Computer Methods and Programs in Biomedicine 192 (2020) 105341

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Offline computer-aided diagnosis for Glaucoma detection using fundus images targeted at mobile devices

José Martins[a], Jaime S. Cardoso[b], Filipe Soares[a,*]

[a] Fraunhofer Portugal AICOS, Rua Alfredo Allen 455/461, Porto 4200-135, Portugal
[b] INESC TEC and Faculty of Engineering of the University of Porto, Portugal

ARTICLE INFO

Article history:
Received 22 July 2019
Revised 14 December 2019
Accepted 14 January 2020

Keywords:
Computer-aided diagnosis (CAD)
Fundus images
Glaucoma
Deep learning

ABSTRACT

Background and Objective: Glaucoma, an eye condition that leads to permanent blindness, is typically asymptomatic and therefore difficult to be diagnosed in time. However, if diagnosed in time, Glaucoma can effectively be slowed down by using adequate treatment; hence, an early diagnosis is of utmost importance. Nonetheless, the conventional approaches to diagnose Glaucoma adopt expensive and bulky equipment that requires qualified experts, making it difficult, costly and time-consuming to diagnose large amounts of people. Consequently, new alternatives to diagnose Glaucoma that suppress these issues should be explored.

Methods: This work proposes an interpretable computer-aided diagnosis (CAD) pipeline that is capable of diagnosing Glaucoma using fundus images and run offline in mobile devices. Several public datasets of fundus images were merged and used to build Convolutional Neural Networks (CNNs) that perform segmentation and classification tasks. These networks are then used to build a pipeline for Glaucoma assessment that outputs a Glaucoma confidence level and also provides several morphological features and segmentations of relevant structures, resulting in an interpretable Glaucoma diagnosis. To assess the performance of this method in a restricted environment, this pipeline was integrated into a mobile application and time and space complexities were assessed.

Results: Considering the test set, the developed pipeline achieved 0.91 and 0.75 of Intersection over Union (IoU) in the optic disc and optic cup segmentation, respectively. With regards to the classification, an accuracy of 0.87 with a sensitivity of 0.85 and an AUC of 0.93 were attained. Moreover, this pipeline runs on an average Android smartphone in under two seconds.

Conclusions: The results demonstrate the potential that this method can have in the contribution to an early Glaucoma diagnosis. The proposed approach achieved similar or slightly better metrics than the current CAD systems for Glaucoma assessment while running on more restricted devices. This pipeline can, therefore, be used to construct accurate and affordable CAD systems that could enable large Glaucoma screenings, contributing to an earlier diagnose of this condition.

1. Introduction

Glaucoma is a group of chronic eye diseases and a leading cause of irreversible blindness worldwide [1].

The rise of intraocular pressure (IOP) inside the eye is considered the main cause of Glaucoma. This pressure damages the optic nerve and is usually related to an inability of the eye to properly manage the balance between the amount of fluid that is produced and the amount that is drained [2]. Although it is confirmed that the raised intraocular pressure is a causal risk factor for Glaucoma and the only one that can be treated, it is neither sufficient nor necessary for the diagnosis of the disease [3].

In the majority of Glaucoma cases, no early symptoms or pain occur, and for that reason, Glaucoma is often called a 'silent thief of sight'.

Furthermore, the usual approaches to diagnose Glaucoma are performed by ophthalmologists and consist of a comprehensive eye examination. This encompasses an external examination of the eye, ocular mobility, examination of the pupil, slit lamp examination, tonometry, gonioscopy, and perimetry [3]. These methods require trained professionals and expensive equipment; therefore, due to the high costs incurred by them, some are only performed

* Corresponding author.
E-mail address: filipe.soares@fraunhofer.pt (F. Soares).

https://doi.org/10.1016/j.cmpb.2020.105341
0169-2607/

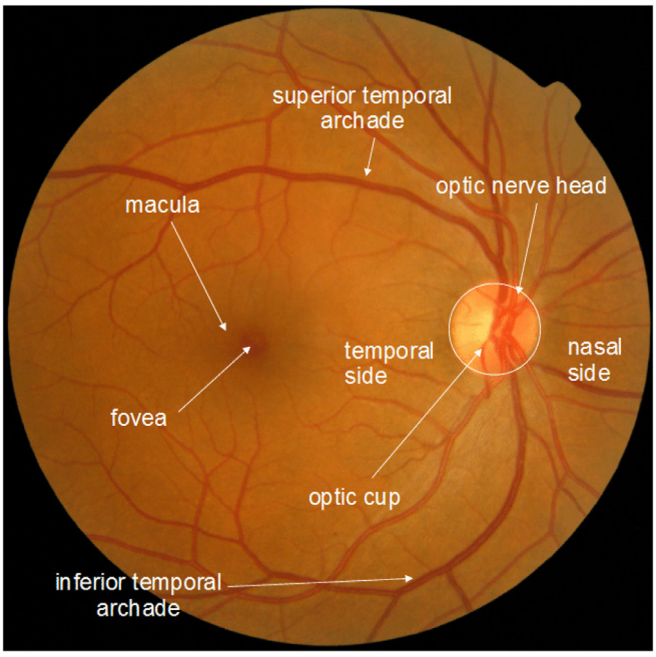

Fig. 1. Fundus image [5].

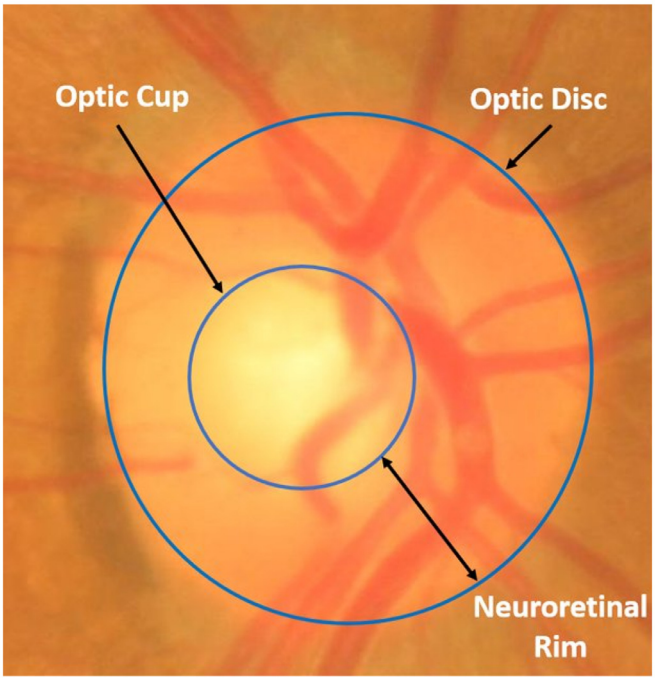

Fig. 2. Fundus Image ROI [6].

if Glaucoma is already suspected, which contributes to an increasing number of undiagnosed Glaucoma cases.

This number of undiagnosed cases is a big concern in the healthcare community since, despite being irreversible, Glaucoma can be effectively slowed down with appropriate treatment, pharmaceutical or surgical [4], therefore reducing the risk of total blindness. Hence, achieving an early diagnosis of Glaucoma is of utmost importance. This results in the demand for new diagnose techniques that aid or replace the existing ones, turning Glaucoma assessment into a more efficient process.

To respond to this demand, fundus images can be used. These images, illustrated in Fig. 1, have the necessary morphological features to diagnose several eye conditions, like Glaucoma, the topic of this work.

The region of interest (ROI), illustrated in Fig. 2, shows the most relevant structures for Glaucoma diagnosis present in fundus images. These structures are the optic disc and optic cup.

This work contributes to improve Glaucoma management by proposing an interpretable pipeline for Glaucoma detection powered by fundus images that can run offline on mobile devices and achieves better or comparable results to the state-of-the-art techniques in classification and segmentation tasks.

2. Related work

The first relevant implementations of CAD systems for Glaucoma detection with fundus images started to arise in 2008 [2]. Since then a significant collection of approaches has been proposed.

In 2016, A. Singh et al. [7] achieved an accuracy of 0.947. They started by identifying the center of the disc, performing the disc segmentation afterward. Then, the intuition that blood vessels represent noisy pixels that affect the performance of the system led to the removal of these vessels from the obtained optic disc images. After this step, feature extraction was accomplished using first level discrete wavelet decomposition that resulted in a feature vector of 18 features. For feature selection, two approaches were tried, genetic algorithms and Principal Component Analysis (PCA). The final feature vectors of each technique were then tested with several classifiers (Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Naive Bayes, and Artificial Neural Networks (ANN)). Since the fundus images are from a local dataset of 63 images, which is rather small, leave-one-out cross-validation was also performed to account for overfitting issues. The best performing models were the SVM and KNN, with PCA feature selection that resulted in only 2 components. Both of these classifiers obtained 0.947 accuracies.

More recently, Deep Learning approaches are proving to be capable of surpassing the existing techniques and are being heavily applied.

H. Fu et al. [1] proposed a new segmentation approach of the optic disc and cup. The novelty of this work consists in the joint segmentation of the optic disc and optic cup since most approaches achieve this segmentation separately. To deal with this joint segmentation the work also proposes a new joint loss metric. The first step of this method is the transformation of the fundus image to the polar coordinate system, in order to enlarge the cup representation. Next, the segmentation is performed with a CNN, consisting of four key parts, the multi-scale input layer, U-shape architecture, a side output layer, and the mentioned joint loss metric. Several tests were performed in different conditions and it was concluded that the polar transformation attains a relevant performance gain. Glaucoma screening was then performed calculating the cup to disc ratio value and an AUC of 0.899 in the private SCES dataset was obtained.

A more recent approach by Z. li et al. [8] uses the Inception-v3 architecture, that has already proven its success in other tasks. The images were resized to 299 $\times$ 299 and the RGB values were normalized between 0 and 1 before feeding the images to the network. A minibatch gradient descent of 32 was used in combination with the ADAM optimizer and a 0.002 learning rate. The key aspect of this approach is its dataset since a total of 70000 fundus images were downloaded from an unlabeled dataset and 48116 of those images were selected after verifying that they include a visible optic disc. After this, 21 ophthalmologists performed the classification of these images in an online platform and multiple interpretations of each image were obtained for quality control. The result was

Table 1
Image count per dataset. G and NG represent Glaucoma and No-Glaucoma, respectively.

Dataset	G	NG	Total
Origa [9]	168	482	650
Drishti-GS [10]	70	31	101
RIM-ONE r1 [11]	40	118	158
RIM-ONE r2 [12]	200	255	455
RIM-ONE r3 [12]	148	170	318
iChallenge [13]	80	720	800
RIGA [14]	749		749

39745 images, where 31745 were used as a training set and 8000 as a validation set. The system achieved an AUC of 0.986, 0.956 sensitivity and 0.920 specificity, which demonstrates the relevance of the dataset size and provides further evidence about the success of deep learning techniques for Glaucoma detection.

These approaches, however, do not target mobile devices and despite the better results achieved recently with deep learning techniques, these solutions have little or no concerns with respect to the interpretability of the decision.

3. Proposed methodology

The proposed methodology explores deep learning techniques to build a Glaucoma assessment pipeline with a focus on interpretability and space and time complexities.

3.1. Datasets

To construct a robust dataset for the two main tasks of this work, segmentation and classification, several publicly available datasets, listed in Table 1, were merged.

Firstly, a dataset for segmentation tasks of both the optic disc and optic cup was created merging the Origa, Drishti-GS, iChallenge, RIM-ONE r3, and RIGA datasets, resulting in a total of 2618 images. In the datasets where several annotations were provided for the same image, the considered ground truth was the region of agreement between the annotators.

Lastly, for the classification task, Origa, Drishti-GS, RIM-ONE (r1,r2,r3) and iChallenge were merged, resulting in 2482 images. Furthermore, only two classes were considered, Glaucoma and No-Glaucoma, since most datasets only included a binary classification for Glaucoma.

Additionally, when considering the RIM-ONE r3 dataset, which consisted of stereo images, each of these images was divided into two, and considered as a separate case. With this in mind, the number of images in this dataset increased from 159 to 318.

3.2. Augmentation

Real-time data augmentation was used since it is robust and easy to implement, not requiring additional disk space to save the augmented images and providing different images in every batch of data.

The augmentation pipeline consists of four main steps that are applied in random order and from augmentation to augmentation some of them might even be skipped. The first step consists in the application of blur to the image, this blur can be either a gaussian, average or a median blur. Secondly, contrast normalization is performed by moving the pixel values away or closer to mid-scale. Then, changes in the brightness and sharpness of the images are also employed.

3.3. Segmentation

When considering Glaucoma, the optic disc and optic cup are the most important structures to segment, since they can power the calculation of several morphological features and the crop of the ROI in the full fundus image.

To target the first task, an architecture entitled GFI-ASPP-Depth was developed to perform joint segmentation of the optic disc and optic cup in a ROI image.

This network has a U-shaped structure where the encoding path consists of four depth levels. In each depth level, two depth-wise separable convolution blocks, inspired in the MobileNet architectures [15,16], are employed with 3×3 kernels, the number of filters starts at 32 and is then doubled on every transition to the next depth level.

The transition to the next depth level is performed through average pooling with a pool size of 2. Average pooling was selected in detriment of max pooling since the literature reports better results on similar tasks [17]. Furthermore, two inputs are used in the network in the first two depth levels and skip connections between the encoder and decoder path are present at every depth level to provide more spatial context to the decoding path.

When it comes to the transition from the encoder to the decoder path, an ASPP module [18–20] was chosen, consisting of four parallel padded atrous convolutions with dilation rates of 1, 2, 4 and 7. Additionally, image-level features obtained by performing a global average pooling on the original feature map are also used. This is done because when the dilation rate increases, the number of filters weights that are applied to the valid feature region (region without the padded zeros) declines. The inclusion of image-level features helps to mitigate this problem and include more global context in the network [19].

Then, the result of each level of the ASPP module and the image-level features are concatenated and the decoding path starts. This is very similar to the encoding path, however, in here, the transition to an upper level leads to the decrease in the number of filters in each convolution by a factor of 2 and is done with nearest interpolation. Finally, in the first two depth levels of the decoder, two outputs are generated and then averaged to form the final segmentation prediction.

This architecture consists of a total of 1,152,131 parameters and is represented in Fig. 3.

Additionally, given that the segmentation of the optic disc is a simpler task than the joint optic disc/cup segmentation or the optic cup segmentation alone, a lighter and smaller architecture entitled GFI-SPP-Depth-simple was created. This network only performs optic disc segmentation that can then be used to quickly obtain a ROI image.

This network is similar to GFI-ASPP-Depth, but starts with 16 filters in the first depth layer instead of 32 that then increases and decreases in the same way as GFI-ASPP-Depth. Moreover, the transition between the encoder and decoder path consists solely in the concatenation of global average pooling performed on the original feature map, the result of a convolution block applied to the original feature map and the original feature map itself. Resulting in an architecture with 271,929 parameters.

Both networks were trained using an Nvidia GPU Tesla V100 after resizing the images to 224×224 and performing data normalization. The multi-label dice loss was selected [1], together with the ADAM optimizer with a learning rate of 0.0001. Additionally, the learning rate was reduced during the learning process by a factor of 2 if the validation loss was not improved in the last 10 epochs. Furthermore, the model was trained for 200 epochs, with early stopping if no improvements are verified in the validation loss after 20 epochs to mitigate overtraining.

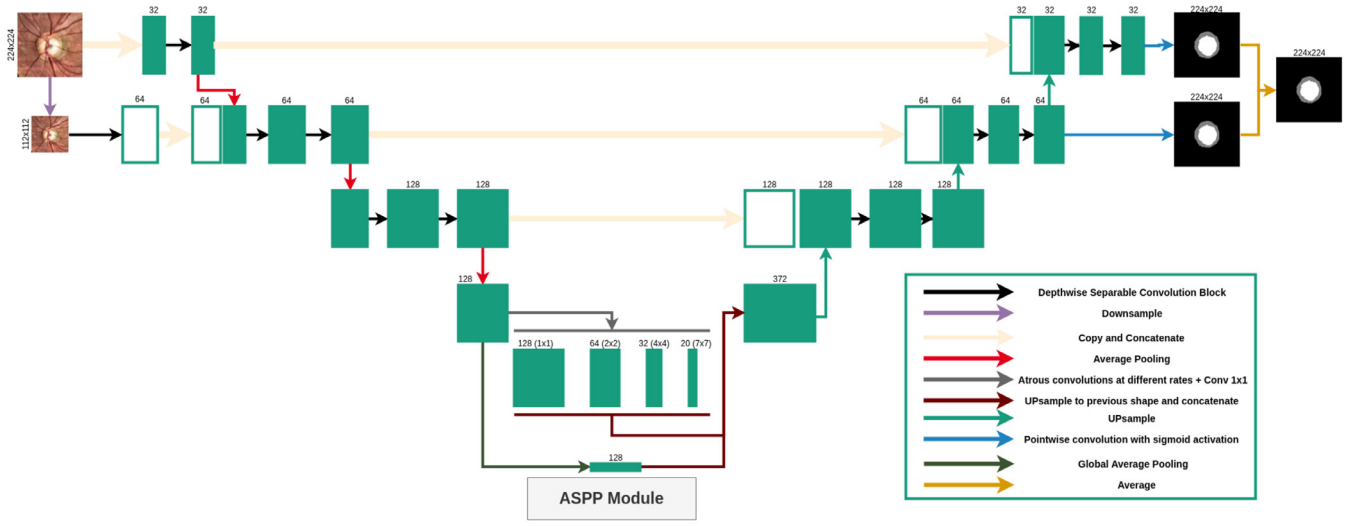

Fig. 3. GFI-ASPP-Depth architecture diagram.

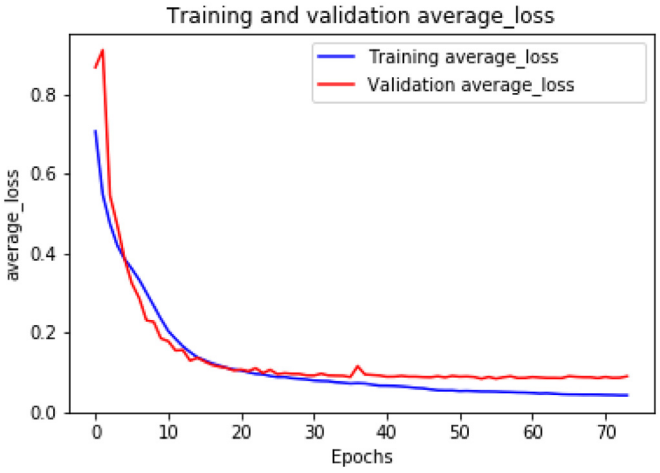

Fig. 4. Evolution of the loss during a GFI-ASPP-Depth training session.

Table 2
Segmentation performance comparison with state-of-the-art methods trained with the ORIGA dataset. Achieved results obtained on the test set of the dataset described in 3.1.

Method	IoU disc	IoU cup
Transcribed from [1]		
R-Bend [22]	0.871	0.605
ASM [23]	0.882	0.687
Superpixel [24]	0.898	0.736
LRR [25]	–	0.756
QDSVM [26]	0.89	–
U-net [27]	0.885	0.713
M-net [1]	0.917	0.715
M-net + PT [1]	0.929	0.77
Achieved results		
M-net (self-trained) [1]	0.87	0.70
GFI-ASPP-Depth	0.91	0.74
GFI-SPP-Depth-Simple	0.89	–

Considering GFI-ASPP-Depth in particular, the input images fed to the network consist of the ROI images enhanced with CLAHE [21] and the model was trained with a batch size of 16. This model was used for joint segmentation of the optic disc and optic cup, both with 0.5 weights in the multi-label dice loss.

In the GFI-SPP-Depth-simple case, a batch size of 16 was again used, but the input images were the center cropped full fundus images.

The validation approach consists of a stratified train/val/test split with 80/10/10 proportions of the original dataset. Fig. 4 discloses the evolution of the training and validation loss for a given split in a training session with the GFI-ASPP-Depth model. The type of convergence was identical in the GFI-SPP-Depth-simple model.

Moreover, Table 2 reveals the achieved results of the developed networks along with the results of other methods from the literature.

One issue when comparing these models with the literature is that the used datasets and, consequently, the validation approaches often differ significantly, making a direct comparison unfair in some cases. To mitigate this issue, the M-net model [1] without polar transformation (PT), was trained with the segmentation dataset used in the developed models. Afterward, the results of this model can be directly compared with GFI-ASPP-Depth and GFI-SPP-Depth-Simple.

Furthermore, the state-of-the-art results extracted from [1], compare conventional methodologies with deep learning approaches and are comparable between themselves. Hence, by comparing the obtained M-net result with the result in the referred paper and then with the other state-of-the-art methods, it is possible to establish a point of comparison between all models.

The results indicate that the developed models outperform the most recent methods while being considerably faster and smaller. For instance, in comparison with the M-net model, the number of parameters is around 7 times smaller in GFI-ASPP-Depth and 31 times smaller in GFI-SPP-Depth-Simple.

Regarding wrong segmentations obtained by the developed models, the lack of contrast and the presence of reflections in the fundus image were identified as the main causes.

3.4. Morphological features

The segmentation predictions of the optic disc and cup obtained by GFI-ASPP-Depth can then be used to calculate several morphological features.

In this work, the cup to disc ratio was calculated based on both the vertical length (VCDR) and the area of the cup and the disc (CDR). The enlargement of this ratio is known as 'cupping' [2] and can be classified into mild (CDR up to 0.4), moderate (CDR be-

tween 0.5 and 0.7) and severe (CDR above 0.7) Glaucoma classes [4]. Rim to disc area ratio (RDAR), was also calculated as it also provides an interpretation of the optic nerve head shape.

Furthermore, the ISNT rule states that, in normal eyes, the Neuro-Retinal Rim (NRR) width in the inferior section (I) of the ROI is greater than in the superior section (S) which in turn is greater than the area in the nasal section (N) which in turn is greater than the area in the temporal section (T); this can be translated by the following inequation.

$$I > S > N > T \quad (1)$$

It has been declared that the compliance of this rule is useful in differentiating normal from Glaucomatous optic nerves and is not affected by race [28]. Therefore, ISNT values and rule compliance were also computed.

3.5. Classification

To obtain a Glaucoma confidence level, a classification network, entitled GFI-C, was created using MobileNetV2 [16] feature extractor as a backbone.

After the MobileNetV2 block, global average pooling is performed to flat the feature maps and then two fully connected layers, interleaved by heavy dropouts, are performed, resulting in an architecture with 2,299,521 parameters.

To better assess the impact of the MobileNetV2 backbone, other backbones have also been experimented, namely, VGG16 and VGG19 [29], InceptionV3 [30] and ResNet50 [31].

All the models were trained using a Nvidia GPU Tesla V100, batch size of 64, learning rate of 0.00001, and resized 224x224 ROI cropped images with CLAHE transformation as input. The validation approach was the same followed in the segmentation models. Furthermore, callbacks to reduce the learning rate by a factor of 2 after 10 epochs and interrupt the training session after 20 epochs, if no improvements are achieved in the validation loss, were also used to prevent overtraining. With regards to the backbones, ImageNet weights were used on all networks and all the layers were trained.

Fig. 5 reveals the evolution of the training and validation loss for a given split in a training session with the GFI-C architecture.

Table 3 discloses the achieved results on the test set and the results of other methods from the literature. The proposed approach accomplishes very similar metrics to the most recent method (DL 6), which consists of a network with around 30 million parameters, and is, therefore, more suitable for integration in computational restricted devices.

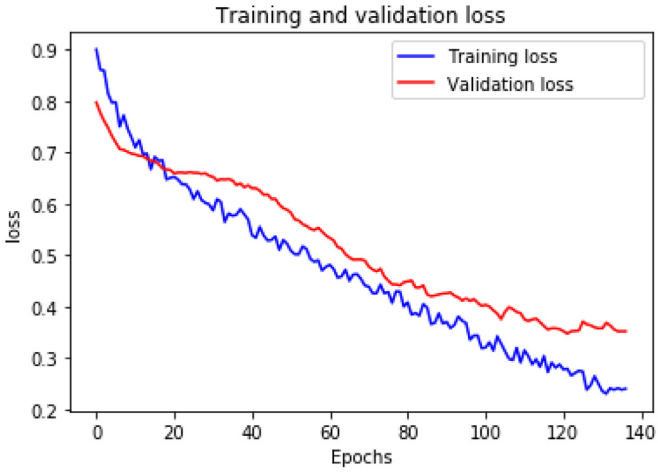

Fig. 5. Evolution of the loss during GFI-C training session.

3.6. CAD pipeline

The developed segmentation and classification models along with the calculation of the morphological features were then assembled to construct an interpretable Glaucoma CAD pipeline.

The pipeline, illustrated in Fig. 6, starts with an acquired full fundus image; this image is then center cropped and the GFI-SPP-Depth-simple model is employed to segment the disc in step *'(1)'*. This segmentation is then used to transform the full fundus image in a ROI image during step *'(2)'*, followed by the CLAHE transformation. The CLAHE ROI image is then utilized for two different tasks that build the pipeline outputs. First, in step *'(3)'*, GFI-ASPP-Depth is used to obtain an optic disc and cup segmentation. This powers the calculation of several morphological features, that contribute to the interpretability of the decision. Additionally, step *'(5)'*, consists of running the images outputted from step *'(2)'* through the GFI-C network, resulting in a Glaucoma confidence level.

Finally, the decision is built, consisting of the Glaucoma confidence level and the calculated morphological features.

3.7. Mobile integration

To assess the performance of the developed pipeline in a restricted environment, it was integrated into a mobile application.

Table 3
Classification performance comparison with state-of-the-art methods. In the methods column, *ML* and *DL*, indicate whether the method consists in a machine learning or deep learning approach, respectively. Considering the datasets column, the number of No-Glaucoma cases is marked with '–' and the number of Glaucoma cases is marked with '+'. The metrics Acc, Sens and Spec stand for accuracy, sensitivity, and specificity, respectively. Achieved results obtained on the test set of the dataset described in 3.1.

Method	Datasets	Acc	Sens	Spec	AUC
Transcribed from [32]					
ML 1 - Maheshwari et al. [33]	RIM-ONE r2 + Private (280+/285–)	0.81-0.98	–	–	–
ML 2 - Acharya et al. [34]	Private (559+/143–)	0.95	–	–	–
DL 1 - Al-Bander et al. [35]	RIM-ONE r2 (200+/255–)	0.88	0.85	0.89	–
DL 2 - Fu et al. [36]	ORIGA (168+/482–)	–	0.84	0.92	0.91
DL 3 - Li et al. [8]	Private (48116)	–	0.95	0.92	0.98
DL 4 - Christopher et al. [37]	Private (5633+/9189–)	–	0.88	0.95	0.91
DL 5 - Mitsuhashi et al. [38]	Private (1364+/1768–)	–	–	–	0.96
DL 6 - Antón et al. [32]	Private + RIM-ONE r1,r2,r3 + Drishti-GS (494+/1819–)	0.88	0.87	0.89	0.94
Achieved results					
DL 7 - **GFI-C**	Dataset in 3.1 (706+/1776–)	0.87	0.85	0.88	0.93
DL 8 - **GFI-C-VGG16**	Dataset in 3.1 (706+/1776–)	0.86	0.81	0.88	0.95
DL 9 - **GFI-C-VGG19**	Dataset in 3.1 (706+/1776–)	0.87	0.83	0.89	0.95
DL 10 - **GFI-C-InceptionV3**	Dataset in 3.1 (706+/1776–)	0.90	0.80	0.94	0.95
DL 11 - **GFI-C-ResNet50**	Dataset in 3.1 (706+/1776–)	0.88	0.83	0.90	0.93

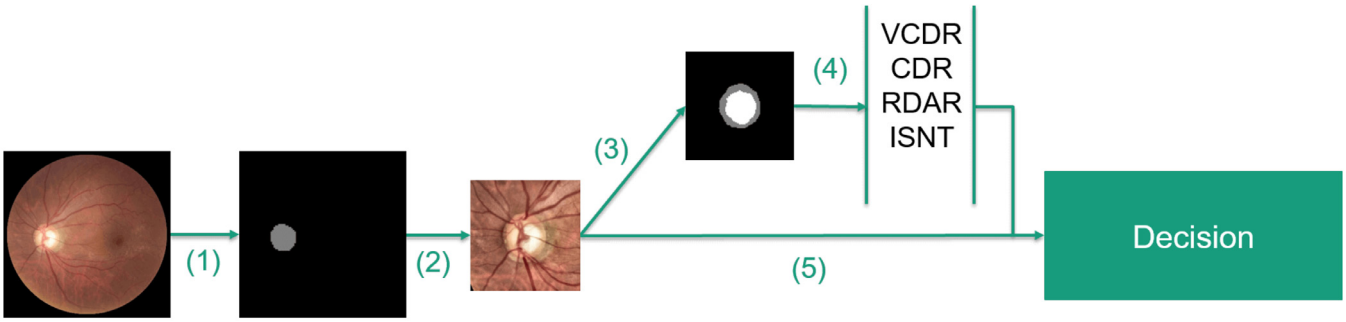

Fig. 6. Proposed pipeline.

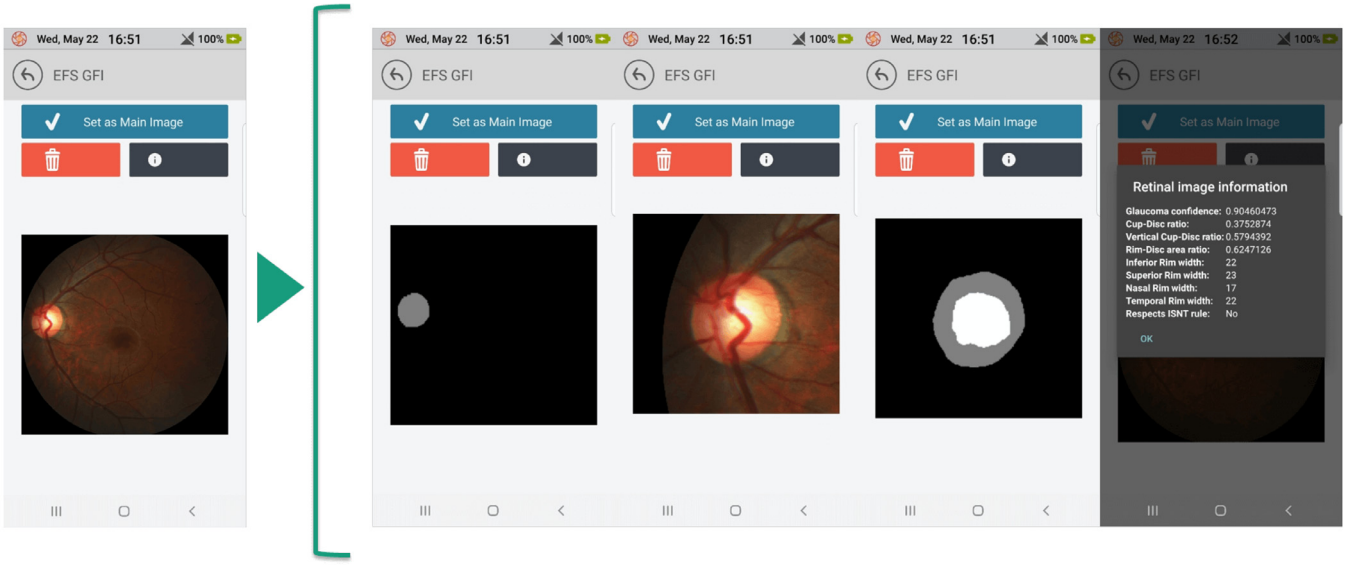

Fig. 7. Pipeline integration in a mobile environment.

Table 4
Metrics of the models used in the Glaucoma assessment pipeline (Obtained on a Samsung Galaxy S8).

Model	CPU (ms)	GPU (ms)	Size (mb)
GFI-SPP-Depth-simple	≈ 250	≈ 215	1.1
GFI-SPP-Depth-simple_quantized	≈ 280	≈ 220	0.3
GFI-ASPP-Depth	≈ 580	≈ 450	4.6
GFI-ASPP-Depth_quantized	–	–	1.7
GFI-C	≈ 80	≈ 25	9.0
GFI-C_quantized	≈ 110	≈ 120	2.3

This integration consisted on the creation of a background service that used TensorFlow Lite to perform inference with the developed models on a mobile environment. This service is run for a given fundus image and notifies the user when the pipeline is completed.

Afterward, the user can see the pipeline results, including the first disc segmentation, the cropped image and the joint optic disc/cup segmentation. Furthermore, the result of the morphological feature calculation, as well as the glaucoma confidence level, are also exposed in a dialog view. This visualization of the pipeline results contributes to the interpretability of the Glaucoma diagnosis, as can be seen in Fig. 7.

In Table 4, several time and size metrics of the developed classification and segmentation models after being converted to the TensorFlow Lite format are exposed. The time metrics were obtained by performing 3 runs in isolation on the same mobile device and averaging times.

By analyzing the table, it is concluded that running the models on GPU is preferred than quantizing and running on the CPU since the times are better and no performance losses will occur. The selected models are highlighted in the table, and the pipeline can run in under two seconds per fundus image.

3.8. Interpretability

Interpretability is one of the main Achilles heels of deep learning models, since, despite having exhibited superior performances in plenty of tasks, it is difficult to make sense of these types of models when compared to machine learning approaches [39]. This mainly occurs because the feature representation is not engineered by the developer, but by the network itself, turning its translation into representative values challenging.

Also, this topic is not only important to further understand the inner workings of a model and ease the development process, but it is also essential, or even mandatory, for regulations and other bureaucracies, needed in case these methods are intended to be integrated into commercial solutions, especially in a medical context.

The proposed solution has three interpretability measures, some of them have already been partially addressed, but below a complete description is provided.

Pipeline intermediate results: These results are evidenced in Fig. 7 and were integrated into the system to allow quick confirmation of the pipeline outcome, since wrong segmentations may

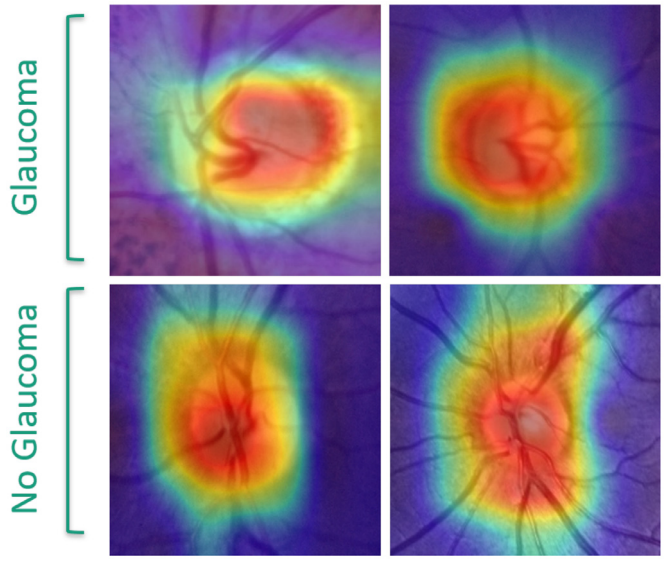

Fig. 8. Grad-cam activation maps of the GFI-C network.

eventually happen. Having the possibility to visualize those results enables the user to discard them right away, if necessary.

Morphological features: These features make the model more interpretable since they have a defined and studied meaning among the ophthalmologist community. Despite not being directly used in the model that outputs the Glaucoma confidence level, the combination of the two output values turns the decision more trustworthy.

Activation maps: The used activation maps are gradient class activation maps (Grad-CAM) [40]. These maps were calculated in the GFI-C network and an illustration of the results with images from both classes can be found in Fig. 8. In the figure, the images are overlayed with the calculated Grad-CAM maps and blue tones indicate the area was not important to the classification, while redder tones indicate a bigger influence of that region for the final decision. By analyzing the images, it is noticeable that the network is looking to the same structures that ophthalmologists inspect, such as the optic disc, cup, retinal vessels topology, and more interesting, on the left Glaucoma case, a focus on the Peripapillary Atrophy (PPA) region is visible, which is an identified Glaucoma risk factor [2,41–43]. All these insights contribute to increase the confidence in the network predictions. These maps are not available in the mobile pipeline integration due to Tensorflow Lite limitations, since the library is only targeted at inference. Nonetheless, they are available in a desktop environment and can be useful to make the pipeline more believable and help in regulatory processes.

4. Conclusions

The main novelty of this work is the fact that the developed Glaucoma assessment pipeline runs offline in mobile devices in seconds.

Moreover, this was achieved while having comparable, and sometimes even better, results in the segmentation and classification tasks of the developed system, contributing to the enhancement of the state-of-the-art in these topics as well.

Despite the aforementioned, the used dataset still poses some challenges, such as class imbalance, and the limited number of samples for deep learning approaches. Therefore, new experiments with more data should be investigated and evaluated to confirm this line of work.

In conclusion, the developed pipeline can be used to enable massive Glaucoma screenings in settings that were not possible before since it does not require an internet connection and runs on affordable mobile devices. This can increase the chance of detecting Glaucoma at an earlier stage and contribute to decrease both individual and economic burdens caused by the disease.

Declaration of Competing Interest

The authors whose names are listed immediately below certify that they have NO affliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgments

This work was supported by the North Portugal Regional Operational Programme (NORTE 2020), Portugal 2020; and the European Regional Development Fund (ERDF) from European Union through the project Symbiotic technology for societal efficiency gains: Deus ex Machina (DEM), NORTE-01-0145-FEDER-000026.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2020.105341.

References

[1] H. Fu, J. Cheng, Y. Xu, D.W.K. Wong, J. Liu, X. Cao, Joint optic disc and cup segmentation based on multi-label deep network and polar transformation, IEEE Transactions on Medical Imaging 37 (7) (2018) 1597–1605, doi:10.1109/TMI.2018.2791488. arXiv:1801.00926

[2] Y. Hagiwara, J.E.W. Koh, J.H. Tan, S.V. Bhandary, A. Laude, E.J. Ciaccio, L. Tong, U.R. Acharya, Computer-aided diagnosis of glaucoma using fundus images: a review, Comput. Methods Programs Biomed. 165 (2018) 1–12, doi:10.1016/j.cmpb.2018.07.012.

[3] R. Thomas, K. Loibl, R. Parikh, Evaluation of a glaucoma patient, Indian J. Ophthalmol. 59 (7) (2011) 43, doi:10.4103/0301-4738.73688.

[4] T. Saba, S.T.F. Bokhari, M. Sharif, M. Yasmin, M. Raza, Fundus image classification methods for the detection of glaucoma: a review, Microsc. Res. Tech. (2018) 1–17, doi:10.1002/jemt.23094.

[5] J. Morales, Assessment of Iris Reflection Artifacts and Alignment in Fundus Images, Faculty of the Graduate School of St. Mary's University, 2011 Ph.D. thesis. doi: 10.13140/RG.2.2.26094.41280

[6] B. Al-Bander, B.M. Williams, W. Al-Nuaimy, M.A. Al-Taee, H. Pratt, Y. Zheng, Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis, Symmetry 10 (4) (2018), doi:10.3390/sym10040087.

[7] A. Singh, M.K. Dutta, M. ParthaSarathi, V. Uher, R. Burget, Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image, Comput. Methods Programs Biomed. 124 (2016) 108–120, doi:10.1016/j.cmpb.2015.10.010.

[8] Z. Li, Y. He, S. Keel, W. Meng, R.T. Chang, M. He, Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs, Ophthalmology 125 (8) (2018) 1199–1206, doi:10.1016/j.ophtha.2018.01.023.

[9] Z. Zhang, F.S. Yin, J. Liu, W.K. Wong, N.M. Tan, B.H. Lee, J. Cheng, T.Y. Wong, 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 Society, EMBC'10, 2010, pp. 3065–3068, doi:10.1109/IEMBS.2010.5626137.

[10] J. Sivaswamy, S.R. Krishnadas, G. Datt Joshi, M. Jain, A.U. Syed Tabish, Drishti-GS: retinal image dataset for optic nerve head(ONH) segmentation, in: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), 2014, pp. 53–56, doi:10.1109/ISBI.2014.6867807.

[11] F. Fumero, S. Alayon, J.L. Sanchez, J. Sigut, M. Gonzalez-Hernandez, RIM-ONE: an open retinal image database for optic nerve evaluation, IEEE Symp. Comput.-Based Med. Syst. (2011) 1–6, doi:10.1109/CBMS.2011.5999143.

[12] F. Fumero, J. Sigut, S. Alayon, M. González-Hernández, M.G. de la Rosa, Interactive tool and database for optic disc and cup segmentation of stereo and monocular retinal fundus images, in: Short Papers Proceedings - WSCG 2015, Pilsen, Czech Republic, 2015, pp. 91–97.
[13] ichallenge, (http://ichallenge.baidu.com/). (Accessed on 07/12/2019).
[14] A. Almazroa, W. Sun, S. Alodhayb, K. Raahemifar, V. Lakshminarayanan, Optic disc segmentation for glaucoma screening system using fundus images, Clinical Ophthalmology 11 (2017) 2017–2029, doi:10.2147/OPTH.S140061.
[15] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: efficient convolutional neural networks for mobile vision applications, CoRR (2017). arXiv:1704.04861 10.1016/S1507-1367(10)60022-3.
[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: inverted residuals and linear bottlenecks, CoRR (2018), doi:10.1134/S0001434607010294. CoRRarXiv:1801.04381
[17] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, January, 2017, pp. 6230–6239, doi:10.1109/CVPR.2017.660. arXiv:1612.01105v2.
[18] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs., IEEE Transactions on Pattern Analysis and MachineIntelligence 40 (4) (2016) 834–848, doi:10.1109/TPAMI.2017.2699184. http://www.ncbi.nlm.nih.gov/pubmed/28463186
[19] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, CoRR (2017). http://arxiv.org/abs/1706.05587
[20] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11211 LNCS, 2018, pp. 833–851, doi:10.1007/978-3-030-01234-2_49. arXiv:arXiv:1802.02611v3
[21] K. Zuiderveld, Graphics gems iv, in: P.S. Heckbert (Ed.), Graphics Gems IV, Academic Press Professional, Inc., San Diego, CA, USA, 1994, pp. 474–485. http://dl.acm.org/citation.cfm?id=180895.180940
[22] G.D. Joshi, J. Sivaswamy, S.R. Krishnadas, Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment, IEEE Trans. Med. Imaging 30 (6) (2011) 1192–1205, doi:10.1109/TMI.2011.2106509.
[23] F. Yin, J. Liu, S.H. Ong, Y. Sun, D.W. Wong, N.M. Tan, C. Cheung, M. Baskaran, T. Aung, T.Y. Wong, Model-based optic nerve head segmentation on retinal fundus images, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2011, pp. 2626–2629, doi:10.1109/IEMBS.2011.6090724.
[24] J. Cheng, J. Liu, Y. Xu, F. Yin, D.W.K. Wong, N.M. Tan, D. Tao, C.Y. Cheng, T. Aung, T.Y. Wong, Superpixel classification based optic disc and optic cup segmentation for glaucoma screening, IEEE Trans. Med. Imaging 32 (6) (2013) 1019–1032, doi:10.1109/TMI.2013.2247770.
[25] Y. Xu, L. Duan, S. Lin, X. Chen, D. Wing, K. Wong, Optic cup segmentation for glaucoma detection using low-rank superpixel representation, Med. Image Comput. Comput. Assist. Interv. 17 (Pt 1) (2014) 788–795.
[26] J. Cheng, D. Tao, D.W.K. Wong, J. Liu, Quadratic divergence regularized SVM for optic disc segmentation, Biomed. Opt. Expr. 8 (5) (2017) 2687, doi:10.1364/boe.8.002687.
[27] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9351, 2015, pp. 234–241, doi:10.1007/978-3-319-24574-4_28.
[28] C. Sciences, The ISNT rule and differentiation of normal from glaucomatous eyes, JAMA Ophthalmol. 124 (11) (2014) 1579–1583, doi:10.1001/archopht.124.11.1579.
[29] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556(2015). http://www.robots.ox.ac.uk/.
[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, CoRR (2015), doi:10.1109/CVPR.2016.308. abs/1512.00567 http://arxiv.org/abs/1512.00567
[31] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, CoRR (2015), doi:10.1109/CVPR.2016.90. http://arxiv.org/abs/1512.03385
[32] A. Antón, G. Fatti, M.J. Ledesma-Carbayo, A. Herranz, B. Liefers, A. Santos, J.J. Gómez-Valverde, C.I. Sánchez, Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning, Biomed. Opt. Expr. (2019), doi:10.1364/boe.10.000892.
[33] S. Maheshwari, R.B. Pachori, U.R. Acharya, Automated diagnosis of glaucoma using empirical wavelet transform and correntropy features extracted from fundus images, IEEE J. Biomed. Health Inf. 21 (3) (2017) 803–813, doi:10.1109/JBHI.2016.2544961.
[34] U.R. Acharya, S. Bhat, J.E. Koh, S.V. Bhandary, H. Adeli, A novel algorithm to detect glaucoma risk using texton and local configuration pattern features extracted from fundus images, Comput. Biol. Med. 88 (May) (2017) 72–83, doi:10.1016/j.compbiomed.2017.06.022.
[35] B. Al-Bander, W. Al-Nuaimy, M.A. Al-Taee, Y. Zheng, Automated glaucoma diagnosis using deep learning approach, in: 2017 14th International Multi-Conference on Systems, Signals and Devices, SSD 2017, 2017-Janua, 2017, pp. 207–210, doi:10.1109/SSD.2017.8166974.
[36] H. Fu, J. Cheng, Y. Xu, C. Zhang, D.W.K. Wong, J. Liu, X. Cao, Disc-aware ensemble network for glaucoma screening from fundus image, CoRR (2018) 1–9, doi:10.1109/TMI.2018.2837012.
[37] M. Christopher, A. Belghith, C. Bowd, J.A. Proudfoot, M.H. Goldbaum, R.N. Weinreb, C.A. Girkin, J.M. Liebmann, L.M. Zangwill, Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs, Scientific Reports 8 (1) (2018) 16685, doi:10.1038/s41598-018-35044-9. http://www.nature.com/articles/s41598-018-35044-9
[38] K. Mitsuhashi, Y. Fujino, M. Tanito, N. Shibata, H. Murata, R. Asaoka, M. Matsuura, Development of a deep residual learning algorithm to screen for glaucoma from fundus photography, Scient. Rep. 8 (1) (2018) 1–9, doi:10.1038/s41598-018-33013-w.
[39] Q. Zhang, S.-C. Zhu, Visual interpretability for deep learning: a survey, CoRR 19 (1423305) (2018) 27–39.
[40] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017-Octob, 2017, pp. 618–626, doi:10.1109/ICCV.2017.74. arXiv:1610.02391v3.
[41] R.A. Shruti Gorasia, A review paper on detection of glaucoma using retinal fundus images, Int. J. Res. Appl. Eng. Technol. 4 (I) (2016) 166–170.
[42] C. Muramatsu, Y. Hatanaka, A. Sawada, T. Yamamoto, H. Fujita, Computerized detection of peripapillary chorioretinal atrophy by texture analysis, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2011, pp. 5947–5950, doi:10.1109/IEMBS.2011.6091470.
[43] J.B. Jonas, Clinical implications of peripapillary atrophy in glaucoma, Curr. Opin. Ophthalmol. 16 (2) (2005) 84–88, doi:10.1097/01.icu.0000156135.20570.30.