Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 196 (2022) 485–492

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

Optic disc and cup segmentations for glaucoma assessment using cup-to-disc ratio

Alexandre Neto[a,c], José Camera[b,c], Sérgio Oliveira[a], Ana Cláudia[a], António Cunha[a,c]*

[a]*Universidade de Trás-os-Montes e Alto Douro, Vila Real 5000-80, Portugal*
[b]*Universidade Aberta, Lisboa 1250-100, Portugal*
[c]*Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, Porto 3200-465, Portugal*

Abstract

Glaucoma is a silent disease that shows symptoms when severe, leading to partial vision loss or irreversible blindness. Early screening permits treating patients in time. For glaucoma screening, retinal images are very important since they enable the observation of initial glaucoma lesions, which typically begins with the cupping formation in the optic disc (OD). In clinical settings, practical indicators such as Cup-to-Disc Ratio (CDR) are frequently used to evaluate the presence and stage of glaucoma. The ratio between the cup and the optic disc can be measured using the vertical or horizontal diameter, or the area of the two. Mass screening programs are limited by the high costs of specialised teams and equipment. Current deep learning (DL) methods can assist the glaucoma mass screening, lower the cost and allow it to be extended to larger populations. With DL methods in the OD and optic cup (OC) segmentation, is possible to evaluate the presence of glaucoma in the patient more quickly based on cupping formation in the OD, using CDR.

In this work, is assessed the contribution of Multi-Class and Single-Class segmentation methods for glaucoma screening using the 3 types of CDR. U-Net architecture is trained using transfer learning models (Inception V3 and Inception ResNet V2) to segment the OD and OC and then evaluate glaucoma prediction based on different types of CDRs indicators. The models were trained and evaluated on main public known databases (REFUGE, RIM-ONE r3 and DRISHTI-GS). The segmentation of both OD and OC reach Dice over 0.8 and IoU above 0.7. The CDRs were computed to glaucoma assessment where was reach sensitivity above 0.8, specificity of 0.7, F1-Score around 0.7 and AUC above 0.85. Finally, conclusions of segmentation methods showing adequate performance to be used in practical glaucoma screening.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the CENTERIS –International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

* Corresponding author. Tel.: +351931636373.
E-mail address: acunha@utad.pt

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the CENTERIS –International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021
10.1016/j.procs.2021.12.040

Keywords: Glaucoma; Segmentation; Cup-to-disc ratio; Screening.

1. Introduction

Glaucoma is an asymptomatic neurodegenerative, chronic and irreversible disease in which the optic nerve is progressively injured. It is considered the second leading cause of blindness globally, with about 60 million cases in 2010 [1]. The undetected glaucoma prevalence is as high as 50% even in high-income areas like North America and Australia, increasing to 90% in middle and low-income regions such as Asia and Africa, a consequence of the inadequate screening tools and strategies to detect these glaucomatous lesions. Without more individuals accessing routine eye examinations, glaucoma still goes as undetected [2]. Screening is a crucial point in fighting this disease because diagnosing glaucoma at early stages gives the chance to stop its progression. Studies have revealed that morphological changes in the optic disc (OD) and in the optic cup (OC) reflect damage to the optic nerve, leading to glaucoma. The Cup-to-Disc Ratio (CDR) criterion, which is an indicator widely used by experts to detect glaucoma, is a method that does the ratio between the cup and the vertical disc [3], [4]. Computed-aid diagnosis solutions for screening glaucoma are in need in situations such as mass screening and medical care, and even more in countries with a significant lack of qualified specialists [5]. Human experts' manual delineation of OD and OC boundaries is highly subjective and time-consuming, which is impractical for use in mass screening. Automatic segmentation approaches can be more objectives and faster than humans [6].

In this paper, Deep Learning (DL) methods based on automatic segmentation of the OD and OC for glaucoma screening in retinal images using transfer learning techniques are evaluated for application in clinical practice in glaucoma assessment.

As the outline for this paper, in section 2, is presented a revision of the public databases with glaucoma images and the DL methods for glaucoma screening based on segmentation approaches. In section 3, the methodology implemented for glaucoma screening based on segmentation is presented, and in section 4, the results are described and discussed for OD and OC segmentation and compared with the literature review. Finally, in section 5, were taken conclusions.

2. Literature Review

Glaucoma detection can be achieved through the computation of CDR based on automatic segmentation of the OD and OC on retinal images. In this section, the main public databases of retinal images available and used by researchers (section 2.1.) and the recent methods published for glaucoma screening based on the segmentation approach (section 2.2.) are summarised and the CDR criterium (section 2.3.) explained.

2.1. Public Databases

The databases with retinal images reported in the literature publications are summarised in Table 1.

Table 1. Public databases for glaucoma.

Database	Glaucoma/Normal	OD/OC GT	Total
ACRIMA	396/309	No	705
DRIONS-DB	-	No	110
DRISHTI-GS	70/31	Both	101
HRF	27/18	Both	45
ONHSD	-	Only OD	99
ORIGA	168/482	No	650
REFUGE	40/360	Both	400
RIM-ONE r3	74/85	Both	159
Sjchoi86-HRF	101/300	No	601

All of these are public databases. The different databases reveal heterogeneity in their characteristics from lighting, the field of view and resolutions, and there are huge variations between them in these same aspects. Also, and as can be seen in the table, there is a lack of cohesion in the structuring of the databases from the clinical annotations, in some cases masks only for OD or cup and different types of file format representations for the masks, as well as a lack of specific information, in some cases, the sample label.

2.2. Glaucoma screening based on the optic disc and optic cup segmentation

Several methods were published in the literature for segment OD and OC. It follows an overview of the best methods recently published.

The study developed by Al-Bander [6] was proposed a method with a DenseNet incorporated with a Fully Convolutional Network (FCN) with U-shaped architecture. This deep network encourages feature re-use and reduces the number of parameters to improve the segmentation of the OD and OC. The approach of Al-Bander has used five databases of colour fundus images: ORIGA, DRIONS-DB, DRISHTI-GS, ONHSD and RIM-ONE. For the pre-process, only the green channel of the colour images was considered since the other colour channels contain less useful information. The images were then cropped in the Region of Interest (ROI), and to increase the number of images artificially was done augmentation processes with vertical flips and random crops. For the OD segmentation, the model reached better results in Dice and IoU for the DRISHTI-GS database than RIM-ONE, and the cup segmentation was confirmed the same thing. Still, lower values of Dice and IoU compared to the OD segmentation.

Studies from Sevastopolsky [7] paper used a U-Net architecture to segmentation the OD and cup. The proposal U-Net has fewer filters in all convolutional layers and does not increase filters for decreasing resolution. For this model, was used three databases: DRIONS-DB, DRISHTI-GS and RIM-ONE r3. The images were then pre-processed with a contrast limited adaptive histogram equalisation and then cropped by bounding boxes in the ROI. The segmentation of OD reached good results and for the cup segmentation on both databases. Dice reach values above 0.8, proving to be a good model.

Qin [8] proposes neural network constructs utilising the FCN and the Inception building blocks in the GoogleNet. The FCN is the main body of the deep neural network architecture of this method. They add several convolution kernels for feature extraction after deconvolution, based on the Inception structure in GoogLeNet. Qin's experiments used two databases: REFUGE and one from the Second Affiliated Hospital of Zhejiang University School of Medicine. For this technique, a fully automatic method using Hough Circle Transform recognises and cuts the image to obtain an image of the Region of Interest (ROI). The images data are increased by rotating, flipping and adjusting the contrast and then using the Laplacian for image enhancement. Since the red channel contains less useful information for this paper is only used blue-green channels. In the OD and the cup segmentation, the model reached values above 0.9 for the Dice and the IoU.

However, most of the studies do not have a practical application for the segmentation results in the glaucoma screening process. The segmentation can help in mass screening by using the output segmentation for CDR calculation, which is a well define and used indicator for glaucoma assessment by specialists.

2.3. Cup-to-Disc ratio

Glaucoma progression can be assessed base on the ratio of measures of OD and OC. The CDR is a clinical method that compares the ratio of the cup to disc diameters. If the vertical CDR (VCDR, equation 1) and horizontal CDR (HCDR, equation 2) are more than 0.5, the eye can be considered a threat of abnormality. Otherwise, it is considered a normal eye [9]. Alternatively, considering the criteria by Andres Diaz Pinto [9], the assessment can be done through the area CDR (ACDR, equation 3) using the threshold of 0.3.

$$VCDR = \frac{V_{cup}}{V_{disc}} \quad (1) \qquad HCDR = \frac{H_{cup}}{H_{disc}} \quad (2) \qquad ACDR = \frac{A_{cup}}{A_{disc}} \quad (3)$$

The work developed by Andres Diaz Pinto [9] presented an automatic algorithm that uses several colour spaces and the Stochastic Watershed transformation to segment the cup and then obtains handcrafted features such as the VCDR, HCDR and ACDR. Diaz method was evaluated on 53 images obtaining a specificity and sensitivity of 0.81 and 0.87.

After the segmentation, Al-Bander [6] calculated the VCDR varying the thresholds and compared with the expert's glaucoma diagnosis, achieving an AUC of 0.74, close to the 0.79 achieved using the ground truth segmentations. After that, it was made the same approach but HCDR achieving an AUC of 0.78 close to the 0.77 achieved by the expert's annotation, higher results than the results obtained with the VCDR.

This work will use segmentation models to compute the different types of CDRs for the glaucoma assessment, and evaluate the contributions for the practical application in glaucoma screening, comparing to works that do not use all the types of CDRs or do not use DL models for CDR calculation.

3. Methodology

In this work, glaucoma screening methods based on the segmentation approach are assessed. The pipeline used for this work is described in Fig.1.

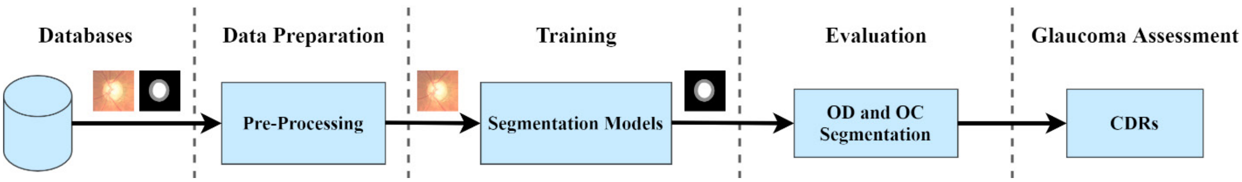

Fig. 1. The pipeline of the work.

Retina and mask images from public databases are pre-processed and organised in datasets. These datasets are used to train Multi-Class (MC) and Single-Class (SC) segmentation models capable of segmenting the OD and OC in retinal images. The MC segmentation does the two tasks at once while the SC has one model train to segment only the OD and the other train to segment the OC. The segmentations from the models are then evaluated and used to compute the CDRs. In the end, the use of CDRs as indicators for glaucoma assessment are evaluated.

3.1. Databases

Was used data from the public databases mention in Table 1, and selected the databases with OD and OC masks with easy access that have the respective labels:

- The RIM-ONE r3 database has a balance proportion between normal and glaucomatous samples with 85 healthy images and 74 glaucomatous images. This database has enormous variations in the quality of the images in the illumination and contrast with low light images that turn difficult the identification the OD and the cup and images with good illumination and contrast that help identify the retinal components.
- The DRISHTI-GS has a bigger representation of glaucoma samples (70 images) than healthy samples (31 images). Compared to the RIM-ONE, the DRISHTI-GS has more homogeneous illumination and contrast that helps to identify and segment the OD and the cup.
- The REFUGE database is composed of 400 images (only access to the validation set), with a lower representation of glaucoma samples compared to the healthy (40 glaucomatous images and 360 normal images).

3.2. Data Preparation

All three databases were merged in a unique database (DB) where each image was centralised in the OD region and then cropped to focus the Convolutional Neural Networks (CNNs) in the ROI. The cropped images have 512x512 resolutions. Was applied augmentations processes such as rotations, zooms and shifts variations in DB to avoid the model overfitting. 70% of the images were used for training, 15% for validation, and 15% for test.

3.3. Training

For the segmentation approach in the glaucoma screening, the pre-trained models selected based on the literature review and the high performance was the Inception V3 and Inception ResNet V2 (for simplification, S1 and S2, respectively). The availability of a huge dataset as ImageNet to train the model with high capacity leads to many pre-trained models used for feature encodes. The encoder in a U-Net is a stack of convolution layers combined with activations functions and pooling layers that can have the architecture that is frequently employed for feature extraction with pre-trained models. The use of U-Net is based on the high performance shown in the literature review.

3.4. Evaluation

The metrics for the evaluation of the segmentation model were the IoU and the Dice coefficient. The IoU metric measures the accuracy of an object detector on a particular database. It measures the common area between the predicted and expected regions, divided by the total area of the two regions. The Dice coefficient is a statistic used to gauge the similarity of two samples, in this case, between predicted and ground truth segmentation. Although IoU and Dice are inter-related, both were used for the sake of comparison with other publications.

3.5. Glaucoma Assessment

Segmentations were used to compute the CDRs and assess glaucoma. Was used the criterium of Diaz [9] where the VCDR and HCDR can be considered normal if they are less than 0.5 and glaucomatous otherwise. For the ACDR, the samples with a ratio below 0.3 can be considered normal and glaucomatous otherwise.

For the evaluation of the glaucoma assessment using CDRs was used the Sensitivity (Sen) and the specificity (Sep), where the Sen measures the proportion of positives that are correctly identified and the Sep measures the proportion of negatives that are correctly identified. The CDRs equations are described in section 2.3. in the equations (1), (2) and (3). The F1-Score (F1) conveys the balance between the precision and the recall where the precision is the number of TP divided by the number of all positives and the recall is the number of TP divided by the number of all samples that should have been identified as positive. The F1 is the harmonic mean of these two.

4. Results and Discussion

4.1. Training setup

The U-Net encoder weights were frozen for 20 epochs during the pre-training step and 100 epochs during the fine-tuning step (the encoder layers were unfrozen). The learning rate started at 10^{-4} with Adam optimiser and binary cross-entropy as the loss function. To prevent stagnation of the learning rate on the plateau, the callback of reducing the learning rate on the plateau as a factor of 0.90 and only is saved the best training weights.

4.2. Segmentation results

The MC segmentation global results for both models were 0.94 (± 0.05) of Dice and 0.89 (± 0.08) of IoU. Table 2 presents the results separately for the OD and OC of the MC segmentation to enable comparison with the literature review works, as well as the results of the SC models for both segmentation tasks.

Table 2. Results for the OD and OC segmentation for our approach compared with the literature review results.

		OD		OC	
Database	Method	IoU	Dice	IoU	Dice
RIM-ONE	[6]	0.83	0.90	0.56	0.69
	[7]	0.89	**0.95**	0.69	0.82
Drishti-GS	[6]	0.90	**0.95**	0.71	0.83
	[7]	-	-	0.75	0.85
Refuge	[8]	**0.92**	**0.95**	**0.90**	**0.92**
DB (Our)	MC S1	0.73 (± 0.18)*	0.83 (± 0.17)*	0.72 (± 0.17)*	0.82 (± 0.15)*
	MC S2	0.72 (± 0.18)*	0.82 (± 0.17)*	0.71 (± 0.18)*	0.82 (± 0.16)*
	SC S1	0.89 (± 0.15)*	0.93 (± 0.13)*	0.71 (± 0.21)*	0.81 (± 0.21)*
	SC S2	0.91 (± 0.08)*	**0.95 (± 0.05)***	0.74 (± 0.18)*	0.83 (± 0.16)*

* average (± standard deviation)

For MC segmentation, both models had results below the ones mentioned in the literature review for the OD segmentation. Either way, the Dice coefficient from both models is above 0.80 with 0.17 of standard deviation, which shows a good performance from the models. The models showed difficulties in 6 samples with bad image quality jeopardising the mean of the Dice coefficient. Overall, the segmentation output for most of the samples is accurate and close to the ground truth masks, with some exceptions that will be represented in Fig. 2. For the SC models for the OD segmentation, the results of the Dice coefficient are above 0.9 with lower values of standard deviation comparing to the MC segmentation methods. The SC models show difficulties in the same samples that MC models had lower results.

The results for the OC segmentation are at the level of the works mentioned and close to the results of the OD segmentation, showing consistency in the segmentation for both classes. The OC Dice coefficient for both models in the MC approach is above 0.8 with 0.17 of standard deviation as well, showing difficulties in the same samples mention before. The SC models have results similar to the MC models for the OC segmentation, with Dice above 0.8 as well. The results for the segmentation in both approaches reach levels of performance good enough to use the predict masks of the OD and OC to calculate the CDRs. The CDRs computations are the main point since they can indicate glaucoma presence or not. The segmentation must be good enough so the use of CDR is reliable.

In Fig. 2 are illustrated some predictions for both models and the respective ground truth masks (Masks GT). Fig.2 (a) and (b) are an example of images represent good results that are close to the Mask GT, showing the good performance of the models. On the other hand, in (c) and (d), both models of the MC and SC approach reveal difficulties to correctly segment the OD and the OC since the input images do not contain good enough quality to identify and segment the components. The bad results in all the approaches used were detected in the RIM-ONE samples, since, as mention before, this dataset has heterogeneous conditions of light, with some darker retinal images that make the segmentation even more complicated. Even for a specialised professional, this could be a hard image to evaluate and segment the OD and OC manually.

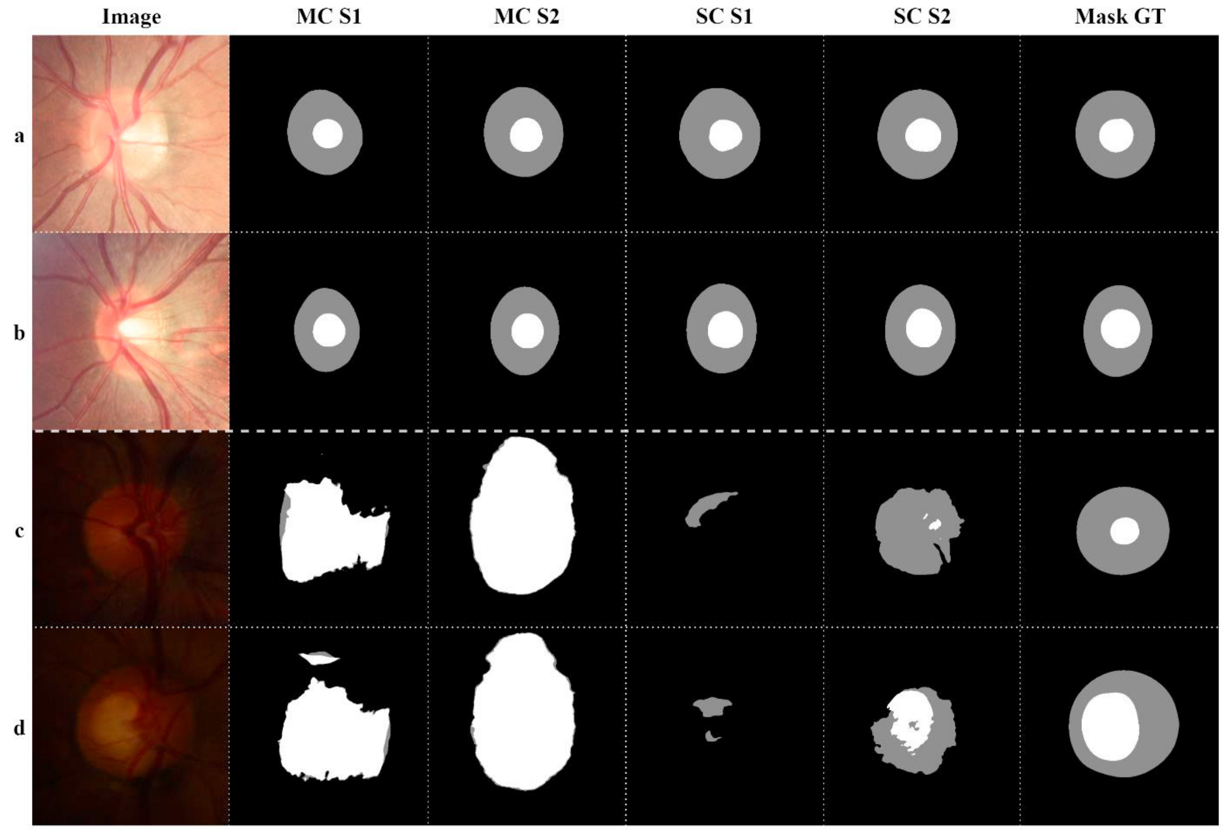

Fig. 2. Examples of good results (a) and (b) of segmentation for both models and bad results (c) and (d), compared to the Mask GT of the respective images.

4.3. Glaucoma assessment

The CDRs (ACDR, HCDR and VCDR) are computed as an indicator for glaucoma assessment and the results are presented in Table 4.

Table 3. Results of glaucoma classification with CDRs calculation for S1, S2 and Masks GT.

	Masks GT			MC S1			MC S2			SC S1			SC S2		
	ACDR	HCDR	VCDR	ACDR	HCDR	VCDR	ACDR	HCDR	VCDR	ACDR	HCDR	VCDR	ACDR	HCDR	VCDR
Sen	0.82	0.86	0.86	0.82	0.82	0.89	0.86	0.93	0.93	0.69	0.73	0.85	0.85	0.85	**0.96**
Spe	0.81	0.64	0.75	**0.86**	0.64	0.79	0.83	0.61	0.74	**0.86**	0.64	0.74	0.84	0.56	0.70
F1	0.71	0.62	0.69	**0.75**	0.6	0.74	**0.75**	0.63	0.71	0.67	0.54	0.67	**0.75**	0.57	0.70
AUC	0.88	0.83	0.90	0.91	0.87	0.93	0.91	0.89	0.91	0.85	0.79	0.88	0.93	0.86	**0.96**

The segmentation masks of the OD and OC of both models are used to calculate the ratio between them.

In this work, all the CDRs were calculated, including the vertical, horizontal and ratio between the areas of the OD and OC as mentioned. For the VCDR and HCDR, the criterium used was normal for a CDR < 0.5 and glaucomatous for CDR ≥ 0.5 and the ACDR normal if < 0.3 and glaucomatous if ≥ 0.3 based on Diaz [9] work. The same criterium is used to the Masks GT to have a direct comparison between the results from our models and the segmentation made by ophthalmologists to see how reliable the segmentation made by S1 and S2 models is. The results for both models in the MC and SC approach came close to or even higher than the results using the Masks GT, which indicates that the segmentation is close to each other or at least provides close CDRs. Overall, both models reach slightly better results than the results of the CDRs using the Masks GT. Compared to the Diaz [9] and Al-Bander [6] work, our results for both models reach higher values of Sen, Spe and AUC.

Using the same CDR criterium for the segmentation outputs of both models and the Masks GT shows similar metrics results, reflecting that the automatic segmentation can help to facilitate these tasks providing similar results when evaluated using the same indicator as the CDR. These results show as well how correlated the CDR and glaucoma presence is, revealing that when used with the right criterium, CDR can be a good form of glaucoma assessment. All the different types of CDRs have good and similar results, with a slightly better performance of the VCDR overall.

5. Conclusions

Along with clinical data and the visual field test, the CDR is a widely used parameter in screening to analyze and identify the glaucomatous papilla, for that purpose the segmentation methods to classify the samples based on the CDRs were applied. The segmentation of the OD is easier than the segmentation of the OC, which most of the time is difficult to define the limits making it harder for the clinicians to perform this task. So, for this reason, the CNNs prove to be already a good help to facilitate a subjective and hard task that highly depend on the experience of the ophthalmologist. The CDRs computed through the segmented masks were very close to the Masks GT, reinforcing that the CNNs can do an evaluation similar to the one done by a clinician. The model that reaches better results overall for these tasks was Inception V3 as the backbone of the U-Net, with slightly better performance for the different CDRs. Either way, some predictions do not reach the expected results but as showed, that can be correlated to the bad quality of light and contrast in images that even for a specialist can make the manual segmentation even harder. Both MC and SC segmentation approaches reach good enough results for the use of the OD and OC masks for glaucoma assessment using the different types of CDR. The use of CDRs prove to be a good practical application of the segmentation since the results of glaucoma assessment using these indicators have high AUC results (lower AUC of 0.79 and higher AUC of 0.96).

References

[1] M. L. Claro, R. Veras, L. Santos, M. Frazão, A. Carvalho Filho, and D. Leite, “Métodos computacionais para segmentação do disco óptico em imagens de retina: uma revisão,” Rev. Bras. Comput. Apl., vol. 10, no. 2, pp. 29–43, 2018.

[2] S. MacIver, D. MacDonald, and C. L. Prokopich, "Screening , Diagnosis , and Management of Open Angle Glaucoma : An Evidence-Based Guideline for Canadian Optometrists," Cjo, vol. 79, no. 1, pp. 1–72, 2017.

[3] J. Cheng et al., "Superpixel classification based optic cup segmentation," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8151 LNCS, no. PART 3, pp. 421–428, 2013.

[4] S. Sreng, N. Maneerat, K. Hamamoto, and K. Y. Win, "Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images," Appl. Sci., vol. 10, no. 14, 2020.

[5] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network," Pattern Recognit. Image Anal., vol. 27, no. 3, pp. 618–624, 2017.

[6] B. Al-Bander, B. M. Williams, W. Al-Nuaimy, M. A. Al-Taee, H. Pratt, and Y. Zheng, "Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis," Symmetry (Basel)., vol. 10, no. 4, 2018.

[7] A. Sevastopolsky, S. Drapak, K. Kiselev, B. M. Snyder, J. D. Keenan, and A. Georgievskaya, "Stack-U-Net: Refinement Network for Image Segmentation on the Example of Optic Disc and Cup," 2018.

[8] P. Qin, L. Wang, and H. Lv, "Optic disc and cup segmentation based on deep learning," Proc. 2019 IEEE 3rd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2019, no. Itnec, pp. 1835–1840, 2019.

[9] A. Diaz, S. Morales, V. Naranjo, P. Alcoceryz, and A. Lanzagortayz, "Glaucoma diagnosis by means of optic cup feature analysis in color fundus images," Eur. Signal Process. Conf., vol. 2016-Novem, pp. 2055–2059, 2016.