Deep Learning Glaucoma Detection Models in Retinal Images Capture by Mobile Devices

Roberto Flavio Rezende[1], Ana Coelho[1], Rodrigo Fernandes[1], José Camara[3], Alexandre Neto[1,2], and António Cunha[1,2(✉)]

[1] Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801 Vila Real, Portugal
acunha@utad.pt

[2] Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, 3200-465 Porto, Portugal

[3] Departamento de Ciências e Tecnologia, Universidade Aberta, 1250-100 Lisbon, Portugal

Abstract. Glaucoma is a disease that arises from increased intraocular pressure and leads to irreversible partial or total loss of vision. Due to the lack of symptoms, this disease often progresses to more advanced stages, not being detected in the early phase. The screening of glaucoma can be made through visualization of the retina, through retinal images captured by medical equipment or mobile devices with an attached lens to the camera. Deep learning can enhance and increase mass glaucoma screening. In this study, domain transfer learning technique is important to better weight initialization and for understanding features more related to the problem. For this, classic convolutional neural networks, such as ResNet50 will be compared with Vision Transformers, in high and low-resolution images. The high-resolution retinal image will be used to pre-trained the network and use that knowledge for detecting glaucoma in retinal images captured by mobile devices. The ResNet50 model reached the highest values of AUC in the high-resolution dataset, being the more consistent model in all the experiments. However, the Vision Transformer proved to be a promising technique, especially in low-resolution retinal images.

Keywords: Glaucoma Screening · Deep Learning · Transfer Learning

1 Introduction

Glaucoma is a silent disease that arises from increased intraocular pressure and consequently leads to irreversible partial or total loss of vision. This loss of vision occurs due to the destruction of the ganglion cells, which belong to the optic nerve, a structure that connects the eye to the occipital brain and which is responsible for conducting images from the retina to the brain [1, 2]. Aqueous humor is a transparent liquid composed of water and dissolved salts. Its function is to nourish the cornea and the crystalline lens, besides regulating intraocular pressure. The liquid is located in the anterior and posterior chamber of the eyeball [1, 3]. The reason why aqueous humor production is so important

A. Cunha et al. (Eds.): MobiHealth 2022, LNICST 484, pp. 3–13, 2023.
https://doi.org/10.1007/978-3-031-32029-3_1

is that if this liquid is not drained in the same quantities as it is produced, the ocular pressures increases and irreversible damages the optic nerve, leading to the development of glaucoma. After all, in glaucoma patients, these fibers are atrophied, making it impossible to conduct images to the brain. Secondarily, there is the appearance of scotomas in the visual field and with the evolution of the disease, glaucoma causes progressive loss of vision [1, 3, 4]. Figure 1 show the effects of high intraocular pressure and the visual consequences.

Fig. 1. Normal and glaucomatous eye. Consequences of glaucoma in vision loss (adapted from Hagiwara et al. [4]).

The retina can be directly examined by using an ophthalmoscope or can be examined indirectly through retinal fundus images. The retinal fundus images allow for the detection of indicators and parameters normally correlated to the appearance and development of cupping, such as disc diameter, peripapillary atrophy notching and cup-to-disc ratio (CDR). Retinal images turn the process very easy to access the data, duplication, archiving and delivery, which help in more immediate results in medical centers where is performed automatic or manual screening [5].

Since ophthalmology in general, and particularly the screening of glaucomatous diseases, is heavily based on image analysis, one of the emerging research areas in recent years is the interpretation of images using automated computational methods. In this field of visual analysis by computers for the identification of ophthalmic diseases the use of deep learning (DL) algorithms has stood out [6].

The use of deep learning in the identification of pathologies, especially those that are diagnosed through a visual pattern, offers the potential to improve the accuracy and speed of exam interpretation in addition to lowering their costs [7].

This study intends to compare the performance of two deep learning models in glaucoma classification, namely the ViT and ResNet50 architecture, and evaluate if this technique be beneficial to models created for glaucoma classification in low-resolution images.

2 Related Work

2.1 Convolutional Neural Network

In fields of investigation where information is collected more rapidly than analyzed and require experienced individuals to perform these investigations, DL strategies show up as tools for handling and analyzing a great amount of information. A typical method of DL for image processing is the convolutional neural network (CNN) [8, 9].

There is available a large variety of CNNs and in this section, some will be explained according to the ones that will be used in this work. One type of CNN is the residual network (ResNet) which was created with shortcut connections to increase the depth of CNN without gradient degradation. The shortcuts assist the gradient flowing easily during the backpropagation, leading to accuracy gains during the training. Normally this type of network is composed of 4 blocks, with convolutional blocks inside being the difference among the different versions of ResNet (50, 101 and 152) the number of consecutive convolutional blocks. With the use of residual blocks the gradient gradation problem is resolved [10].

Recently, a new type of neural network emerged as an alternative to CNNs, relying on self-attention mechanisms, called Vision Transformers (ViT). The ViT model, when compared to a traditional CNN, has a weaker inductive bias leading to greater reliance on the model regularization or data augmentation when training on smaller datasets. The training phase of ViT depends on measuring the relation among pairs of input tokens. The tokens, in the computer vision field, can be made by using image patches. The relation between these patches is learned through providing attention to the network [11].

Since classification DL models can be black boxes, explainable artificial intelligence (XAI) methods were created to provide transparency and explainability explainable to assist experts and non-experts in understanding which features are related and influence the models' decisions [12]. One of the methods that can be used for XAI is the Grad-CAM which uses the specific gradient information for each class from the last convolutional layer, to design a coarse location map of the important areas in the image [13]. The other method, designed for explainability in attention models is the attention rollout mechanism that quantifies how the information flows through in transformer blocks, doing the average of weights across all the heads [14].

2.2 Deep Learning for Glaucoma Detection

Machine learning techniques have become indispensable in many areas. With these techniques, computers are increasingly equipped with the capability to act without the need of explicitly being programmed, building models which can train from data and make decisions based on that data.

Chen [15] proposed a CNN-based DL algorithm to detect glaucoma presence in retinal fundus images, these images being present in the ORIGIN and SCES datasets. This CNN is composed of six layers comprising four convolutional layers and two fully connected layers. A dropout layer was added, and data augmentation was implemented to further increase the results of glaucoma detection. An area under the curve (AUC) of 0.831 and 0.887 for each of the databases was achieved, respectively.

In Raghavendra [16] study a CNN (18 convolution layers) was used for the same purpose of detecting glaucoma in retinal fundus images. For this were collected 1426 retinal fundus images composed of 589 normal cases and 837 with glaucoma train the network.

Chai [17] designed a multivariate neural network model, which was inspired by domain knowledge, to extract hidden features from the retinal images simultaneously while extracting important areas of images. The performance of the algorithm was quite high, as they achieved an accuracy of 0.9151, sensitivity of 0.9233 and specificity of 0.9090. Benzebouchi [18], used two DL models with multimodal data (RGB and grayscale images). The two models are composed of two convolutional layers and one fully convolutional layer. It turned out that it would be better if the two models were combined.

In Suguna [19], the authors used a pre-trained model called EyeNet, which was originally trained on fundus images for the classification of a different eye disease, to detect glaucoma presence in retinal fundus images. The contribution is that the model is more suited to the problem since the EyeNet model was trained to classify diabetic retinopathy, which is a very related problem. The similarities between the two disease domains largely influence the amount of knowledge transfer. Using pre-trained models of similar domains outperforms pre-trained models that used generic datasets, unrelated to the glaucoma classification. The proposed method reached better results when compared to pre-trained models based on generic datasets.

Alghamdi [20] proposes a framework for automatic glaucoma diagnosis based on three CNN models, each one with different methods for learning, comparing the results with ophthalmologists. In the first phase, the authors start with a pre-trained model which was trained in unrelated data and fine-tunes with the glaucoma data. Then, a semi-supervised model is developed to train using label and unlabeled data. This method demonstrates the effectiveness in glaucoma detection from the CNN models when tested in two different datasets (RIM-ONE and RIGA). All the models reached better performance when compared to the two ophthalmologists.

3 Materials and Methodologies

This project intends to compare the performance of two models in glaucoma classification, namely the ViT and ResNet50 architecture. These models will be first pre-trained in public high-resolution retinal images, with and without PCA color augmentation. After this, the weights of these models were used to pre-trained the same models in the private dataset of low-resolution retinal images (collected by mobile devices), with and without PCA color augmentation. In the end, both the glaucoma assessment in public and private databases will be evaluated and will be concluded if this transfer learning technique can benefit the models created for glaucoma classification in low-resolution images. To further explained the results, XAI techniques will be implemented as well. In this section, will be described the datasets used for this work (public and private), the pre-processing methods and the evaluation metrics to test the glaucoma assessment of each model.

3.1 Public Dataset

RIM-ONE consists of 169 optic nerve head (ONH) images obtained from 169 complete retinal fundus images of different patients (10 images were discarded because were not directly related to glaucoma) [21].

DRISHTI-GS is composed of a total of 101 images of which 50 images are for training and 51 images are for testing [22].

REFUGE is composed of 1200 retinal fundus images, acquired by ophthalmologists or technicians, from patients sitting upright (only 400 images are available for training) [23].

In Table 1 are described the proprieties of the three public databases used.

Table 1. Public retinal image datasets.

Database	RIM-ONE	Drishti-GS	REFUGE
Normal	85	31	360
Glaucoma	74	70	40

In the model developed for the training, the set consists of 128 images classified as glaucoma and 332 images classified as normal. For the testing set, 28 images are glaucoma and 72 normal, for validation the same number of images with same glaucoma/normal ratio. The percentage is then 70% training, 15% test and 15% validation.

3.2 Private Dataset

The "Private Dataset" is a dataset consisting of 491 images (356 normal and 135 with glaucoma). This dataset was provided by an ophthalmologist to help this study, given that the images used in this work must contain low quality examples and this is exactly what this dataset provides. Figure 2 illustrates some examples of low-resolution retinal images captured by mobile devices.

For training, testing and validation the percentages were the same as for the public dataset, so for training 260 normal and 95 glaucoma images were used, for testing 46 normal and 20 glaucoma images, and finally for validation 50 normal and 20 glaucoma images, making a percentage of about 70%, 15% and 15% respectively.

3.3 Image Pre-processing

Resize Images: Since the selected images did not all have the same resolution, a filter was applied to these same images, and all images were resized to a resolution of 512 × 512 pixels.

PCA Color Augmentation

Color augmentation principal components analysis (PCA), changes RGB channel intensities, using PCA of the pixel colors [24]. Specifically, PCA is performed on RGB pixel

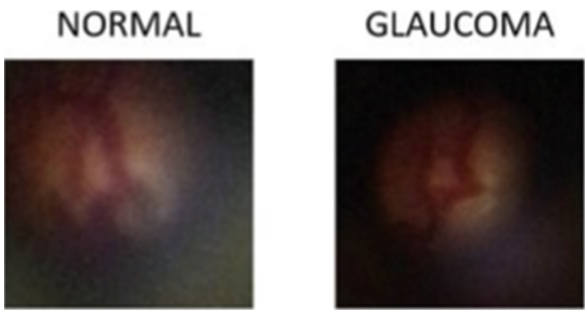

Fig. 2. Examples of normal and glaucoma images in private retinal images collected by mobile devices.

values in the entire data set. The principal component multiples are added to each image, with proportional to the corresponding eigenvalues times a random variable drawn from a Gaussian distribution with mean = 0 and standard deviation = 0.1.

3.4 Evaluation Metrics

Evaluation of a classification model is done by comparing the classes predicted by the model with the true classes of each example. All classification metrics have the common goal of measuring how far the model is from perfect classification, but they do this in different ways.

The accuracy (1) tells how many examples were classified correctly, regardless of class. Follows the accuracy equation:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Sensitivity (2) is also known as recall. This metric evaluates the method's ability to successfully detect results classified as positive. It can be obtained by the equation:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Specificity (3), on the other hand, evaluates the ability of the method to detect negative results. We can calculate it using the equation:

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{3}$$

Precision (4) is a metric that evaluates the number of true positives over the sum of all positive values. It is calculated by the following formula:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

F-measure, F-score or F1 score (5) is a harmonic mean calculated based on precision and revocation. It can be obtained from the equation:

$$\text{F1 Score} = 2 \times \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \tag{5}$$

4 Results and Discussion

Deep learning models can help to increase the speed of mass glaucoma screening and maintain the performance between exams and examiners. The models' results in the high-resolution dataset are described in Table 2. Using deep learning algorithms for glaucoma detection in high-resolution images demonstrates high performance, with AUCs above 0.70. The standard CNN model (ResNet50) reached the highest performance among the models selected. The use of PCA pre-processing increased considerably the ResNet50 performance, thus, this pre-processing technique enhances the feature visualization related to glaucoma presence. Regarding the ViT models, the results did not reach the same standards as the ResNet50. Despite the high values of specificity and precision, the sensitivity was below 0.5. In this case, instead of enhancing the effectiveness of the ViT, the pre-processing using the PCA method downgraded the results.

Table 2. Model results in high-resolution retinal images with or without PCA pre-processing.

Models	Acc	Sen	Spec	Prec	F1	AUC
ViT	0.60	0.49	0.89	0.92	0.64	0.72
ViT PCA	0.58	0.42	**1.00**	**1.00**	0.59	0.70
ResNet50	0.80	**0.86**	0.64	0.86	**0.86**	0.78
ResNet50 PCA	**0.81**	0.82	0.79	0.86	**0.86**	**0.82**

After training and testing in high-resolution datasets, the weights of these models were used to initialize the training with low-resolution retinal images collected from mobile devices. The purpose is to use transfer learning techniques from a close and similar domain problem instead of using generic datasets. The results of the models in the images from mobile devices are presented in Table 3.

Table 3. Model results in low-resolution retinal images with or without PCA pre-processing.

Models	Acc	Sen	Spec	Prec	F1	AUC
ViT	0.67	0.25	0.85	0.42	0.31	0.58
ViT PCA	**0.76**	**0.40**	**0.90**	**0.62**	**0.48**	**0.83**
ResNet50	0.62	0.20	0.80	0.31	0.24	0.78
ResNet50 PCA	0.63	0.30	0.84	0.55	0.39	0.82

In this case, the PCA technique proved to be a good method for improving feature visualization in low-resolution retinal images. Both the ViT and ResNet50 with PCA pre-processing reached AUCs above 0.8. The ViT model in this dataset was the one reaching better results due to the fact of pair-wise pixel relations processing of the self-attention

mechanisms which can detect more important features in low-resolution images. However, the ResNet50 showed consistency either in the high-resolution and low-resolution images, always with AUC above 0.75 with or without the PCA pre-processing. After testing the models, explainability methods were implemented to highlight important features related to the model decision. In Fig. 3 are some examples of ResNet50 predictions with Grad-CAM activation maps for low-resolution images and in Fig. 4 some examples of ViT predictions with attention rollouts activation maps in low-resolution retinal images as well. As can be seen in Fig. 3, the main features highlighted are within the optic disc, especially in the region with veins.

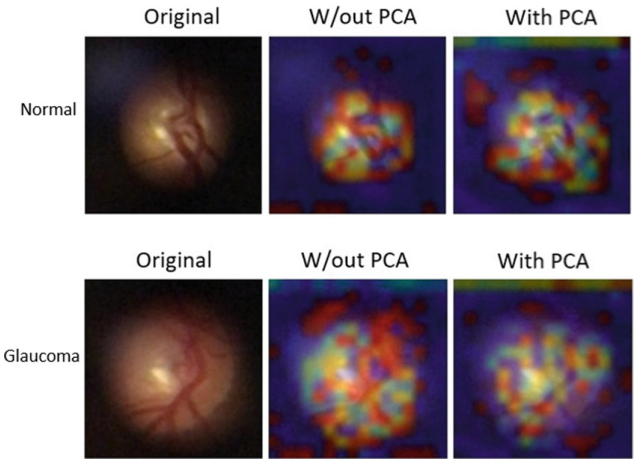

Fig. 3. Examples of Grad-CAM explainability in normal and glaucoma cases for ResNet50 model.

In Fig. 3, in the bottom row, in the case without the PCA, the model focus on the region corresponding to the optic cup. The correlation between a large optic cup and glaucoma presence is big. Thus, the ResNet50 model focus on an important feature to predict glaucoma presence.

Regarding the attention maps from the ViT models, the same can be concluded in Fig. 4. In both examples, either in normal or glaucoma cases, can be seen that the model highlights the optic disc region. In this case, in the normal case is possible to see that the model highlights the optic cup. The optic cup, in the normal case, is small, which does not demonstrate signs of glaucoma presence. Compared to the same example in Fig. 3 is possible to see that the models focus on similar features in the same image.

These recent ViT methods which rely on the self-attention mechanism prove to have promising results. However, the ResNet50 reached more stable results throughout the different experiments made, reaching at least 0.75 of AUC with or without PCA pre-processing in high and low-resolution retinal images. Despite this, the ViT model has shown great potential in detecting glaucoma in retinal images collected by mobile devices due to the process of pixel relation of the model. The domain transfer learning uses more similar datasets, inserted in the same type of problem, allowing the transfer of more related knowledge to another model for a better weight initialization. The PCA

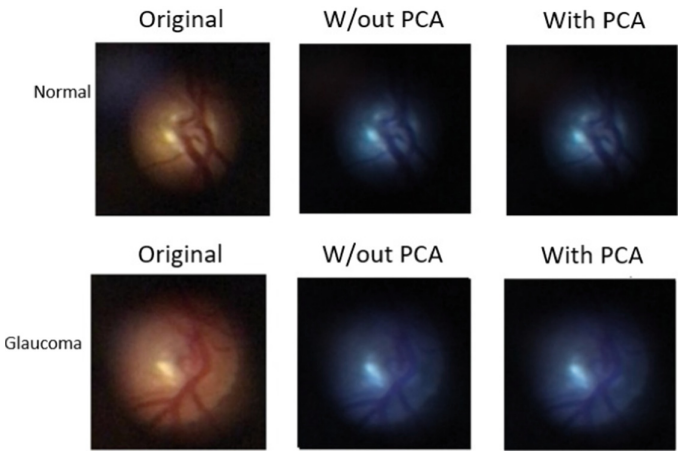

Fig. 4. Examples of attention rollout explainability in normal and glaucoma cases for ViT model.

pre-processing technique shows more relevance in the low-resolution dataset, where the models using this pre-processing method reached higher values of AUC, thus, this technique can enhance feature visualization in the retinal image.

5 Conclusions

The purposed methods in this study proved to be promising since the knowledge transfer of datasets related to the same domain can improve the model training, enabling to detect glaucoma in low-resolution images collected by mobile devices. The emerging techniques of transformers and self-attention mechanisms in ViT models reveal to have promising performance, especially in the low-resolution retinal images, where the model reached the highest results due to the process of the relation between pixels. In this specific case for the mobile retinal images, the PCA pre-processing reveal to have importance, enhancing feature visualization and improving the model results. The explainability methods, in the end, highlighted correlated features to the corresponding class of the image.

Future works could verify if adaptations in ViT models could increase the effectiveness of glaucoma detection by using shifted patch tokenization or/and locality self-attention are very useful, especially in small datasets. Also, more suitable local explainability methods will be explored such as local interpretable model-agnostic explanations.

Acknowledgements. This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project LA/P/0063/2020.

References

1. Claro, M.L., Veras, R., Santos, L., Frazão, M., Carvalho Filho, A., Leite, D.: Métodos computacionais para segmentação do disco óptico em imagens de retina: uma revisão. Rev. Bras. Comput. Apl. **10**(2), 29–43 (2018). https://doi.org/10.5335/rbca.v10i2.7661
2. Bajwa, M.N., et al.: Two-stage framework for optic disc localization and glaucoma classification in retinal fundus images using deep learning. BMC Med. Inform. Decis. Mak. **19**(1), 136 (2019). https://doi.org/10.1186/s12911-019-0842-8
3. Stella Mary, M.C.V., Rajsingh, E.B., Naik, G.R.: Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey. IEEE Access **4**, 4327–4354 (2016). https://doi.org/10.1109/ACCESS.2016.2596761
4. Hagiwara, Y., et al.: Computer-aided diagnosis of glaucoma using fundus images: a review. Comput. Methods Programs Biomed. **165**, 1–12 (2018). https://doi.org/10.1016/j.cmpb.2018.07.012
5. Camara, J., Neto, A., Pires, I.M., Villasana, M.V., Zdravevski, E., Cunha, A.: A comprehensive review of methods and equipment for aiding automatic glaucoma tracking. Diagnostics **12**(4), 935 (2022). https://doi.org/10.3390/diagnostics12040935
6. Mayro, E.L., Wang, M., Elze, T., Pasquale, L.R.: The impact of artificial intelligence in the diagnosis and management of glaucoma. Eye **34**(1), 1–11 (2019). https://doi.org/10.1038/s41433-019-0577-x
7. Litjens, G., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005
8. Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K.: Convolutional neural networks: an overview and application in radiology. Insights Imaging **9**(4), 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9
9. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X.: Deep learning and its applications to machine health monitoring. Mech. Syst. Signal Process. **115**, 213–237 (2019). https://doi.org/10.1016/j.ymssp.2018.05.050
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition arXiv, arXiv:1512.03385 (2015). Accessed 05 June 2022
11. Dosovitskiy, A., et al.: An image is worth 16 × 16 words: transformers for image recognition at scale. arXiv, arXiv:2010.11929 (2021). Accessed 02 June 2022
12. Neto, A., Camara, J., Cunha, A.: Evaluations of deep learning approaches for glaucoma screening using retinal images from mobile device. Sensors **22**(4), 1449 (2022). https://doi.org/10.3390/s22041449
13. Linardatos, P., Papastefanopoulos, V., Kotsiantis, S.: Explainable AI: a review of machine learning interpretability methods. Entropy **23**(1), 1–45 (2021). https://doi.org/10.3390/e23010018
14. Abnar, S., Zuidema, W.: Quantifying attention flow in transformers (2020). http://arxiv.org/abs/2005.00928. Accessed 18 July 2022
15. Chen, X., Xu, Y., Kee Wong, D.W., Wong, T.Y., Liu, J.: Glaucoma detection based on deep convolutional neural network. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, pp. 715–718 (2015). https://doi.org/10.1109/EMBC.2015.7318462
16. Raghavendra, U., Fujita, H., Bhandary, S.V., Gudigar, A., Tan, J.H., Acharya, U.R.: Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. Inf. Sci. **441**, 41–49 (2018). https://doi.org/10.1016/j.ins.2018.01.051
17. Chai, Y., Liu, H., Xu, J.: Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. Knowl.-Based Syst. **161**, 147–156 (2018). https://doi.org/10.1016/j.knosys.2018.07.043

18. Benzebouchi, N.E., Azizi, N., Bouziane, S.E.: Glaucoma diagnosis using cooperative convolutional neural networks. Int. J. Adv. Electron. Comput. Sci. **5**(1), 31–36 (2018)
19. Suguna, G., Lavanya, R.: Performance assessment of EyeNet model in glaucoma diagnosis. Pattern Recogn. Image Anal. **31**(2), 334–344 (2021). https://doi.org/10.1134/S1054661821020164
20. Alghamdi, M., Abdel-Mottaleb, M.: A Comparative study of deep learning models for diagnosing glaucoma from fundus images. IEEE Access **9**, 23894–23906 (2021). https://doi.org/10.1109/ACCESS.2021.3056641
21. Fumero Batista, F.J., Diaz-Aleman, T., Sigut, J., Alayon, S., Arnay, R., Angel-Pereira, D.: RIM-ONE DL: a unified retinal image database for assessing glaucoma using deep learning. Image Anal. Stereol. **39**(3), Article no. 3 (2020). https://doi.org/10.5566/ias.2346
22. Sivaswamy, J., Krishnadas, S.R., Datt Joshi, G., Jain, M., Syed Tabish, A.U.: Drishti-GS: retinal image dataset for optic nerve head (ONH) segmentation. Presented at the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp. 53–56 (2014). https://doi.org/10.1109/ISBI.2014.6867807
23. Fu, H.: REFUGE: Retinal Fundus Glaucoma Challenge. IEEE (2019). https://ieee-dataport.org/documents/refuge-retinal-fundus-glaucoma-challenge. Accessed 16 Nov 2022
24. Bargoti, S., Underwood, J.: Deep fruit detection in orchards (2017). http://arxiv.org/abs/1610.03677. Accessed 28 July 2022