochs	Accuracy	Sensitivity	Specificity
$Model_1$	45	63.6%	-	-
$Model_2$	91	66.4%	-	-
VGG16	80	74.3%	62%	86%
$VGG16noFC_1$	75	72.70%	68%	77.60%
$VGG16noFC_2$	80	83.68%	54.47%	93.65%

They believe that they have implemented the most efficient CNN techniques with 93.65% in specificity and 83.68% accuracy. But, VGG16noFC2 only achieves 54.47% in sensitivity.

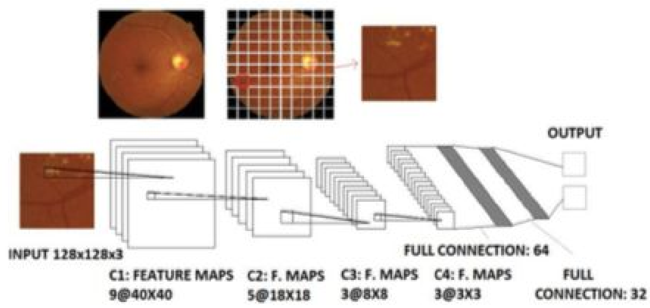
By expanding the retina image database, make a fusion of 2 networks using fully connected layers, it will improve the network architecture and better cost function to database.

Propose CNN with 4 convolutional layers is one of the method to achieve the great performance of DR classification (Andonová, Pavlovičová et al. 2017). They split images based to type of anomalies and used accuracy criteria and cross validation approach to estimate the efficiency of classification. The researchers put some interest on anomalies or exudates. The aim is to look a suitable and effective automatic method for early detection of DR. CNN as good tool for complicated object detection. CNN is part of deep learning and deep neural networks that usually used when it come with huge data and image classification.

1200 fundus retinal images that used are provided from MESSIDOR. The dataset is used for analysis the algorithms for diagnoses purposes and detection of exudates of retinal images. All input was marked by using ophthalmologist. Based on CNN theory, to getting accurate results in classification, it requires huge training data and learning representatives.

First, the input images into blocks with 128 x 128 pixels then split the block to 2 clusters, block with exudates and without exudates. Dataset also will process using many transformations and modifications. To having the images to stage that clearly separate between background and anomalies, researches do some preprocessing images. Adaptive histogram equalization used to increase the improve the contrast. AHE method is by changing each pixel by transformation function consequent from nearby section. Next method is using gauss noise to increase the quality and highlight scratched areas. Then RGB images is converted to grayscale images which can increase the anomalies. In addition, to having the good vision of anomalies on the retinal images is to used green channel in RGB model. Because there are some parts from exudates detection can decrease quality of accuracy, dataset that contains optical disc and blood vessels were deleted from database. Many transformation also applied in this work such as rotation, flipping and cropping.

After do all the preprocessing images, they defined 4 convolutional layers which contains neurons that all connected to sub regions of input / output from previous layer. Max pooling layer also an important step of CNN to reduces the size of input image. After applying all the layer, it will come 2 fully connected layers that have full connections to every action in the previous one. And second fully connected layer will join all the features to classifier.



Before going to classification, cross validation is used to validate CNN. By using 5-fold cross validation, 1 subset is used for testing and the rest for train convolutional. In addition, classification was applied by using CNN. It processed the data with cross validation techniques. Here the result.

TABLE II. RESULTS FROM THE TRAINING PROCESS

Datasets	Accuracy ACC [%]	
	Normal validation	*Cross validation*
Original RGB images	79.69	81.024
Marked images (optic disc and blood vessels supression)	79.79	82.812
Overlapping images	87.84	83.458
Green channel images	77.87	78.846
Gauss noise images	78.78	78.136
Mixed transformations	80.47	**85.072**

TABLE III. RESULTS FROM THE TESTING PROCESS

Datasets	Accuracy ACC [%]	
	Normal validation	*Cross validation*
Original RGB images	77.65	77.59
Marked images (optic disc and blood vessels supression)	77.65	77.13
Overlapping images	84.01	79.81
Green channel images	76.77	76.76
Gauss noise images	77.68	76.37
Mixed transformations	79.92	**82.51**

As can be seen from many articles that already reviewed, all of them using Convolutional Neural Network to perform classification DR. This is because if they are using hand crafted features, it will require highly trained clinician and time consuming. In addition, all the method identifies some features such as microaneurysms, hemorrhages or exudates to detect diabetic retinopathy. There is some method to identify all the features that stated above and some method only focus on 1 or 2 features in their approach. The dataset that used on their method comes from variant provider such as Kaggle, EyePacs, etc. Then, in the preprocessing images, every technique will rescale the original input images to lower pixel such as 128 or 256 or 512 pixels that can be perform better with CNN and use data augmentation to increase the class size also to make the model immune to lighting conditions and several orientations.

After preprocessing the data, it will go through the network for classification. There are several types of network used in the above articles. For example, AlexNet, VGG-16, GoogLeNet and so on. All network mostly performs well on the classification. Next, there is an experiment that used join fine-tuned CNN to perform classification which perform a better accuracy than all others. Furthermore, some process apply a five-fold cross validation when doing classification which also gives a better accuracy.

Conclusion

To conclude from all the articles that we have referred above, it can be seen that from the first article until the last, each process has its own advantages and disadvantages. We cannot surely say that a particular method is the best method. All of the CNN approach is experimental. Meaning they can be improved further. Most of them performs quite well in terms of accuracy, sensitivity and specificity. Some perform better than the others. It also can be seen that the dataset has a huge impact on the results. Every method has their own dataset. All the dataset needed some sort of pre-processing steps. This pre-processing step is highly crucial variation in the dataset can highly impact the classification process. Some dataset contained good quality images, while some did not. This pre-processing step is required for the training phase.

So far we have seen from the reviewed articles above that, in the training phase there can be three approaches: training from scratch, use train CNN and unsupervised pre-training dataset. All the trained dataset then goes into the network layer. Each method described in each article has different number of network layers. The network layer basically filters the images to extract intricate features that needs to be classified. Different articles focus on different feature extraction. Some just focus on finding the exudates[Andonová, Pavlovičová et al. 2017]. Some focuses on finding the micro-aneurysms, exudate and hemorrhages [Pratt, Coenen et al. 2016]. Based on the feature extraction the classifier usually inspects the location and parameter of the extracted features and then classifies the image into the nearest class. In the classification phase, there is an article that use 5 -fold cross validation to increase the accuracy, specificity and sensitivity.

Based on all the result from the articles most of them achieve a good percentage of accuracy, sensitivity and specificity but they are using different datasets. The number and quality of images in the dataset varies substantially. They are also training and testing with different number of images. So, we cannot simply define one of them as the best method just by analysing the results.

References

1. *Acharya, U., Lim, C., Ng, E., Chee, C., Tamura, T.. Computer-based detection of diabetes retinopathy stages using digital fundus images.*
   P I Mech Eng H 2009;223(5):545–553.
2. Andonová, M., et al. (2017). Diabetic retinopathy screening based on CNN. 2017 International Symposium ELMAR.
3. Tennakoon, R.B., Mahapatra, D., Roy, P., Sedai, S., & Garnavi, R. (2017). Image Quality Classification for DR Screening Using Convolutional Neural Networks.
4. García, G., et al. (2017). Detection of Diabetic Retinopathy Based on a Convolutional Neural Network Using Retinal Fundus Images. Artificial Neural Networks and Machine Learning – ICANN 2017, Cham, Springer International Publishing.
5. Doshi, D., et al. (2016). Diabetic retinopathy detection using deep convolutional neural networks. 2016 International Conference on Computing, Analytics and Security Trends (CAST).
6. Ghosh, R., et al. (2017). Automatic detection and classification of diabetic retinopathy stages using CNN. 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN).
7. Li, X., et al. (2017). Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI).
8. Pratt, H., et al. (2016). "Convolutional Neural Networks for Diabetic Retinopathy." Procedia Computer Science **90**: 200-205.
9. Sun, J., et al. (2017). Retinal Image Quality Classification Using Fine-Tuned CNN. Fetal, Infant and Ophthalmic Medical Image Analysis, Cham, Springer International Publishing.
10. Tajbakhsh, N., et al. (2016). "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" IEEE Transactions on Medical Imaging **35**(5): 1299-1312.
11. Verma, K., et al. (2011). Detection and classification of diabetic retinopathy using retinal images. 2011 Annual IEEE India Conference.
12. Zhou, L., et al. (2018). "Deep multiple instance learning for automatic detection of diabetic retinopathy in retinal images." IET Image Processing **12**(4): 563-571.

.