Diabetic Retinopathy Detection through Image Analysis Using Deep Convolutional Neural Networks

Jordi de la Torre, Aida Valls & Domenec Puig
Departament d'Enginyeria Informàtica i Matemàtiques
Universitat Rovira i Virgili

Contact Information:

Email:
jordi.delatorre@gmail.com
aida.valls@urv.cat
domenec.puig@urv.cat

Abstract

Diabetic Retinopathy is one of the main causes of blindness and visual impairment for diabetic population. The detection and diagnosis of the disease is usually done with the help of retinal images taken with a mydriatic camera. In this paper we propose an automatic retina image classifier that using supervised deep learning techniques is able to classify retinal images in five standard levels of severity. In each level different irregularities appear on the image, due to micro-aneurisms, hemorrages, exudates and edemas. This problem has been approached before using traditional computer vision techniques based on manual feature extraction. Differently, we explore the use of the recent machine learning approach of deep convolutional neural networks, which has given good results in other image classification problems. From a training dataset of around 35.000 human expert classified images, different convolutional neural networks with different input size images are tested in order to find the model that performs the best over a test set of around 53000 images. Results show that it is possible to achieve a quadratic weighted kappa classification score over 0.75 not far from human expert reported scores of 0.80.

Introduction

Diabetic Retinopathy (DR) is a leading disabling chronic disease and one of the main causes of blindness and visual impairment in developed countries for diabetic patients. Studies reported that 90% of the cases can be prevented through early detection and treatment [1]. Eye screening through retinal images is used by physicians to detect the lesions related with this disease.

The DR disease is standardly classified [2] in the next fixe classes:

0. - No apparent retinopathy

1. - Mild Non-Proliferative Diabetic Retinopathy (NPDR)

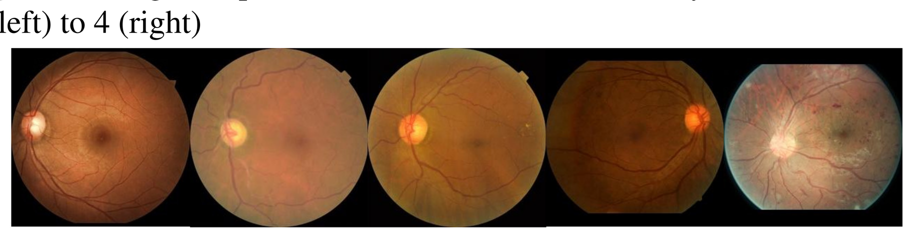
2. - Moderate NPDR

3. - Severe NPDR

4. - Proliferative DR

In this paper we propose an automated classification of retinal images into these 4 classes using Deep Learning techniques.

Figure 1: Image samples of the five different DR severity classes sorted from 0 (left) to 4 (right)



Methodology

Evaluation function

The performance of the classification model is measured using the quadratic weighted kappa (QWK) index. See equation 1.

$$\kappa = 1 - \frac{\sum_{i=1}^{C}\sum_{j=1}^{C}\omega_{i,j}O_{i,j}}{\sum_{i=1}^{C}\sum_{j=1}^{C}\omega_{i,j}E_{i,j}} \quad \text{where} \quad \omega_{i,j} = \frac{(i-j)^2}{(C-1)^2} \tag{1}$$

The interpretation of kappa values can be done using the guide of table 1.

Table 1: Table Interpretation of kappa, after Landis and Koch (1977)

Kappa	Strength of agreement
<0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very good

The Data

The training set contains a total of 35.126 high resolution images; 25.810 of class 0, 2.443 of class 1, 5.292 of class 3, 873 of class 3 and 708 of class 4. The test set contains a total of 53.576 high resolution images; 39.533 of class 0, 3.762 of class 1, 7.861 of class 2, 1.214 of class 3 and 1.206 of class 4. Notice that it is highly imbalanced.

All the images are classified by ophthalmologists according to the standard severity scale presented before [2].

From the training data we can extract the frequency of combined occurrence of the disease in both eyes of the same patient. Table 2 show this frequencies.

Table 2: Frequencies of combined occurrence of classes (left eye: rows, right eye: columns)

Eyes	C0	C1	C2	C3	C4	Sum
C0	12155	407	295	3	11	12871
C1	435	600	171	2	4	1212
C2	336	222	1998	96	50	2702
C3	3	1	87	307	27	425
C4	10	1	39	40	263	353
Sum	12939	1231	2590	448	355	17563

Training and Testing procedure

Logarithmic loss function is used for optimization. Leaky ReLU[3] is used as activation function. In all layers a batch normalization [4] is applied before the activation function. Dropout [5] (p=0.5) is performed before the final classification layer. A random initialization based in the Kaiming&He approach[6] is used for all the networks.

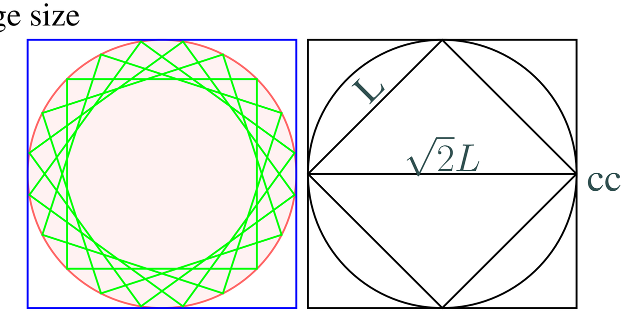
All models are optimized using stochastic gradient descent with Nesterov momentum. All the high resolution images are resized to the fixed input network size before training. For every epoch, a data augmentation technique based on cropping, rotation, mirroring and brightness & contrast correction is applied.

At test time five 72°rotated evaluations are averaged to get the class classification of every image (see figure 2).

Combining the information coming from both eyes (eq.2), we can improve the classification scores.

$$\begin{aligned} P_L &= P(Left|Right)P(Right) \\ P_R &= P(Right|Left)P(Left) \end{aligned} \tag{2}$$

Figure 2: Ensemble used at test time and geometric relationship between crop and image size



Results

In table 3 are shown the best results obtained from the use of different input sizes.

Table 3: Best classification results achieved with one eye and two eyes information

		One eye information			Two eyes information		
Layers	Input size	κ_{test}	FN	FP	κ_{test}	FN	FP
12	(3,128,128)	0.488	11.6%	11.5%	0.555	11.2%	12.9%
14	(3,256,256)	0.636	4.4%	28.7%	0.661	4.4%	28.7%
16	(3,384,384)	0.668	7.9%	14.9%	0.722	11.2%	4.0%
16	(3,512,512)	0.725	5.0%	11.9%	0.752	6.5%	7.0%

In table 4 we show the architecture of the best performing network.

Table 4: Best performing architecture

Layer	Type	Characteristics	Output Size
0	Input	3 RGB channels	(3,512,512)
1	Conv	8 fil 3x3 1,1 str, 1,1 pad	(8,512,512)
2	Conv	16 fil 3x3 1,1 str, 1,1 pad	(16,512,512)
3	Conv	16 fil 3x3 1,1 str, 1,1 pad	(16,512,512)
-	Maxpool	2,2 size, 2,2 str	(16,256,256)
4	Conv	32 fil 3x3 1,1 str, 1,1 pad	(32,256,256)
5	Conv	32 fil 3x3 1,1 str, 1,1 pad	(32,256,256)
-	Maxpool	2,2 size, 2,2 str	(32,128,128)
6	Conv	64 fil 3x3 1,1 str, 1,1 pad	(64,128,128)
7	Conv	64 fil 3x3 1,1 str, 1,1 pad	(64,128,128)
-	Maxpool	2,2 size, 2,2 str	(64,64,64)
8	Conv	128 fil 3x3 1,1 str, 1,1 pad	(128,64,64)
9	Conv	128 fil 3x3 1,1 str, 1,1 pad	(128,64,64)
-	Maxpool	2,2 size, 2,2 str	(128,32,32)
10	Conv	128 fil 3x3 1,1 str, 1,1 pad	(128,32,32)
11	Conv	128 fil 3x3 1,1 str, 1,1 pad	(128,32,32)
-	Maxpool	2,2 size, 2,2 str	(128,16,16)
12	Conv	128 fil 3x3 1,1 str, 1,1 pad	(128,16,16)
13	Conv	128 fil 3x3 1,1 str, 1,1 pad	(128,16,16)
-	Maxpool	2,2 size, 2,2 str	(128,8,8)
14	Conv	256 fil 8x8 1,1 str, 1,1 pad	(256,8,8)
15	Fully connected	256 elem	(256)
16	Softmax	256 to 5 elem	5

Conclusions

In this paper is shown that deep learning techniques are a promising technique for solving medical imaging problems like the diabetic retinopathy detection. Having enough data this method is able to perform near human level expertise achieving κ values of 0.752 not far from the κ achieved by human experts, around 0.80.

Forthcoming Research

Future work will be centered on testing higher resolution input images, newer schemes such as residual networks, the use of alternative cost functions that encode the prior information of the ordering of the classes and more elaborated methods for combining the information coming from both eyes.

References

[1] J. Torrents-Barrena, J. Melendez, A. Valls, P. Romero, and D. Puig. Screening for diabetic retinopathy through retinal colour fundus images using convolutional neural networks. In *Artificial Intelligence Research and Development - Proceedings of the 18th International Conference of the Catalan Association for Artificial Intelligence. IOSPress*, pages 259–262, 2015.

[2] C.P. Wilkinson, F.L. Ferris 3rd, R.E. Klein, P.P. Lee, C.D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, and J.T. Verdaguer. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*, 110(9):1677–82, 2003.

[3] G. E. Dahl, T. N. Sainath, and G. E. Hinton. Improving deep neural networks for LVCSR using rectified linear units and dropout. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8609–8613, 2013.

[4] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Deep Learning Symposium (NIPS 2015)*, 2015.

[5] P. Baldi and P. Sadowski. Understanding dropout. In *Advances in Neural Information Processing Systems (NIPS)*, volume 26, pages 2814–2822, 2013.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015.

Acknowledgements

This work is supported by the URV grant 2014PFR-URV-B2-60 and the Spanish research projects PI15/01150 and PI12/01535 (Instituto de Salud Carlos III). The authors would like to thank to the California Healthcare Foundation and EyePACS for providing the images used in this study.