Automatic Diabetic Retinopathy Classification

María A. Bravo and Pablo A. Arbeláez

Universidad de los Andes, Bogotá, Colombia

ABSTRACT

Diabetic retinopathy (DR) is a disease in which the retina is damaged due to augmentation in the blood pressure of small vessels. DR is the major cause of blindness for diabetics. It has been shown that early diagnosis can play a major role in prevention of visual loss and blindness. This work proposes a computer based approach for the detection of DR in back-of-the-eye images based on the use of convolutional neural networks (CNNs). Our CNN uses deep architectures to classify Back-of-the-eye Retinal Photographs (BRP) in 5 stages of DR. Our method combines several preprocessing images of BRP to obtain an ACA score of 50.5%. Furthermore, we explore subproblems by training a larger CNN of our main classification task.

Keywords: Diabetic retinopathy, convolutional neural networks, retinal vessel segmentation, Back-of-the-eye Retinal Photography.

1. INTRODUCTION

Diabetic retinopathy (DR), or diabetic eye disease, affects most patients with diabetes type 1 and some with diabetes type 2. This disease is mainly caused by diabetic microangiopathy, a damage to the small blood vessels that oxygenate the retina in the posterior part of the eye of a diabetic patient. The damage in the eye is due to a metabolic disorder that increases the levels of glucose in a person's blood, causing the person to present high levels of blood pressure, which in turn affect the circulatory system of the retina and the light sensitive lining at the back of the eye. This condition, known as DR, can lead to leaking blood and other fluids that cause swelling of retinal tissue and clouding of vision. Patients with DR develop dark spots in the field of vision, blurred vision, difficulty to see at night and in some cases, total blindness.[1]

DR is classified into two main stages. In the non-proliferative stage of the disease, symptoms are mild or non-existent. During this stage: (1) the blood vessels in the retina are weakened due to the accumulation of glucose and the increase in blood pressure, causing tiny bulges called microaneurysms that protrude from the walls of vessels; (2) exudates and edemas become present because of the increase in the permeability of vessels; and (3) some hemorrhages can be detected. Figure 1 shows a back-of-the eye image exhibiting these symptoms. The second proliferative stage is characterized by a neoformation of weak and fragile vessels in the retina and the crossing between veins and arteries causing wrong blood flow, combined with the symptoms already present during the non-proliferative stages. The need of oxygen causes new vessel formation to irrigate the retina.[2] In some severe cases, DR can lead to diabetic macular oedema when the fluid reaches the macula. Macular oedema can be detected by the presence of retinal thickening and swelling within one disc diameter of the center of the fovea.[3,4]

DR is a serious public health problem, being one of the main causes of blindness worldwide and its prevalence is set to continue rising. Today, it is estimated that diabetes affects 2 to 6% of the population of industrialized countries and 16% of adults over the age of 65.[5] Of the global diabetic population, 40-50% could be afflicted by retinopathy.[6] An estimate of 50–65 new cases of blindness per 100,000 people are diagnosed every year because of DR.[3]

Traditionally, DR has been categorized as incurable and unpreventable. Nowadays, new technology like *Argón's* laser and *pars plana* has opened the possibility of treatment, but early detection is essential for effectiveness.[1] Early detection of sight-threatening retinopathy allows laser therapy to be performed to prevent or delay visual loss and may be used to encourage improvement in diabetic control. Unfortunately, there are almost no visual symptoms in the early stages of DR,

Further author information, E-mail: ma.bravo641@uniandes.edu.co

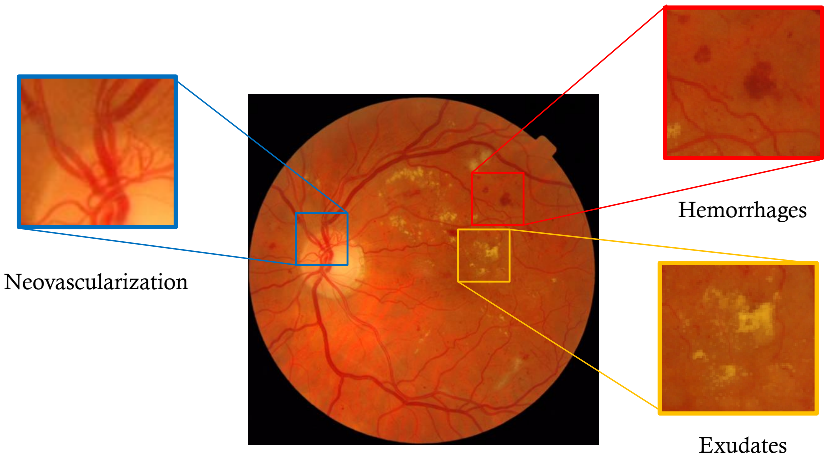

Figure 1: DR visible symptoms. This is an eye fundus image with a high degree of DR, it presents the different characteristic symptoms of this disease: neovascularization (top left), hemorrhages (top right) and exudates (bottom right).

making it difficult to give a diagnosis in time. There exist several diagnostic methods for this disease such as fluorescein angiography, optical coherence tomography, and the most common and less expensive method used, the Back-of-the-eye Retinal Photography (BRF) or Digital Fondus Photography. During fluorescein angiography, a yellow fluorescein dye is introduced into the patient's bloodstream and then several retinal photographs are taken. For optical coherence tomography, light waves are used to capture images of tissues inside the eye that allow to measure the amount of retinal swelling. To take a BRF, the pupil is dilated and then a picture is taken with a high-resolution camera.[3]

At the present time, the detection of DR is a time-consuming and manual process that requires a trained clinician to examine and evaluate the diagnostic images. The diagnosis can take several days depending on the doctor's availability and the number of patients. Experience has proven that there exists a need for an automatic method of DR screening.[7] Recently, image processing and analysis has progressed detecting DR, but this is still a new field, with room for improvement in both processing time and accuracy.

2. RELATED WORK

Initially, diabetic retinopathy detection methods worked extracting handcrafted features of interest in the image. The first approaches in this field aimed at detecting diabetic retinopathy with respect to the textural features of retinal images. For example, Yun *et al.*[8] proposed a method based on detecting and analyzing microaneurysms and hemorrhages. Nayak *et al.*[9] used a combination of hard exudates area and blood vessels by morphological techniques. Reza *et al.*[10] developed a rule-based algorithm based on hard exudates and cotton wool spots. These pathological regions were segmented using image processing techniques such as thresholding, morphological reconstruction, and boundary tracing. Finally, recent works[4] in this topic have centered the attention in finding exudates localization and in combining them to determine the correct stage of the disease.

In 2015, thanks to the California Healthcare Foundation that created a challenge with an available dataset,[11] the development of automatic algorithms to classify and detect different degrees of DR grew rapidly. All the groups that finished the challenge competition on top used deep learning and CNNs. The images were labeled according to the degree of DR present, from 0 (healthy) to 4 (the highest degree of DR).

Ben Graham, one of the participants of the challenge, obtained the first place[12] under the name of ”Min Pooling" with a score of 0.8496. This method involved SparseConvNet, a convolutional neural network whose input is the whole image. To train the network, it was necessary to do data augmentation: scaling, skewing and rotating the images. Min Pooling rescaled the images to guarantee eyes of the same radius (300 or 500 pixels), subtracted the local average color, and clipped the image to 90% of its size to remove boundary effects. The final competition model combined three different networks: two convolutional networks, using fractional max-pooling[13] involving layers of spatial pooling with fractional parameters, and the other using very deep convolutional neural networks.[14] To give the final class, Min Pooling trained a random forest using the predicted scores and other correlated information such as the score of the other person's eye, the variance of the original image and the variance of the preprocessed image. In comparison, in our method we train a single CNN and use smaller images of 225 and 299 pixels for our algorithm.

The team "o_O" earned second place with 0.8448 using a total of three neural network architectures. For this method, images were resized to 768x768 and augmented in number by applying translation, stretching, rotation, flipping and color augmentation. The algorithm used two types of neural networks with raw images as their input. For each network, three sets of weights were calculated, and the authors combined the scores of both eyes, right and left eyes. The final score was the average of the six outputs. To train this network they used a regression loss, so their score was not a probability vector for each category but a single number, usually between 0 and 4, that was rounded to obtain the predicted classification. Since the database was greatly unbalanced, images were first sampled so that each class were represented equally and, as training progressed, they decreased sampling of rare classes. In order to train the final weights of the CNN, the algorithm was first trained with smaller networks, and these weights were used to initialize the more complicated networks. Instead, our single method comparison uses a single CNN and explores considering the problem as a classification task with a softmax classification layer and subproblems considering fewer classes.

The third-place group was the "Reformed Gamblers" with a score of 0.8394. This method combined the results of nine different convolutional networks, including the work of Benjamin Graham[13] and Karen Simonyan and Andrew Zisserman.[14] For data augmentation, images were randomly cropped to 85-95% of the original size, horizontally flipped, rotated between 0 and 360 degrees, and finally scaled to the desired model input size (between 385 and 767, depending on the network). For non-linearity of the networks, the algorithm used leaky rectifier unit. To facilitate training the algorithm started with pretrained weights and with very small learning rates, which were then increased after several epochs. The output of each network was a number between 0 and 4 and the algorithm used a simple linear model to combine the output of both eyes. Unlike "o_O", for the final classification, the authors did not simply round the output but performed a grid search of possible cutoffs to maximize the kappa score.[15]

All three approaches used information from both eyes to make the predictions. Another important point was the size of the input image. Both o_O and Reformed Gamblers tried inputs of different sizes and agreed that there was a significant improvement as you increase the size which in turn increase the processing time. This increase in the score might have being caused by the microaneurysms that are only visible in large scale images. Our method uses some of the experience given by these groups, such as using pretrained weights, preprocessing images and augmentate data to balance the classes. Compared to them, we do not use images with such big sizes and do not consider a relation between both eyes since in many cases different eyes of the same patient belong to different classes. We used a single network and explore different subproblems that are interesting for clinical diagnosis. In previous work,[16] we have tried to classify BRF into degrees of DR disease using feature extraction followed by traditional classifiers such as support vector machines, random forests, nearest neighbors, and convolutional neural networks (CNN) resulting in 43.3%, 39.1%, 39.1%, and 41.6% of average accuracy respectively. This work expands our efforts to solve the early DR diagnosis.

3. CNNS FOR DIABETIC RETINOPATHY IMAGE CLASSIFICATION

3.1 Dataset

The original dataset used in this work was obtained from Kaggle [*], where the California Healthcare Foundation issued a challenge to create an automatic program to detect diabetic retinopathy. This challenge comprises a dataset of high-resolution retina images taken under a variety of imaging conditions with 35126 training images. For annotations, the dataset has the images classified in five groups (0, 1, 2, 3, 4), depending on the disease's severity. Level 0 contains images with no signs of retinopathy, whereas level 4 images show advanced symptoms. The scale of each of the degrees of

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

DR corresponds to no DR, mild, moderate, severe and proliferative, form 0 to 4 respectively. The training set is highly unbalanced as shown in Table 1 and Figure 2, with 25810 level 0 images, 2443 for level 1, 5292 of level 2, 873 in level 3, and for 708 in level 4.[11] In order to allow training and optimization of large capacity models, we split the data into training and validation. From this database, we reserved a balanced set of 1560 images for validation and the rest were used for training.

Table 1: Unbalanced dataset class distribution. Class 0 is almost 73% of all images and class 4 is around 2%.

Class	Name	Number 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%

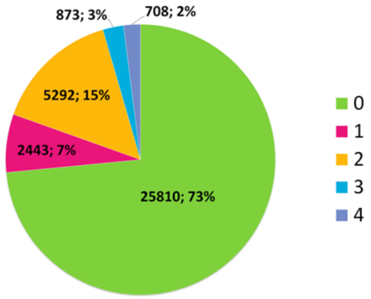

Figure 2: Unbalanced dataset class distribution. Class 0 is almost 73% of all images and class 4 is around 2%.

It is important to note that the database contains challenging images with poor quality (poor lighting and unfocused) which makes it difficult for any algorithm to classify them correctly in the stages of the disease.

3.2 Deep Diabetic Retinal Analysis

We approach the DR classification problem by using CNNs. We start from the VGG16[14] network, originally designed for large-scale natural image classification into 1000 different classes. This architecture consists of thirteen convolutional layers coupled with Rectified Linear Unit (ReLU) activations, five max pooling layers that split the network into six scales, and three Fully Connected (FC) layers (as in Fig. 3). For our task, the last fully connected layer at the end of the network was changed to have a 5-response vector that corresponds to each of the degrees of the disease. Our architecture included two dropout layers, which disable several connections in some iterations of the training process, allowing the model to reduce overfitting. We started our network with the pretrained original VGG16 weights and during the training process all convolutional weights, except for Conv5_3, were fixed. All the FC layers were also finetuned during training. For the last layer, we used a softmax classifier, which calculates the probability for an image of being in each of the five classes.

To train the network, we used a multinomial logistic loss function and implemented it in the same layer as the softmax classifier, Equation (1) shows the definition, where $c \in \{0, ..., 4\}$ are the classes. The combination of the softmax layer and the log-loss is useful for numerical stability.[17] Equation (1) generates a probability vector for each image indicating the class in which it is more likely to be part.

$$l(\mathbf{x}, c) = -log\frac{e^{x_c}}{\sum_{k=0}^{4} e^{x_k}} = -x_c + log\sum_{k=0}^{4} e^{x_k}. \tag{1}$$

For training the networks, we used the publicly available *Caffe*[18] framework and a NVIDIA TITAN X GPU in which the average forward execution time for an image was of 100μs and the training time for an image to pass through the net was

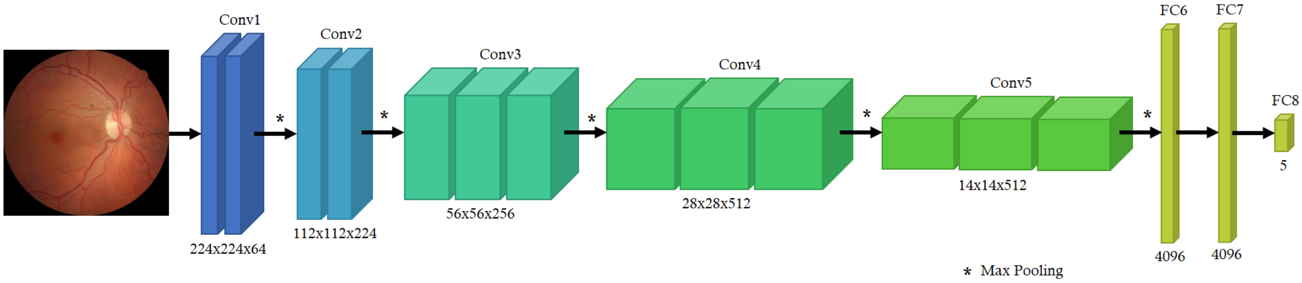

Figure 3: VGG16 fine-tuning architecture used as base network for our method.

around 1ms, using the VGG-16 architecture.

In order to evaluate a deeper architecture, we used Inception-V4[19] for training. Figure 4 shows the architecture of the network. Inception-4 is a bigger network that has more connections between layers; it is deeper and wider, and uses parallel processing blocks in its layers to improve its performance. We trained this network using *Tensorflow* with a batch size of 32, an initial learning weight of 0.001 starting from a pretrained model in IMAGENET. We used a weight decay of 0.00004 and trained for 20 epochs. For our task, we removed the last fully connected layer at the end of the network and changed it for a 5-response vector that corresponds to each of the degrees of the disease. For this method we trained all of the weights of the CNN and made use of the complex structure.

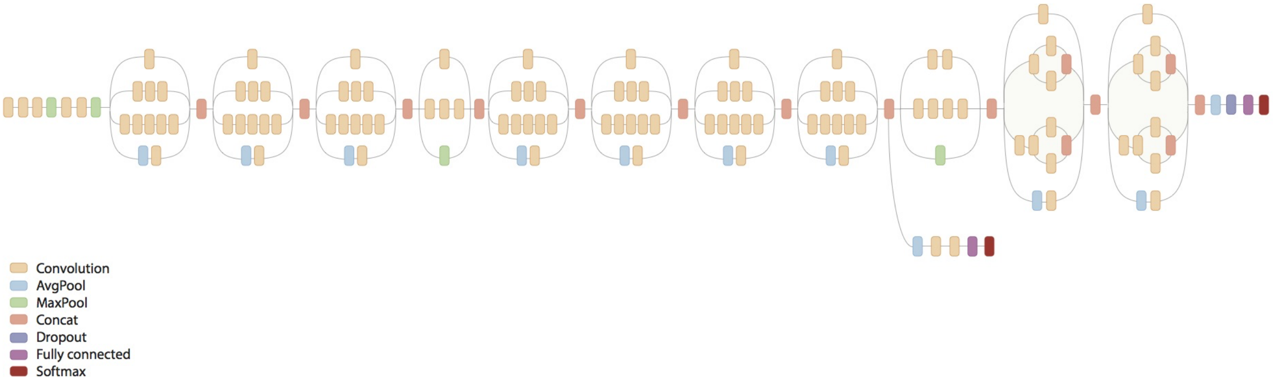

Figure 4: Inception-4 architecture used as base network for our method. Image taken from https://github.com/tensorflow/models/tree/master/inception. This image is better seen in color.

3.3 Preprocessing

Images from the dataset varied in size between 2500x2000 and 4700x3100 pixels. Due to the number of images, we decided to reduce their size to facilitate computational time. We first cropped the images to remove the neutral background and resized them to 500x500 pixels. To determine which was the best input for the CNN, we tried five different preprocessing images shown in Fig. 5:

1. Circle RGB images (Circle): For this set, we used the whole image except for the neutral background and resized it to 224x224.
2. Square images (Square): For this set, we used the largest square image inscribed in the eye and resized it to 224x224.

3. Circle color adjust images (Color centered): For this set, we used the same circle images of the first set and subtracted the local average color using a Gaussian filter. The intensities for each channel were centered at 127.
4. Gray scale images (Gray Scale): For this set, we used the same circle images in gray scale. This experiment was done to determine the importance of colors for the classification task.

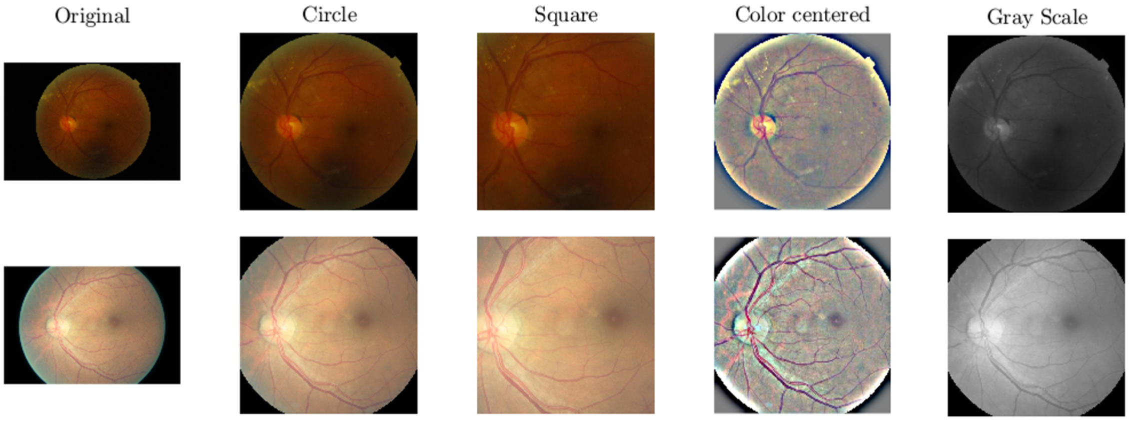

Figure 5: Image preprocessing. Columns correspond to: (1) original image; (2) circle cropped image; (3) square cropped image; (4) color adjust image using a Gaussian filter; and (5) gray scale image.

4. EXPERIMENTAL VALIDATION

4.1 Evaluation Methodology

To evaluate performance of the classification methods, we used a confusion matrix (CM). We counted how many images of each class were classified in each category and calculated the mean of the diagonal of the normalize CM, the Average Classification Accuracy (ACA). In the DR classification problem, we have five categories; therefore, an ACA above 0.2, which is the score of a random guess, shows the algorithm is better than chance.

4.2 Data augmentation

For training and testing the network it is better to have a balanced set and since the dataset was not balanced, we decided to split the database as follows: 1560 images for validation with about 312 images per class; and 33566 images for training (with 25496, 2132, 4980, 561, 397 images per class, corresponding to classes 0 to 4). This split was made making sure both eyes of each subject in the database were in the same set. To balance the training set and to increase the number of images for training, we did data augmentation using rotations from 0 to 360 degrees, zooming in and out from 0 to 20 pixels, and doing horizontal and vertical flips. This augmentation was done randomly for all images in the training set and, for the classes 1, 2, 3, and 4, images were repeated with different augmentation parameters. The resulting balanced training data set had 127480 images.

4.3 Selection of image preprocessing

To choose the best preprocessing image for the classification task, we trained four different CNNs with each of the preprocessing methods of Sec. 3.3 and the same augmentation. Results of training are shown on Table 2. For training these CNNs, we used the pretrained weights of VGG-16 from ILSVRC-2014 and finetuned the fully connected layers with an initialization learning rate of 0.01 that changed every 5 epochs by a power of 10. We trained for 700 iterations (aprox. 14 epochs), with a batch size of 115, using stochastic gradient descent and a momentum of 0.9. We observed similar behaviors for the *circle rgb, color centered*, and *gray scale* sets. Around the fifth epoch, the networks stopped generalizing and started over-fitting to the training set. Best ACA validation scores are recorded in Table 2, best score was 0.483 obtained by using the *color* set. Unlike sets with complete eye images the *square* set started over-fitting around the second epoch and achieved a best ACA score of 0.453 in the 3rd epoch.

Table 2: Preprocessing images experiment. Evaluation of the same CNN VGG-16 trained with the same initialization and parameters but different preprocessing images. Scores correspond to the validation set.

Preprocessing	ACA
Circle RGB	46.3%
Square	45.3%
Color centered	**48.3%**
Gray scale	48.1 %

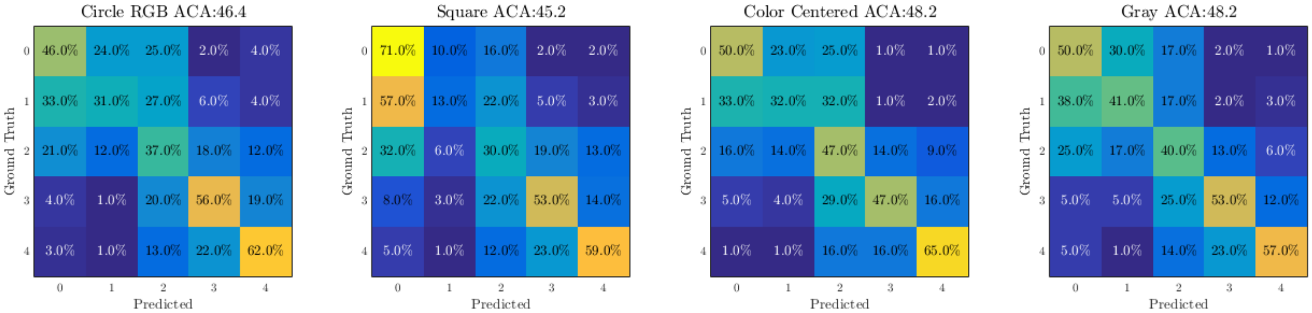

Figure 6: Confusion matrices (CM) of best epoch image preprocessing methods. Each CM corresponds to one preprocessing image experiment.

.

We made a further analysis on the classification scores of the different sets based on their confusion matrices shown in Fig. 6. Results showed that the *color centered* and *gray scale* sets presented a more uniform correct classification score in comparison to the other sets. The *square* image set obtained a very unbalanced classification since most of the predictions were centered in classes 0, 3 and 4. A similar behavior happened for the vessel segmentation images, except for class 2, since it obtained a classification score higher than the other sets. With the *color centered* set, we obtained a more uniform result than using the raw *circle rgb* set, but in both cases, class 1 was the most misclassified by the models. All experiments presented a clear confusion between classes 0, 1 and 2. The confusion between the first three classes was expected since first stages of DR are almost indistinguishable, symptoms are minimum, and differences between stages are not strictly stated. This fact reflects the difficulty of classifying correctly back-of-the-eye images in DR degrees.

We found a clear difference between the two crops of the images; in both experiments, the results were lower for square images compared with circle images. With these results, we concluded that for some images relevant information, such as the optic disc, hemorrhages, and exudates, were localized on the border of the BRF. Since square images cropped part of the border of the eye, they missed important information for DR detection. For this reason, in the rest of the experiments, we only considered the circular rgb, and color centered sets.

4.4 Inception-4 architecture and different pretrained weights

To explore different subproblems of the main task, we trained a deeper network considering first several binary problems: sick (1,2,3,4) vs healthy (0); low (0,1) vs high (2,3,4); and class 0 vs class 1. Then we considered 3-class problems: 0, 1, higher (2,3,4); and 2, 3, and 4. Finally using the best pretrained weights, we retrain the network for the 5-class classification problem. The same images and sets where used for all experiments.

5. RESULTS AND DISCUSSION

5.1 Fusion of Classifiers

As for our final classification method, we combined the scores of the best trained neural networks (the ones trained with circular rgb, gray scale, and color centered sets) using the VGG architecture. We chose, for each image, the class with higher average score. The resulting Confusion Matrix is shown in Figure 7. We observed classes 1 and 2 were extremely difficult to classify correctly, instead class 4 was the easiest. Our final ACA score was 0.505. This combined method yielded the highest score of all because it used multiple sources of information and different networks. The use of different networks trained with different parameters, made the final method more robust to outliers and capable of further generalizing if more information is considered. It is clear from Figure 7 that classes 1 and 2 are the most difficult ones to differentiate and class 4 is the easiest to detect.

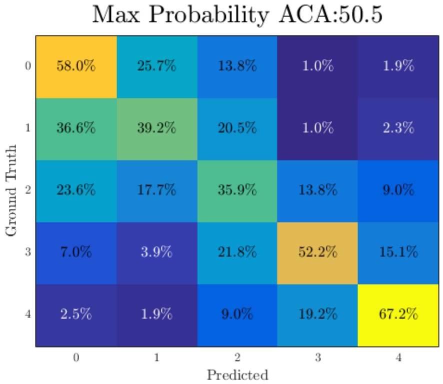

Figure 7: Final classification Confusion Matrix for the fusion of classifiers. Horizontal axes correspond to the Predicted classes and vertical axis to the ground truth.

.

5.2 Inception-4 results

In order to gain insights on the problem and to explore its difficulty, we extended our experiments changing the number of classes by grouping the images that belong to adjacent classes. In these experiments, we used five different label sets of the training dataset and three binary problems: sick (classes 1,2,3,4) vs healthy (class 0); low (0,1) vs high (2,3,4); and class 0 vs class 1. Two 3-class problems: classes 0, 1, and higher (2,3,4); and classes 2, 3, and 4. Finally, we also trained Inception-V4 for the original 5-class problem. We report the results in Table 3. To analyze these results, we should take into account the corresponding scores that we would have if we considered a classification just by chance: for a 2-class problem is 50%, for a 3-class problem is 33.3% and for a 5-class problem is 20%. With the 2-class problem, it was evident that separating class 0 from the rest was not as effective as separating 0 and 1 from 2, 3, and 4. We can support the fact that classes 0 and 1 are much alike. Also, when classifying only classes 0 and 1, the scores obtained where lower. The 3-way classification scores were proportionally better than the binary scores. This time, the model that separated classes 0 and 1 achieved a higher score than the one that classified classes 3, 4 and 5. These results show a strong relation and similitude between DR advanced stages (2, 3 and 4). Finally, our 5-class classification model scored close to our best classification accuracy. Additionally it is important to note that the great majority of scores were slightly higher using the Color preprocessing set compared to the original image.

Table 3: Inception-4 ACA results: Using different problems of 2, 3 and 5 classes.

Problem variation	RGB No preprocessing	Color Preprocessing
binary sick (1,2,3,4) vs healthy (0)	64.63%	72.41%
binary low (0,1) vs high (2,3,4)	83.71%	86.25%
binary class 0 vs class 1	62.71%	62.86%
3 classes: 0, 1, higher (2,3,4)	77.14%	72.18%
3 classes: 2, 3, and 4	68.75%	69.75%
5 classes	44.62%	45.00%

5.3 Qualitative Results

Figure 8 shows examples of classification results. Top left shows two visually similar images; however, the left-hand side image has DR degree 3 and the right-hand side 1. For both cases, our algorithm classified them in degree 1. In our second example, the top right pair of images in Figure 8, both images are healthy but our algorithm classified the left-hand side as degree 2, this prediction might be due to the illumination artifacts and the small brown spots it has. For this image, the second higher probability score was class 0 and the scores between the first and second place differ less than 0.01. The third example, the bottom left pair of images in Figure 8, shows two images in which the left-hand side image looks more seriously ill than the right-hand side image since it exhibits more hemorrhages and exudates. Actually, the left-hand side image is class 3 and the right-hand side is 4. Both images were predicted in class 4. Our final example, the bottom right pair of images in Figure 8, shows two very similar images which might have similar classes. The fact is that the left-hand side image class is 0 but it was predicted class 1 and the right-hand side image has ground truth and predicted class 3.

These few examples show the complexity of this problem. Even when images seem to have the same degree of DR or a consecutive order, the real classification of the images is not easy to guess just by sight by non-specialists, and even sometimes by specialists, but our algorithm predicts good scores for each example.

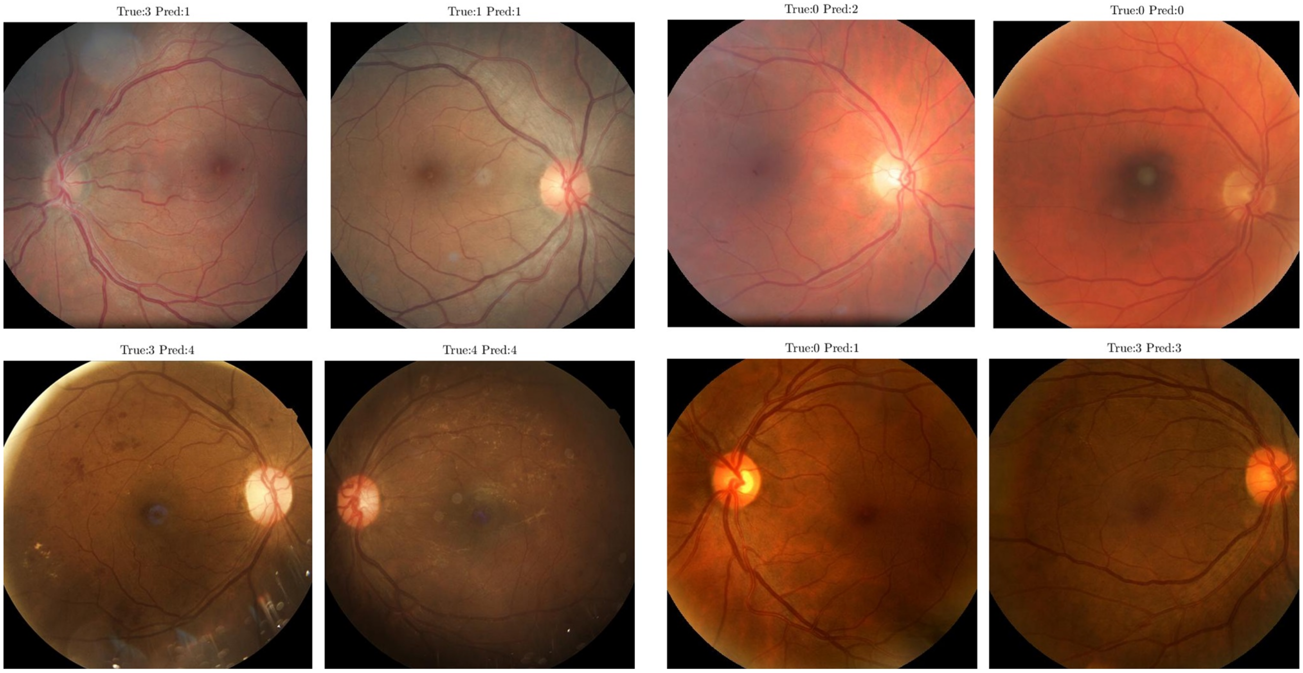

Figure 8: Classification results. Top left pair of images: left side image true class 3 predicted 1, right side image true and predicted class 1. Top right pair of images: left side image true class 0 predicted 2, right side image true and predicted class 0. Bottom left pair of images: left side image true class 3 predicted 4, right side image true and predicted class 4. Bottom right pair of images: left side image true class 0 predicted 1, right side image true and predicted class 3. (Best viewed in color.)

.

5.4 Conclusions

This work explores different approaches for a relevant medical disease detection problem. Our method exploits the benefits of pretrained state-of-the-art architectures for large-scale classification problems and uses this knowledge to impact the medical field. We seek to understand the difficulty of Diabetic Retinopathy detection with one of the most recent networks. With Inception-V4, we reinforced and confirmed the fact that consecutive degrees of DR have minimum visual differences. Furthermore, we present an algorithm capable of approximating DR classification with a 50.5% of accuracy for a 5-class problem.

ACKNOWLEDGMENTS

The authors gratefully acknowledge NVIDIA Corporation for donating the GPUs used in this project.

REFERENCES

[1] Diabética, Á. R. R., "Pontificia universidad católica de chile," *Boletín de la escuela de medicina* **31**(3) (2006).

[2] Fong, D. S., Aiello, L., Gardner, T. W., King, G. L., Blankenship, G., Cavallerano, J. D., Ferris, F. L., and Klein, R., "Retinopathy in diabetes," *Diabetes care* **27**(suppl 1), s84–s87 (2004).

[3] Olson, J., Strachan, F., Hipwell, J., Goatman, K., McHardy, K., Forrester, J., and Sharp, P., "A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy," *Diabetic medicine* **20**(7), 528–534 (2003).

[4] Perdomo, O., Otalora, S., Rodríguez, F., Arevalo, J., and González, F. A., "A novel machine learning model based on exudate localization to detect diabetic macular edema," (2016).

[5] Esteve, M. C., Fernández, M., Goday, A., and Cano, J., "Revisión de las complicaciones crónicas de la diabetes mellitus en españa," *Jano: Medicina y humanidades* (1644), 27 (2007).

[6] Sández, J., Fernández Vigo, J., Díaz, A., Rodríguez, M., Jesús, M., and Barrios, J., "Prevalencia de retinopatía diabética en una población diabética no seleccionada.," *Archivos de la Sociedad Española de Oftalmología* **59**(3), 277–284 (1990).

[7] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA* **316**(22), 2402–2410 (2016).

[8] Yun, W. L., Acharya, U. R., Venkatesh, Y., Chee, C., Min, L. C., and Ng, E., "Identification of different stages of diabetic retinopathy using retinal optical images," *Information Sciences* **178**(1), 106–121 (2008).

[9] Nayak, J., Bhat, P. S., Acharya, R., Lim, C., and Kagathi, M., "Automated identification of diabetic retinopathy stages using digital fundus images," *Journal of medical systems* **32**(2), 107–115 (2008).

[10] Reza, A. W. and Eswaran, C., "A decision support system for automatic screening of non-proliferative diabetic retinopathy," *Journal of medical systems* **35**(1), 17–24 (2011).

[11] Foundation, C. H., "Diabetic retinopathy detection," (September 2015).

[12] Graham, B., "Kaggle diabetic retinopathy detection competition report," (August 6 2015).

[13] Graham, B., "Fractional max-pooling," *arXiv preprint arXiv:1412.6071* (2014).

[14] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).

[15] Jun Xu, John Dunavent, R. K., "Summary of our solution to the kaggle diabetic retinopathy detection competition," (2015).

[16] Bravo, M. A., "Learning classification methods and applications to medical images," Thesis of Mathematics, Universidad de los Andes, Bogotá, Colombia (2016).

[17] Vedaldi, A. and Lenc, K., "Matconvnet: Convolutional neural networks for matlab," in [*Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*], 689–692, ACM (2015).

[18] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T., "Caffe: Convolutional architecture for fast feature embedding," in [*Proceedings of the 22nd ACM international conference on Multimedia*], 675–678, ACM (2014).

[19] Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A., "Inception-v4, inception-resnet and the impact of residual connections on learning," *arXiv preprint arXiv:1602.07261* (2016).