Glaucoma Diagnosis from Eye Fundus Images Based on Deep Morphometric Feature Estimation

Oscar Perdomo[1]*, Vincent Andrearczyk[2], Fabrice Meriaudeau[3], Henning Müller[2], and Fabio A. González[1]

[1]MindLab Research Group, Universidad Nacional de Colombia, *ojperdomoc@unal.edu.co, [2] University of Applied Sciences Western Switzerland (HES-SO), Switzerland, [3] Universiti Teknologi PETRONAS, Malaysia

Abstract. Glaucoma is an ophthalmic disease related to damage in the optic nerve and it is without symptoms in its early stages. Left untreated, it can lead to vision limitation and blindness. Eye fundus images have been widely accepted by medical personnel to examine the morphology and texture of the optic nerve head and the physiologic cup but glaucoma diagnosis is still subjective and without clear consensus among experts. This paper presents a multi-stage deep learning model for glaucoma diagnosis based on a curriculum learning strategy. In curriculum learning, a model is sequentially trained to solve incrementally difficult tasks. Our proposed model includes the following stages: segmentation of the optic disc and physiological cup, prediction of morphometric features from segmentations, and prediction of disease level (healthy, suspicious and glaucoma). The experimental evaluation shows that our proposed method outperforms conventional convolutional deep learning models from the state of the art reported on the RIM-ONE-v1 and DRISHTI-GS1 datasets with an accuracy of 89.4% and an AUC of 0.82 respectively.

Keywords: Deep convolutional neural network, curriculum learning, morphometric features, glaucoma diagnosis, eye fundus images

1 Introduction

Glaucoma is one of the leading causes of vision loss and blindness worldwide [1]. It is defined as an increment of intraocular pressure producing morphological changes in the optic disc (OD) and physiological cup (PC) affecting the ability of the optic nerve to transmit images to the brain [2]. The main problem with glaucoma is associated with a delayed diagnosis causing an irreversible damage to the eye [1,2]. The examination of the optic disc, physiological cup and neuroretinal rim structures is important for an early detection and proper treatment [1].

The ocular tonometry or measurement of intraocular pressure does not quantify the damage or glaucoma progression [3]. Thus, a complete ophthalmoscopy examination through an eye fundus image is widely used to grade and monitor

the disease [3,4]. Additionally, an accurate and objective diagnosis is required to avoid the minimal damage to the eye structure [4]. Thus, the design of computer-aided diagnosis models for automatic disease assessment is important to improve the glaucoma detection and minimize the subjectivity in the diagnosis.

The cup-to-disc ratio (CDR) is the most typical morphometric feature used in the diagnosis of glaucoma. However, locating and segmenting the OD or optic nerve head (ONH) and the physiological cup are not easy tasks. Septiarini et al. proposed an automatic glaucoma detection method extracting statistical features from the intensity in ONH: the mean, smoothness and 3rd moment, and using a k-nearest-neighbor algorithm as a classifier [5]. Pardha et al. reported a region-based active contour model using multiple image channels and gray level properties for optic disc and cup segmentations [4].

Deep learning models, such as Deep Convolutional Neural Networks (DCNNs) have been applied with success to different medical image analysis tasks and, in particular, to automatically discriminate between glaucoma and non-glaucoma patterns in eye fundus images. Al-Bander et al. presented an 8-layer CNN model to automatically extract features of the optic disc from the raw images and a linear Support Vector Machine (SVM) classifier to classify the images into normal or glaucoma subjects [6]. On the other hand, Chen et al. reported a deep learning model that contains four convolutional layers and two fully-connected layers, where a dropout layer and data augmentation strategies are used to improve the performance of glaucoma diagnosis [7]. In addition to this, Orlando et al. fine-tuned two deep learning approaches: OverFeat and VGG pre-trained from non-medical data for automated glaucoma detection on the DRISHTI-GS1 dataset [8]. Sevastopolsky reported a DCNN for automatic OD and PC segmentations, using a modification of the U-Net CNN tested on the DRIONS-DB, RIM-ONE v.3 and DRISHTI-GS1 database [9]. Finally, Abbas presented an unsupervised CNN architecture to extract the features and used a deep belief network model to select the most discriminative deep features with a softmax linear classifier to differentiate between glaucoma and non-glaucoma retinal fundus image [10]. Despite the results obtained, these studies have only been tested on a binary classification task, as many images are not clear cases but in a continuum between healthy and glaucoma.

This paper presents a novel model for automatic analysis of eye fundus images to support glaucoma diagnosis. The model is based on an end-to-end deep convolutional neural network. Deep convolutional neural networks have been highly successful in solving several image analysis tasks. However, they require a large number of labeled samples for training, which is not necessarily the case when dealing with medical images. To mitigate this problem, we devised a curriculum learning strategy that trains the network in stages that solve incrementally more complex tasks [11]. The division of the problem into subtasks allows to better train the different network modules even with a small number of samples. The sequence of tasks is motivated by the current practice for glaucoma diagnosis from eye fundus images by specialists who use morphometric measures estimated from the optical disc and physiological cup segmentations. Thus, the proposed

deep learning method is composed of three stages: OD and PC segmentations, morphometric feature estimation and glaucoma detection, which are sequentially trained using a curriculum learning strategy.

The remainder of this paper is organized as follows: in Section 2, we give a detailed description of the proposed method including the DCNN used for automatic segmentation, the DCNN for extraction of Morphometric Features (MF), and the multilayer perceptron neural network used for glaucoma diagnosis. In Section 3 we define the experimental setup used to split the dataset and the baseline methods to evaluate our proposed method. The results are reported for the three tasks in Section 4. Finally, Section 5 discusses the results, presents the conclusions and future work.

2 Methods

Figure 1 shows the architecture of the deep neural network model for automatic analysis of eye fundus images to support glaucoma diagnosis. The model is organized in three consecutive stages that are sequentially trained using a curriculum learning approach, i.e. at each stage the training process focuses on different learning goals. This learning strategy regularizes the optimization process to converge faster, guiding the search towards better local minima [11]. The network stages were designed following a process analogous to the one followed by experts. The first stage performs the segmentation of the OD and PC using a 15-layer DCCN. The second stage uses as input the two segmentations generated by the first stage, stacking a third image mask corresponding to the union of the OD and PC segmentations to create a 3D-binary mask, which are fed to a 12-layers DCNN. The goal of this stage is to calculate different morphometric features, which are generally used by experts to diagnose glaucoma. Finally, the third stage applies a multilayer neural network to produce the final prediction that classifies the input image into three possible classes: normal, suspicious or glaucoma. The following subsections discuss the details of the three stages.

2.1 DCNN for Automatic Segmentation of Optic Disc and Physiological Cup

The first stage of the model corresponds to a DCNN that receives as input an RGB eye fundus image and calculates a segmentation of the OD and the PC. The DCNN is based on a deep retinal image understanding (DRIU) model using the last four sub-blocks called coarse feature maps [12], but with two additional convolutional layers. The DRIU model contains 13 convolutional layers with different filters sizes and 4 max-pooling layers [12]. The DRIU model is initialized with VGG weights pretrained on ImageNet. It is fine-tuned for 10,000 epochs with a learning rate of $1e-6$, which is gradually decreased as the training process proceeds. A real time data augmentation strategy is implemented to grow the training data. Class weights of 0.1 and 0.9 for background and foreground respectively are used to handle the imbalance of the number of background pixels

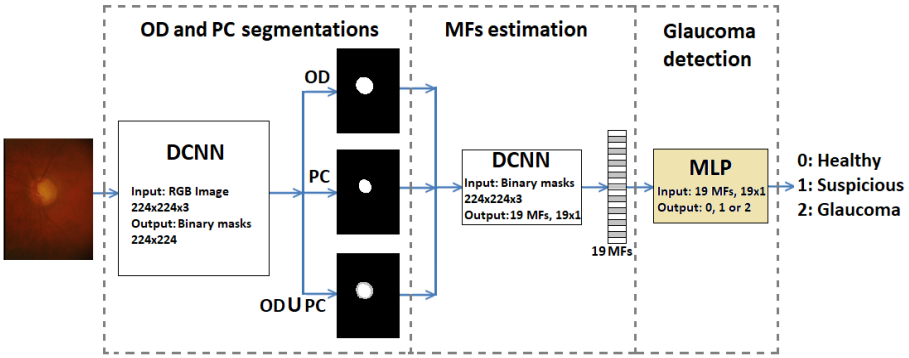

Fig. 1. Block diagram used to segment binary masks (first block), to extract morphometric features (second block), and to classify into healthy, suspicious and glaucoma classes (third block).

(nor disc neither cup pixels) compared to foreground pixels (disc and cup pixels). The coarse DRIU feature maps or the 4^{th}, 7^{th}, 10^{th} and 13^{th} convolutional layers are extracted, stacked and up-sampled to generate a binary mask with size 224×224. A modification of the original DRIU was done adding two convolutional layers in cascade with one-padding and kernels size of 3×3 and 1×1 to improve resolution details at the optic and cup edges, without affecting the final size. The model is trained using binary cross-entropy as loss function with ground-truth segmentations of the OD and the PC provided by experts.

2.2 DCNN for Automatic Morphometric Estimation

The second stage takes as input the segmentations of the OD and the PC, along with a third image corresponding to the union, to calculate 19 morphometric features used by Perdomo et al. [13]. The 19 morphometric features can be divided into four subsets: geometric, distances, axis and ratio. The geometric subset contains areas and perimeters of OD and PC; the distance subset is composed of the superior, inferior, nasal and temporal distances between the OD and PC; the axis subset is defined by major and minor axis of the OD and the PC, finally, the ratio subset includes the eccentricity of the OD and PC and the five ratios between upper OD and PC parameters to seek disproportions and relationships between optic disc and the physiologic cup [13].

The DCNN designed for automatic morphometric feature estimation is composed of five convolutional, five pooling layers and two fully connected layers. The four max-pooling and the global average pooling layers are non-linear size reducers that are applied to reduce the spatial dimensions, to minimize overfitting and the number of parameters in the model. Finally, the first fully-connected layer connects all the neurons obtained from the global average-pooling layer with 512 neurons to the next fully-connected layer with the number of morpho-

metric features to predict. The model is trained using mean absolute error as the loss and, morphometric features directly predicted from the OD, PC and union segmentations as ground-truth.

2.3 Multilayer Perceptron for Glaucoma Classification from Morphometric Features

The final stage corresponds to a multilayer perceptron (MLP) that receives as input the 19 morphometric features and classifies them into three possible classes: normal, suspicious or glaucoma. The MLP is composed of two fully-connected layers with 64 hidden and 3 output units. The batch size, number of epochs and initial learning rate used were explored using a grid search strategy, the best performing parameters found experimentally were 16, 500 and 0.01 respectively.

3 Experimental setup

3.1 Eye Fundus Image Databases for Glaucoma Detection

The DRISHTI-GS1, RIM-ONE_v1 and RIM-ONE_v3 databases are used in this study [14,15]. The DRISHTI-GS1 dataset has been acquired and labeled as healthy and glaucomatous by Aravind Eye Hospital (India), and it contains 101 color fundus images distributed in two subsets: 50 images for training, and the 51 remaining for the testing subset [14]. RIM-ONE_v1 and RIM-ONE_v3 focus on optic nerve head segmentation for glaucoma detection with manual reference segmentations as gold standard with 455 and 159 images respectively, created by ophthalmologists from the Department of Ophthalmology at the Hospital Universitario de Canarias in Spain [15]. The RIM-ONE_v1 dataset was labeled according to a binary classification (healthy vs. glaucomatous), and RIM-ONE_v3 was labeled as a 3-class classification problem (healthy, suspicious and glaucoma).

The proposed method was evaluated in 2 setups: binary-classification task (DRISHTI-GS and RIM-ONE_v1) and 3-class classification (RIM-ONE_v3). Additionally, the proposed method used a stratified sampling to randomly divide two RIM-ONE datasets into three subsets with 60%, 10% and 30% for each class that correspond to training, validation and test sets respectively.

3.2 Evaluation

The proposed method used several performance metrics in each stage. The OD and PC segmentations stage were assessed with the Jaccard index and the Dice coefficient. The MF estimation stage was evaluated using the Mean Average Percentage Error (MAPE) among all the predicted MFs. Moreover, the MFs calculated from the OD and PC segmentations by the experts called Real Morphometric Features (RMFs) were compared to the Estimated Morphometric Features (EMFs) or the MFs calculated from the OD and PC segmentations by the

first stage as reported in Table 1. Finally, the complete proposed method for glaucoma detection was evaluated using accuracy, sensitivity, specificity, Area Under the Curve (AUC), precision, recall, f-score, Kappa coefficient, and Overall Accuracy (OA) performance metrics reported on the test sets.

For the binary task, a combination of an 8-layer CNN model and a linear SVM applied to RIM-ONE_v1 [6], and a fine-tuning of VGG pre-trained on ImageNet applied to DRISHTI-GS1 [8] were chosen as binary-classification baseline models, as reported for the test set in Table 2. The end-to-end data-fusion deep learning model that combines raw color fundus images and RMFs was chosen as the 3-class classification baseline [13] on the RIM-ONE_v3 dataset. Furthermore, a DCNN feeding with color fundus images and a 3D-binary mask described in Section 2 was compared to the proposed method. The proposed approach was implemented with Keras* using a GeForce GTX TITAN X from NVIDIA.

4 Results

The best performance of the proposed model for OD and PC segmentations (first stage) was obtained with a learning rate of $1e-6$, a batch size of 2, a number of samples per epoch of 300, and a number of epochs of 10.000. The Jaccard index (JI) and Dice coefficient (DC) were monitored during training in the first stage (not reported here), and these parameters were evaluated on the test set for OD segmentation with of $JI = 0.9975$ and $DC = 0.9987$ of, and PC segmentation with $JI = 0.9983$ and $DC = 0.9991$ respectively.

For the second stage, the EMFs presented a MAPE during training of 3.57%, and a MAPE in the test set of 6.30% compared to RMFs. Table 1 presents a comparison between the RMFs and EMFs for 3-class classification on the RIM-ONE_v3 dataset using SVM and Random Forest (RF) classifiers as reported in [13].

Table 1. Performance measures for the RMFs and EMFs on the test RIM-ONE_v3 dataset, bold values show the best score for each performance metric.

Method	Source	Precision	Recall	f-score	Kappa	OA
SVM [13]	RMFs	0.63	0.56	0.55	0.35	0.66
RF [13]	RMFs	**0.64**	0.57	0.58	**0.37**	**0.65**
SVM	EMFs	0.54	0.41	0.39	0.23	0.57
RF	EMFs	0.60	**0.64**	**0.61**	0.35	0.64

Finally, we evaluated the proposed method with baseline methods for binary and 3-class classifications (third stage) as reported in Tables 2 and 3 respectively. These tables present the comparison between the methods, information sources and performance metrics evaluated in classification tasks for the two experimental setups respectively.

* http://keras.io

Table 2. Binary-classification performance metrics for baseline models and the proposed method on the test data set. The bold values show the best score for each performance metric.

Method	Source	Accuracy	Sensitivity	Specificity	AUC
Al-Bander et al. [6]	RGB	0.882	0.85	**0.908**	–
Orlando et al. [8]	RGB	–	–	–	0.763
Proposed method	RGB	**0.894**	**0.895**	0.889	**0.82**

Table 3. Comparison of performance metrics for 3-class classification for baseline models and the proposed method on the test dataset. The bold values show the best score for each performance metric.

Method	Source	Precision	Recall	f-score	Kappa	OA
DCNN + RMF [13]	RMF,RGB	0.46	0.56	0.50	0.42	0.68
DCNN	RGB	0.48	0.55	0.51	0.20	0.55
DCNN	OD -PC-Union	0.60	0.60	0.59	0.29	0.60
Proposed method	RGB	**0.76**	**0.72**	**0.69**	**0.48**	**0.70**

5 Discussion and Conclusions

We present a novel method for automatic glaucoma assessment from eye fundus images based on DCNNs. The results show that our method is competitive with the best results reported for each dataset: RIM-ONE-v1 accuracy of 89.4% vs. 88.2% reported by [6], and DRISHTI-GS AUC of 0.82 vs 0.76 reported by [8], as shown in Table 2. The most remarkable characteristic of this model is its architecture and training strategy. The model is organized in stages that follow a conventional process for glaucoma diagnosis based on the calculus of morphometric features. The multistage architecture allows us to train the model using a curriculum learning approach, which gradually trains the model to accomplish subtasks with increasing complexity. This approach allows training a complex deep learning model with a reduced set of training samples, resulting in an improved performance of the model that was corroborated by the experimental evaluation. In particular, the experimental results showed that the multistage architecture along with the curriculum training, produces better results than conventional DCNNs, as reported in Table 3. The resulting model being end-to-end, it is able to directly produce a prediction from the input image without requiring a manual intermediate segmentation required by the conventional diagnosis protocol from eye fundus images.

The work shows that it is possible to involve domain knowledge in deep learning models. Additionally, intermediate results produced by the model (segmentations and morphometric features) can help the interpretability of the model predictions, making them more useful in support of the diagnosis process. We hypothesize that this approach can be extended to other medical image analysis applications and exploring this hypothesis will be the focus of our future work.

References

1. Yih-Chung Tham, Xiang Li, Tien Y Wong, Harry A Quigley, Tin Aung, and Ching-Yu Cheng. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11):2081–2090, 2014.
2. Robert N Weinreb, Tin Aung, and Felipe A Medeiros. The pathophysiology and treatment of glaucoma: a review. *Jama*, 311(18):1901–1911, 2014.
3. Daniel M Stein, Gadi Wollstein, and Joel S Schuman. Imaging in glaucoma. *Ophthalmology clinics of North America*, 17(1):33, 2004.
4. Pardha Saradhi Mittapalli and Giri Babu Kande. Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma. *Biomedical Signal Processing and Control*, 24:34–46, 2016.
5. Anindita Septiarini, Dyna M Khairina, Awang H Kridalaksana, and Hamdani Hamdani. Automatic glaucoma detection method applying a statistical approach to fundus images. *Healthcare informatics research*, 24(1):53–60, 2018.
6. Baidaa Al-Bander, Waleed Al-Nuaimy, Majid A Al-Taee, and Yalin Zheng. Automated glaucoma diagnosis using deep learning approach. In *SSD, 14th International Multi-Conference on*, pages 207–210. IEEE, 2017.
7. Xiangyu Chen, Yanwu Xu, Damon Wing Kee Wong, Tien Yin Wong, and Jiang Liu. Glaucoma detection based on deep convolutional neural network. In *EMBC, 37th Annual International Conference of the IEEE*, pages 715–718. IEEE, 2015.
8. José Ignacio Orlando, Elena Prokofyeva, Mariana del Fresno, and Matthew B Blaschko. Convolutional neural network transfer for automated glaucoma identification. In *12th International Symposium on Medical Information Processing and Analysis*, volume 10160, page 101600U. International Society for Optics and Photonics, 2017.
9. Artem Sevastopolsky. Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. *Pattern Recognition and Image Analysis*, 27(3):618–624, 2017.
10. Qaisar Abbas. Glaucoma-deep: Detection of glaucoma eye disease on retinal fundus images using deep learning. *International Journal of Advanced Computer Science and Applications*, 8(6):41–45, 2017.
11. Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM, 2009.
12. Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, and Luc Van Gool. Deep retinal image understanding. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 140–148. Springer, 2016.
13. Oscar Perdomo, John Arevalo, and Fabio A González. Combining morphometric features and convolutional networks fusion for glaucoma diagnosis. In *13th International Conference on Medical Information Processing and Analysis*, volume 10572, page 105721G. International Society for Optics and Photonics, 2017.
14. Jayanthi Sivaswamy, S Krishnadas, Arunava Chakravarty, GD Joshi, A Syed Tabish, et al. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomedical Imaging Data Papers*, 2(1):1004, 2015.
15. Francisco Fumero, Silvia Alayón, JL Sanchez, J Sigut, and M Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. In *Computer-Based Medical Systems (CBMS), 24th International Symposium on*, pages 1–6. IEEE, 2011.