Training U-Net with Proportional Image Division for Retinal Structure Segmentation

Pedro Victor de Abreu Fonseca[1], Alexandre Carvalho Araújo[1], João Dallyson S. de Almeida[1(✉)], Geraldo Braz Júnior[1], Aristófanes Correa Silva[1], and Rodrigo de Melo Souza Veras[2]

[1] Federal University of Maranhão - UFMA, Applied Computing Group - NCA/UFMA, São Luís, Brazil
jdallyson@nca.ufma.br
[2] Federal University of Piauí - UFPI, Teresina, Brazil

Abstract. Cup and optic disc segmentation has become one of the main objects of study in the field of creating and improving machine learning-oriented models due to the importance of vision for human beings and the ability to assist physicians in diagnosing ocular problems. Within this context, this study presents a new method based on the proportional division of images concerning features extracted from the sample set. These samples go through a pre-processing step involving image resizing before going to deep feature extraction and K-means clustering, thus dividing the set for validation and training. Soon after, the amount of samples is increased through data augmentation before going on to the U-Net training. The proposed method has been evaluated on the public RIM-ONE and DRISHTI-GS datasets, and presented promising results in the segmentation of both structures, with emphasis on obtaining the value of 92.2% of Dice for the segmentation of the optic cup in the DRISHTI-GS test dataset and 95.9% of Dice for the optic disc in the RIM-ONE.

Keywords: Segmentation · Retinal fundus · Proportional Image Division · U-Net

1 Introduction

Vision is one of the most important senses for human beings because we can understand the world around us. However, the organ connected to this sense, the human eye, is susceptible to pathologies leading to partial or total loss of this sense. Part of these problems include Glaucoma and Cataracts, which represent a large part of the cases related to the loss of sight [6,25]. Moreover, according to [31], Glaucoma is a pathology that, due to its silent way of acting, may soon become the leading cause of blindness, estimating a total of 118.8 million people worldwide will be affected.

A. Cunha et al. (Eds.): MobiHealth 2023, LNICST 578, pp. 122–138, 2024.
https://doi.org/10.1007/978-3-031-60665-6_9

These eye diseases can be detected early with examinations that physicians can perform, such as the use of structural tests for a clinical analysis of the optic disc on OCT images. Functional tests are based on standard automated perimetry, which is an analysis of the individual's field of vision [28]. Early treatment can be initiated by analyzing and checking retinal structures such as the cup and the optic disc. Due to the human effort required for these tasks and the growing number of cases linked to these diseases, algorithms that use Deep Learning for the segmentation and classification of the structures of the human eye have been developed and improved to provide concise and robust results, facilitating the work done by professionals in the area [2,5,8,14].

The optic disc and cup are the structures most targeted in the retinal segmentation task. However, such a task proves to be challenging for even the most refined neural networks, with emphasis on the optic cup due to its size, tonality, and the presence of other surrounding structures such as blood vessels [32]. Such structures are arranged in Fig. 1. To deal with the difficulty of this task, image processing techniques have been used, such as extraction of a region of interest (ROI) that aims to highlight only the part of the original image that contains the structures [21,32] and the extraction of the green channel from the image [15]. Important highlighted is the use of architectures using Deep Learning that has created near state-of-the-art learning models.

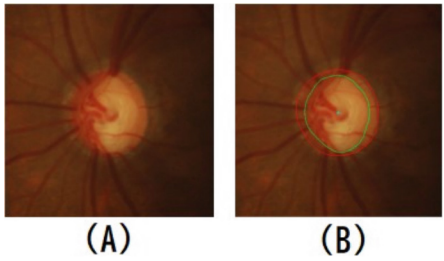

Fig. 1. (A) represents the disc and optic cup area, while (B) delimits these areas. The larger circle is the disc, and the smaller one is the optic cup. Adapted from [24]

We propose an approach that works directly with the optic cup and disc segmentation region to have a balanced set of samples. Differing from related work that has been heavily concerned with preprocessing the images laid out for training [12,16,21], using as architecture the U-Net neural network with encoder from ResNet-34 presented by [32]. The proposed method includes an initial step of proportional separation of the images used in the training/validation of the model, according to features extracted concerning the disc and the optic cup, seeking a balancing in the number of samples arranged in each set of images aiming to improve the generalization of the segmentation model produced.

Our study contributes to segmenting fundus images, precisely the optic cup and optic disc, by sampling the results and obtaining an approach heavily based on sample separation by balancing the sets used for training using feature extraction previously extracted from the samples available for training.

This paper is organized as follows: Sect. 2 presents the related works to the fundus retinal structures segmentation in images. Section 3 details the materials and method of this work, including datasets, the proposed method with proportional division approaches, and the experiment set-up. In Sect. 4, we present the results of the proposed method. Section 5 discusses the results. Finally, Sect. 6 presents the conclusions and suggestions for future work.

2 Related Works

Segmentation of retinal structures is one of the most addressed tasks in the literature, in which several approaches are used to obtain good results. [34] used entropy techniques in union with a uniform set of fundus examination samples used by a convolutional neural network based on boosting filters. As much as modifications to the architectures have been widely applied, [12] preferred to use the standard U-Net architecture without modifications to it by applying image processing techniques working with eye fundus scans in an even smaller proportion when compared to the works cited here. [21] using image processing techniques, extracted a region of interest from the set of fundus exams for segmentation of the optical cup, which soon after is passed to U-Net architecture for segmentation. [26] proposed for the segmentation of retinal structures an architecture named ResFPN-Net, based heavily on multi-scale feature extraction that combined previously extracted features with new ones found. Such methodology allowed the segmentation of the disc and the optic cup to obtain positive results when compared to related works. [16] aimed at the segmentation of the optic cup and the optical disk through operations directly on the masks contained in the training set, applying modifications to the pixels contained in the edges, which were binary values of 0 or 1, to numbers generated through probabilities, as architecture the traditional U-Net was used.

For the classification of glaucoma fundus scans, [29] was able to achieve better results than related works on a sample set using as architecture implementation of U-Net, named U-Net++. In addition, hyperparameter optimization was used with a loss function that combined the binary entropy with the Dice coefficient function, performing a subtraction between the two equations.

Using only the DRISHTI-GS dataset, [30] construed a framework to aid the classification of Glaucoma in fundus examination samples. To accomplish such a task, two custom CNNs were implemented, each used for disk and optic cup segmentation. In addition, image processing techniques, such as edge detection, image dilation, and erosion application, were used to ensure that good results were obtained in segmenting the targeted structures.

As much as our study performs the segmentation of retinal structures separately, works regarding the simultaneous segmentation of the optic cup and the optic disc have been developed as [1]. The authors used a more diversified training concerning the number of samples available for training with the accession of a new set of fundus exams, the ORIGA dataset, in addition to the use of the DenseNet architecture for the simultaneous segmentation of the optic disc

and the optic cup, obtaining results superior to the methods used as a reference in large datasets. As a result of the advance in segmentation-oriented neural networks, hybrid architectures have emerged, increasing the potential of convolutional networks already known by the scientific community. An example is the integration of residual backbones that provide robust and dynamic learning due to the number of parameters and *shortcuts* that the final model presents [32]. Due to these advantages, good results were obtained by [32] integrating the pre-trained residual backbone of ResNet-34, thus avoiding training from the beginning, together with the U-Net, for the segmentation of retinal structures.

3 Materials and Method

3.1 Image Datasets

The public retinal image datasets DRISHTI-GS [23] and RIM-ONE [10] were selected for the experiments. These datasets were chosen because both are widely used for training architectures focused on the segmentation/classification of retinal fundus images. The DRISHTI-GS dataset contains 101 images in its entirety with a resolution of 2896×1944 pixels and with a field-of-view of 30°. In comparison, the RIM-ONE presents 159 images with a resolution of 2144×1424 pixels. Each image was used without cropping, without ROI extraction, thus differentiating from the works used as references [21,32]. The bases were used together for the training process. The Fig. 2 and Fig. 3 shows a sample of the Drishti-GS and RIM-ONE dataset respectively.

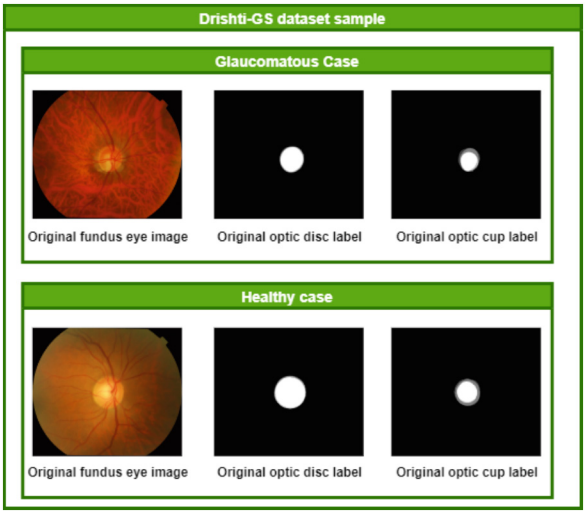

Fig. 2. Original sample of the Drishti-GS dataset. Originally, each mask contained in the dataset is composed of several markings for the optical cup and optical disk.

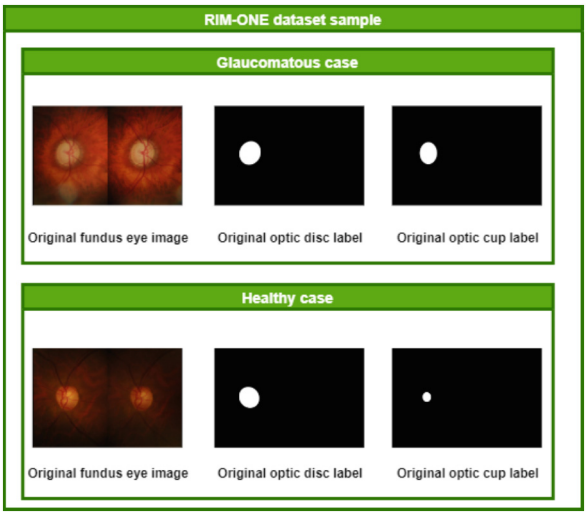

Fig. 3. Original sample of the RIM-ONE dataset. Each sample is composed of a pair of the same fundus eye exam.

3.2 Proposed Method

The proposed method is organized in the following steps: (1) preprocessing, (2) proportional split of samples, which can be done by extracting deep features or working with the average number of pixels for each mask from the training set, (3) data augmentation, (4) U-net with ResNet-34 backbone. Figure 4 presents the methodology steps.

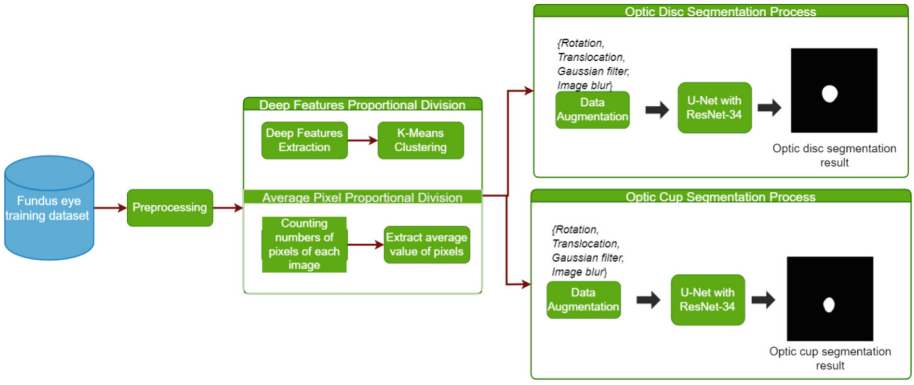

Fig. 4. Proposed Method. The proposed method uses mostly the same operations for both optical cup and optical disk segmentation.

Preprocessing. In the preprocessing step, the images were resized to 256×256 pixels, aiming to test the performance of the proposed method in an even smaller image than [32], which used the dimensions 512×512 pixels, thereby using fewer computational resources for model generation. Due to the change in the dimensions of the samples to a smaller value than the original shape, one can expect more difficulty from the architecture used for training in segmenting the structures, particularly that of the optic cup.

Proportional Division. Both datasets present for each image the disc delimitation and optic cup to be used in training, emphasizing the DRISHTI-GS masks with several labelings performed by experts. Thus, for each image in the DRISHTI-GS dataset, the demarcation with the smallest segmentation region was selected because, as previously mentioned, regions with small segmentation areas are challenging to segment, so to test the proposed methodology, such a choice was made. The RIM-ONE dataset has, for each image, only one label.

The proportional splitting approach was inspired by the work of [9], in which they proposed proportional splitting based on tumor size and pixel intensity to better distribute images from CT scans of the kidneys into the training, validation, and test sets. To define the division based on texture, the authors have defined a threshold (pixel intensity) that separates the reference cases into cases with light and dark tumors. Their goal of segmenting kidney tumors was improved with proportional splitting. In our study, we evaluate proportional splitting based on the area and texture of the regions of interest. In the latter approach, we set the split automatically using deep features and clustering using k-means.

Area Based Proportional Division. The mask annotations area of the training samples was initially used as the separation parameter to divide the training set evenly. The optic cup and the optic disc are represented by a set of white pixels in the masks in both datasets used. The area was obtained considering the number of pixels in each segmentation region. Next, the entire set of images for training is traversed to obtain the average area that references the region of interest contained in the masks. Thereafter, images are categorized as "large region of interest" or "small region of interest" according to their area compared to the average area. The perfect distribution would be 50% of images with the large area and 50% with the small area. However, this task is challenging due to the original arrangement of the DRISHTI-GS and RIM-ONE datasets that are originally unbalanced. Thus, the dataset's proportional splitting shows a variation of 20% concerning the perfect balancing in the worst-case distribution for this approach. The representation of this approach can be seen in Fig. 5.

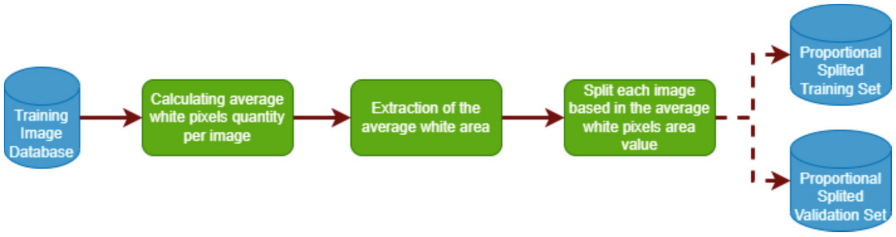

Fig. 5. Flow of operations for the proportional division based on the segmentation area.

Splitting Based on Deep Features + K-Means. This approach differs from the previous one because a network was used for feature extraction on the training set. The purpose of this new method was to verify the effectiveness of the texture features in the activity of the image's proportional division. After several experiments, the ResNet-34 model was used to perform feature extraction, which proved to be more efficient than other architectures previously used for the same task, such as VGG19, ResNet-50, and VGG16. Figure 6 shows steps for this approach.

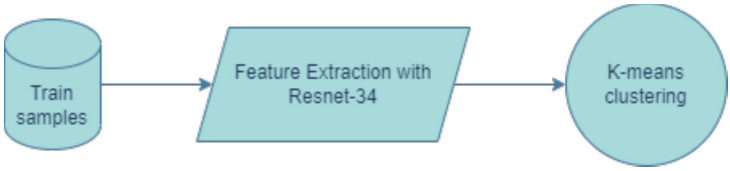

Fig. 6. Feature Extraction Approach using the ResNet-34 model.

After extraction, samples with similar characteristics are clustered using the K-means algorithm [11]. The silhouette coefficient was used to define the value for the number of clusters. This coefficient is calculated using Eq. 1.

$$S(x) = \frac{b(x) - a(x)}{\max(b(x), a(x))} \tag{1}$$

In which x represents an element in the π cluster, a(x) is the average distance of x to all elements in the π cluster, and b(x) represents the average distance of x from all points in another γ cluster. The interval of values for Eq. 1 is $[-1, 1]$, where the closer to 1 the better defined the clusters are [22]. An interval [2–5] of k values was evaluated. At the end of the process, the selected number with the best silhouette coefficient was K = 2.

Data Augmentation. To increase the number of dataset images, new images were generated performing operations of rotation, contrast change, and blur through the transformations of the Albumentation framework [7]. This generator moves the images at 45° angles in a scaling limit of 0.5, rotates the images vertically and horizontally with a probability of 50%, and also, in this same probability, applies filters and lighting effects such as Gaussian noise. It is important to note that due to the 50% value, previously established, each image may or may not be modified with all the operations determined.

To define a test set for the RIM-ONE dataset, of the 159 images initially contained in the same dataset, a percentage of 20% of these images were destined for the creation of a test set, totaling 30 images from the RIM-ONE for post-training evaluation and 129 images from the same dataset for training. Overall, 129 samples from the RIM-ONE combined with 50 images from the DRISHTI-GS, totaling 179 images, proceeded to be distributed into ten folds used during the k-folding technique to ensure a more reliable and robust validation process [19]. After repartitioning, each fold went through a data augmentation step, resulting in a total of 4,180 images summing all the samples in each fold. The test datasets remained unchanged without data augmentation.

Network Architecture. The neural network architecture used was U-Net [20] modified with the path encoder based on ResNet-34, as proposed in [32]. The standard U-net structure is present in the up-sampling decoding layers, while the down-sampling layer features the ResNet-34 model, containing a 7×7 convolutional layer and 64 filters. After that, there are 4 residual blocks, containing each block two 3×3 convolutional layers presenting *shortcuts* between each connection [32].

Each residual block has, respectively, 3 connections in the first block, 4 in the second, 6 in the third and 3 in the last, with each block of 3 connections having batch normalization, and an exponential increase in the number of filters by a factor of 2, thus starting with 64 filters and ending at 512 in the last block. Each decoding block contains a 2×2 convolution transposed layer obtained from the previous layer with a stride of 2, concatenated with a 1×1 convolution for each respective block contained in the down-sampling section. The already concatenated tensor goes through the batch normalization process before proceeding to the next decoding layer. Finally, in the last layer, the last transposed convolution occurs, which returns an image with the same dimensions as the input image. Figure 7 illustrates the complete architecture of the neural network used.

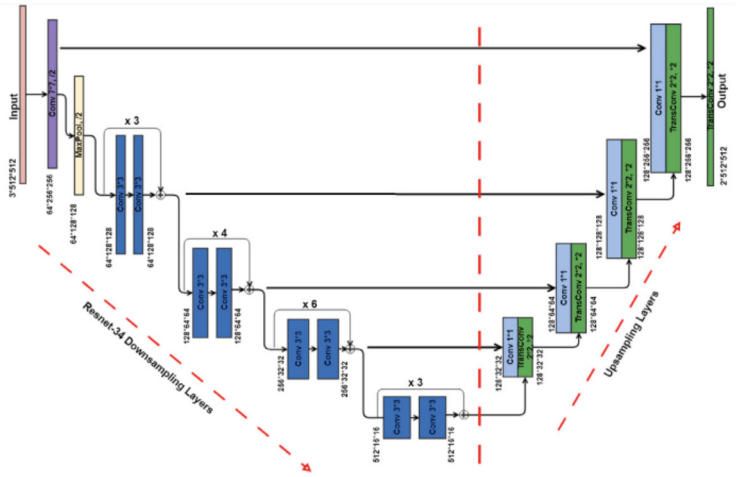

Fig. 7. Hybrid U-Net architecture with the ResNet-34 backbone. Adapted from [32]

3.3 Evaluation Metrics

We used as evaluation metrics the Dice coefficient (DiceScore) (Eq. 2), which aims to measure the similarity between two samples. In this work, the similarity between the generated mask prediction and the reference mask and the Jaccard index (JaccardScore) (Eq. 3) is calculated by dividing the pixels of the prediction concerning the reference mask of the image.

$$DiceScore = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{2}$$

$$JaccardScore = \frac{TP}{TP + FP + FN} \tag{3}$$

In which TP represents pixels correctly labeled as being from the optic cup or optic disc, FP represents pixels misclassified as the optic cup or optic disc, and FN pixels misclassified as not being optic disc or optic cup.

3.4 Experimental Set-Up

In the network training, binary entropy weighted cross entropy was chosen as the loss function [18]. Its purpose is to act with a better proportion of the weights contained in each image and its respective mask, assigning for each class (0 and 1) a corresponding weight. This function is described in Eq. 4. The choice of such a function made it possible to avoid class unbalances.

$$Loss = (\alpha \times \mathbf{t} \times \mathbf{p} \ + \alpha \times ((\mathbf{t} - 1) \times (\mathbf{p} - 1))) - (\alpha - 1) \tag{4}$$

where α remains at the negative value of -1, t represents the target value, the mask of the image according to experts, and p is a prediction image that was generated by the network.

Each *fold* was trained for 30 epochs with a batch size of value 4 with a learning rate set at 1e-4 with a decay of 0.5 every 10000 steps. In addition, the Early Stopping technique was used with the patience set to 3 epochs [13]. These parameters were set after performing training in addition to using the HyperOpt library [4]. All experiments were performed on hardware configured with NVIDIA GeForce GTX 1660 TI video card with 6 GB of RAM.

4 Results

Using the DRISHTI-GS and RIM-ONE test datasets, this section aims to demonstrate the values obtained to measure the accuracy of the model generated by the proposed method. The DRISHTI-GS dataset initially presented 50 images for training and 51 for testing, while the RIM-ONE had 159 images without division. After the augmentation, the number of images in the DRISHTI-GS and RIM-ONE datasets composing the training samples was 1,150 and 3,030, respectively.

Initially, experiments were performed to evaluate the performance of the chosen architecture, comparing the metrics obtained with a set of distinct architectures. It is important to note that the IOU metric in the following tables represents the JaccardScore equation seen in Sect. 3.3. Results for these experiments are shown in Table 1, for the DRISHTI-GS test dataset, and in Table 2, for the RIM-ONE test dataset. Importantly, these results were obtained without extracting Deep Features but using only the proportional division based on the image area. These architectures were chosen to measure the methodology's effectiveness on a deeper residual architecture with more layers (ResNet-50) and one with fewer training parameters (MobileNetV2).

Table 1. Optic disc and optic cup segmentation results. DRISHTI-GS test dataset. (*) means using proportional division based on area.

Architecture	Optic Disc		Optic Cup	
	Dice Score	IOU Score	Dice Score	IOU Score
U-Net with Resnet-50	95.4	92.2	89.0	78.9
U-Net with Resnet-50 (*)	96.0	**93.1**	90.2	80.4
U-Net with MobileNetV2	94.3	91.1	88.0	77.3
U-Net with MobileNetV2 (*)	95.0	91.0	89.2	79.1
U-Net with Resnet-34	96.0	92.2	89.5	79.9
U-Net with Resnet-34 (*)	**96.1**	92.4	**90.4**	**81.6**

After analyzing the values in both tables, it is possible to notice that the proposed U-Net model with ResNet-34 using proportional division is the most effective in segmenting the optic cup and the optic disc in the test sets of both datasets used in the experiments. Notably, the values obtained in the optic cup

Table 2. Optic disc and optic cup segmentation results. RIM-ONE test dataset. (*) means using proportional division based on area.

Architecture	Optic Disc		Optic Cup	
	Dice score	IOU score	Dice score	IOU score
U-Net with Resnet-50	96.0	92.0	84.6	72.5
U-Net with Resnet-50 (*)	96.0	**92.4**	85.0	73.3
U-Net with MobileNetV2	95.4	91.0	81.9	67.5
U-Net with MobileNetV2 (*)	96.0	91.1	83.4	68.7
U-Net with Resnet-34	95.2	91.3	85.1	73.4
U-Net with Resnet-34 (*)	**96.3**	91.8	**85.4**	**73.6**

segmentation stand out in both test samples, thus demonstrating that the application of proportional division proves beneficial regarding the results obtained in the tested networks.

For both optic cup and optic disc segmentation, the pre-trained U-Net model with the Resnet-34 backbone was used by modifying only the images for training and validation. Table 3 and Table 4 show the results obtained on the DRISHTI-GS and RIM-ONE test dataset, respectively, compared to the results of the works used as reference. It is important to point out that our work differs from the related works in a large number of parameters, from the number of folds used in the cross-validation, to the architecture and image base itself, so it is very difficult to make a direct comparison with each of the authors cited in the results tables. This time, there is the inclusion of the proposed methodology that couples proportional splitting with feature extraction and K-means clustering.

Table 3. Optic disc and optic cup segmentation results. DRISHTI-GS test dataset. (*) means using proportional division.

Method	Optic Disc		Optic Cup	
	Dice score	IOU score	Dice score	IOU score
[34]	97.3	91.4	87.1	85.0
[21]	–	–	85	75
[1]	94.9	90.4	82.8	71.1
[32]	96.4	94.2	88.7	78.0
[12]	–	–	95.0	79.0
[26]	91.6	–	73.3	–
[16]	93.0	**95.7**	93.3	**94.4**
[29]	97.0	94.7	95.0	93.2
[30]	**98.7**	93.2	**97.1**	92.1
Area based (*)	96.1	92.4	91.6	83.4
Deep Feature Extraction (*)	96.3	93.2	92.2	84.3

Table 4. Optic disc and optic cup segmentation results. RIM-ONE test dataset. (*) means using proportional division.

Method	Optic Disc		Optic Cup	
	Dice score	IOU score	Dice score	IOU score
[34]	94.2	89.0	82.4	80.2
[21]	95.0	89.0	82.0	69.0
[1]	90.3	82.8	69.0	55.6
[32]	96.1	**92.5**	84.4	74.2
[12]	96.1	88.3	**89.0**	76.2
[26]	87.3	–	75.9	–
[16]	93.2	88.4	78.6	65.9
[29]	-	92.0	–	**89.0**
Area based (*)	**96.4**	91.8	87.2	76.7
Deep Feature Extraction (*)	95.9	91.1	87.8	77.1

Analyzing the results shown in Table 3, the proposed method presents favorable and promising results, approaching the highest values of related works. It is important to note that in most cases, the proportional splitting with deep feature extraction and K-means clustering presented higher metrics when compared to the area method based, highlighting the gain in Dice and Jaccard results in the segmentation of the optic cup. Examples of the predictions generated by the proposed method on the DRISHTI-GS test dataset of both the optic cup and the optic disc can be seen in Fig. 8 and Fig. 9, respectively, while Fig. 10 presents the optic cup segmentation with the lowest Dice value obtained on the DRISHTI-GS test dataset, which was 83%.

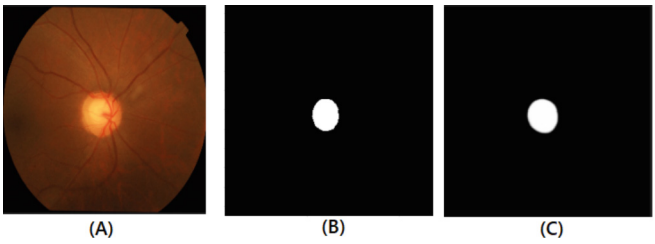

Fig. 8. Segmentation of the optic cup. With (A) meaning the original image, (B) the ground truth and (C) the prediction. Sample of the DRISHTI-GS test dataset.

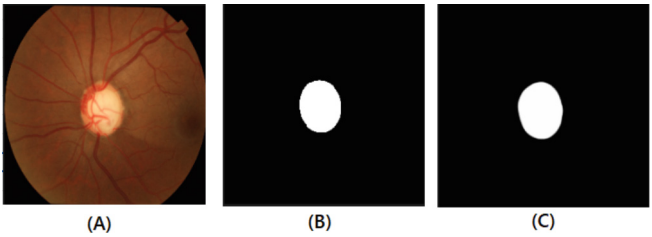

Fig. 9. Optic Disc Segmentation. With (A) meaning the original image, (B) the ground truth and (C) the prediction. Sample of the DRISHTI-GS test dataset.

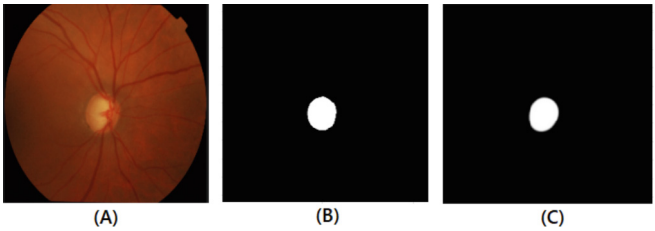

Fig. 10. Optic cup segmentation result with the smallest Dice value of 83%. With (A) meaning the original image, (B) the ground truth and (C) the prediction. Sample of the DRISHTI-GS test dataset.

For the RIM-ONE test dataset, the proposed model follows the same pattern seen in the DRISHTI-GS test samples, with results close to the highest values found in reference works and an increase in the metrics with the use of feature extraction for the segmentation of the optic cup. There was a decrease concerning the metrics obtained of 0.5% points for the Dice metric and 0.7% points for the Jaccard metric. Examples of predictions generated by the proposed method on the RIM-ONE test dataset from both the optic cup and the optic disc can be seen in Fig. 11 and Fig. 12, respectively, while Fig. 13 presents a sample that obtained 69.7% Dice value.

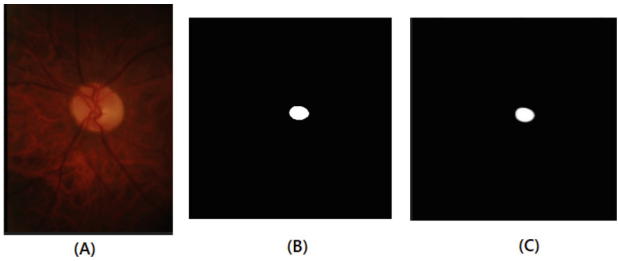

Fig. 11. Segmentation of the optic cup. With (A) meaning the original image, (B) the ground truth and (C) the prediction. Sample of the RIM-ONE test dataset.

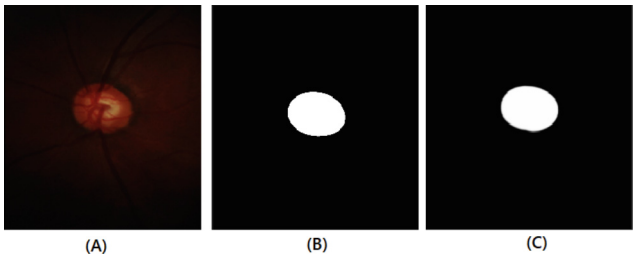

Fig. 12. Segmentation of the optic disc. With (A) meaning the original image, (B) the ground truth and (C) the prediction. Sample of the RIM-ONE test dataset.

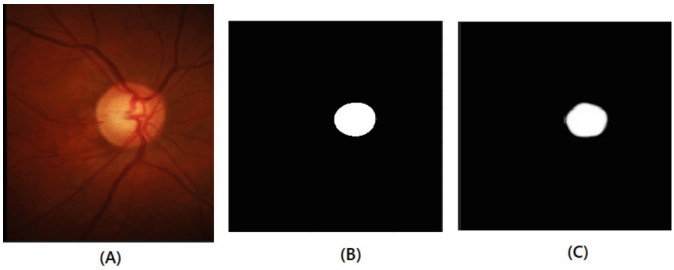

Fig. 13. Optic cup segmentation result with the smallest Dice value of 69.7%. With (A) meaning the original image, (B) the ground truth and (C) the prediction. Sample of the RIM-ONE test dataset.

5 Discussion

After analyzing the set of values obtained through the experiments performed, the application of proportional division showed promising results in the segmentation tasks, both of the disc and the optic cup, with emphasis on the latter, in which the values surpassed those of [32] that used the same architecture for the segmentation of the same structures. Notably, the results were achieved according to the best parameters that made it possible to find the set of best metrics until now.

Furthermore, integrating feature extraction and sample clustering using the K-means technique improved the optic cup and the optic disc segmentation. Overall, the experiments conducted with the U-Net model modified with ResNet-34, with proportional splitting, demonstrated through the results, show that a robust learning procedure using a distributed training set of samples assists in obtaining better results in the segmentation of the optic cup and the optic disc.

The effectiveness of using a balanced training set through a balanced division of the samples was shown. Even if there is room for improvement, our study presents robust results approaching the best values in the tables shown in the results section. However, our method cannot perform an utterly proportional division of the training samples when using the area method-based due to the

unbalanced nature of the RIM-ONE and DRISHTI-GS datasets. However, this methodology comes very close to the balanced division approach, which involves Deep Feature Extraction and K-Means clustering.

6 Conclusion

A good set of training samples is essential to obtain good results when segmenting a retinal structure. In this paper, we compare two methods that aim to work directly on the division of samples from the training set so that we have the most balanced training set at the end of the process. First, we used the division area-based method of the masks available for training. This method showed consistent results in segmenting the optic cup and the optic disc in both datasets, even though it presented an unbalanced sample rate. Next, the separation with Deep Feature Extraction and K-means clustering was used, which showed improvements in the metrics obtained by the previous methodology.

In future works, we intend to expand the set of images by adding new image dataset such as ORIGA [33] and REFUGE [17]. It is also intended to investigate new deep learning architectures, such as the use of a hybrid network of U-Net with EfficientNet that shows promising results in segmentation tasks [3] and, together, a model that also uses the U-Net with an architecture consisting of hierarchical bottlenecks that prove to be useful for multi-segmentation tasks [27].

Acknowledgement(s). The authors acknowledge the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001, Brazil, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA), Brazil, for the financial support.

Disclosure Statement. No potential conflict of interest was reported by the author(s).

References

1. Al-Bander, B., Williams, B.M., Al-Nuaimy, W., Al-Taee, M.A., Pratt, H., Zheng, Y.: Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis. Symmetry **10**(4), 87 (2018)
2. Araújo, J.D.L., et al.: Glaucoma diagnosis in fundus eye images using diversity indexes. Multimed. Tools Appl. **78**(10), 12987–13004 (2019)
3. Baheti, B., Innani, S., Gajre, S.S., Talbar, S.N.: Eff-UNet: a novel architecture for semantic segmentation in unstructured environment. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1473–1481 (2020)
4. Bergstra, J., Yamins, D., Cox, D.D., et al.: Hyperopt: a python library for optimizing the hyperparameters of machine learning algorithms. In: Proceedings of the 12th Python in Science Conference, vol. 13, p. 20. Citeseer (2013)

5. Bilal, A., Sun, G., Mazhar, S., Imran, A., Latif, J.: A transfer learning and U-net-based automatic detection of diabetic retinopathy from fundus images. In: Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, pp. 1–12 (2022)
6. Bourne, R., et al.: Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the global burden of disease study. Lancet Glob. Health **9**(2), e130–e143 (2021)
7. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: fast and flexible image augmentations. Information **11**(2), 125 (2020)
8. Claro, M., et al.: An hybrid feature space from texture information and transfer learning for glaucoma classification. J. Vis. Commun. Image Represent. **64**, 102597 (2019)
9. da Cruz, L.B., et al.: Kidney tumor segmentation from computed tomography images using DeepLabv3+ 2.5 D model. Expert Syst. Appl. **192**, 116270 (2022)
10. Fumero, F., Sigut, J., Alayón, S., González-Hernández, M., González de la Rosa, M.: Interactive tool and database for optic disc and cup segmentation of stereo and monocular retinal fundus images. In: 23rd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2015), pp. 91–97. Václav Skala-UNION Agency (2015)
11. Hartigan, J.A., Wong, M.A.: Algorithm as 136: a k-means clustering algorithm. J. Roy. Stat. Soc. Ser. C (Appl. Stat.) **28**(1), 100–108 (1979)
12. Joshua, A.O., Nelwamondo, F.V., Mabuza-Hocquet, G.: Segmentation of optic cup and disc for diagnosis of glaucoma on retinal fundus images. In: 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), pp. 183–187. IEEE (2019)
13. Li, M., Soltanolkotabi, M., Oymak, S.: Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks. In: International Conference on Artificial Intelligence and Statistics, pp. 4313–4324. PMLR (2020)
14. Lima, A., Júnior, G.B., de Almeida, J.D., de Paiva, A.C., Veras, R.: An automated CNN architecture search for glaucoma diagnosis based on neat. Multimed. Tools Appl. **81**(10), 13441–13465 (2022)
15. Lima, A., Maia, L.B., dos Santos, P.T.C., Junior, G.B., de Almeida, J.D., de Paiva, A.C.: Evolving convolutional neural networks for glaucoma diagnosis. In: Anais do XVIII Simpósio Brasileiro de Computação Aplicada à Saúde. SBC (2018)
16. Mangipudi, P.S., Pandey, H.M., Choudhary, A.: Improved optic disc and cup segmentation in glaucomatic images using deep learning architecture. Multimed. Tools Appl. **80**(20), 30143–30163 (2021)
17. Orlando, J.I., et al.: Refuge challenge: a unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Med. Image Anal. **59**, 101570 (2020)
18. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
19. Rodriguez, J.D., Perez, A., Lozano, J.A.: Sensitivity analysis of k-fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 569–575 (2009)

20. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
21. Sevastopolsky, A.: Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. Pattern Recognit Image Anal. **27**(3), 618–624 (2017)
22. Shutaywi, M., Kachouie, N.N.: Silhouette analysis for performance evaluation in machine learning with applications to clustering. Entropy **23**(6), 759 (2021)
23. Sivaswamy, J., Krishnadas, S., Chakravarty, A., Joshi, G., Tabish, A.S., et al.: A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomed. Imaging Data Pap. **2**(1), 1004 (2015)
24. Sivaswamy, J., et al.: Drishti-GS: Retinal image dataset for optic nerve head (ONH) segmentation. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp. 53–56 (2014). https://api.semanticscholar.org/CorpusID:18432155
25. Steinmetz, J.D., et al.: Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to vision 2020: the right to sight: an analysis for the global burden of disease study. Lancet Glob. Health **9**(2), e144–e160 (2021)
26. Sun, G., et al.: Joint optic disc and cup segmentation based on multi-scale feature analysis and attention pyramid architecture for glaucoma screening. Neural Comput. Appl. 1–14 (2021)
27. Tang, S., Qi, Z., Granley, J., Beyeler, M.: U-net with hierarchical bottleneck attention for landmark detection in fundus images of the degenerated retina. In: Fu, H., Garvin, M.K., MacGillivray, T., Xu, Y., Zheng, Y. (eds.) OMIA 2021. LNCS, vol. 12970, pp. 62–71. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87000-3_7
28. Tatham, A.J., Weinreb, R.N., Medeiros, F.A.: Strategies for improving early detection of glaucoma: the combined structure-function index. Clin. Ophthalmol. **8**, 611–621 (2014)
29. Tulsani, A., Kumar, P., Pathan, S.: Automated segmentation of optic disc and optic cup for glaucoma assessment using improved UNet++ architecture. Biocybern. Biomed. Eng. **41**(2), 819–832 (2021)
30. Veena, H., Muruganandham, A., Kumaran, T.S.: A novel optic disc and optic cup segmentation technique to diagnose glaucoma using deep learning convolutional neural network over retinal fundus images. J. King Saud Univ.-Comput. Inf. Sci. **34**(8), 6187–6198 (2022)
31. Wu, Y., et al.: Measures of disease activity in glaucoma. Biosens. Bioelectron. **196**, 113700 (2022)
32. Yu, S., Xiao, D., Frost, S., Kanagasingam, Y.: Robust optic disc and cup segmentation with deep learning for glaucoma detection. Comput. Med. Imaging Graph. **74**, 61–71 (2019)
33. Zhang, Z., et al.: ORIGA-light: an online retinal fundus image database for glaucoma analysis and research. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pp. 3065–3068. IEEE (2010)
34. Zilly, J., Buhmann, J.M., Mahapatra, D.: Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. Comput. Med. Imaging Graph. **55**, 28–41 (2017)