Diagnose like an Ophthalmologist - Glaucoma Detection in the AIROGS Challenge

Firas Khader[a], Christoph Haarburger[b], Jörg-Christian Kirr[b], Marcel Menke[c,b], Johannes Stegmaier[d], Daniel Truhn[a]

[a]Dpt. of Diagnostic and Interv. Radiology, RWTH Aachen University Hospital, Aachen, Germany
[b]CheckupPoint AG, Riedt, Switzerland
[c]Department of Ophthalmology, Kantonsspital Aarau, Aarau, Switzerland
[d]Institute of Imaging and Computer Vision, RWTH Aachen, Aachen, Germany

Abstract

Glaucoma is an optic neuropathy that is characterized by the progressive degeneration of the optic nerve, leading to visual impairment. While it is the main cause of irreversible blindness worldwide, it typically remains asymptomatic until at a late stage. Glaucoma is usually diagnosed during routine eye examination, which includes fundoscopic evaluation. Here, we describe our approach for glaucoma detection in the Artificial Intelligence for RObust Glaucoma Screening challenge. We first use object detection to focus on the most relevant part of the ocular fundus image and then use an ensemble of classification networks. Ungradability is rated via a combination of out-of-distribution methods and reliability both in the object detection stage and in the classification stage.

Introduction

The term glaucoma summarizes various eye diseases in which the optic nerve is damaged. Glaucoma leads to vision impairment and is the main cause of irreversible blindness worldwide [1]. Early diagnosis is pivotal since damage to the optical nerve that has already occured cannot be reversed. A common cause is excessive intraocular pressure. Lowering the intraocular pressure can help to delay or stop the gradual loss of vision. Therefore, early diagnosis is the best strategy to prevent glaucoma-induced vision impairment. Due to a global shortage of ophthalmologists, a screening programme is not practically feasible. Therefore, the goal of the AIROGS challenge[1] is to develop machine learning models that can detect glaucoma on real-world ocular fundus images.

Material and Methods

The AIROGS dataset [2] consists of 113,893 RGB ocular fundus images from 60,357 subjects from heterogeneous ethnicities. The data was acquired from ~500 different locations. Human experts labeled all images into two classes: *referrable glaucoma* and *no referable glaucoma*. The dataset was partitioned into a training set (N=101,442) and a test set consisting of ~11,000 images. Images that were considered ungradable by the experts were excluded from the training set *but remained in the test set*. The goal of the challenge was, to provide four scores:

1. Likelihood for referable glaucoma (O1)
2. Binary decision on referable glaucoma presence (O2)
3. Binary decision on whether an image is ungradable (O3)
4. Scalar value that is positively correlated with likelihood for ungradability (O4)

Ophthalmologists almost exclusively regard the optical disc when diagnosing glaucoma. Hence, we mimicked this diagnostic pathway and pursued a two-step approach which consisted of optic disc detection and cropping of a high resolution image of the optic disc followed by a classification pipeline with an ensemble of convolutional neural networks and transformers on the cropped image.

[1] https://airogs.grand-challenge.org

For cropping of the optic disc we manually labeled a box around the optic disc on 3221 fundoscopy images. Performing this labeling procedure took about 7 hours in total. The set was split into 80% training and 20% validation sets and a YOLOv5 [3] object detector network was trained to extract the ROI (Region of Interest). Subsequently, the object detector network was used to extract these boxes on all fundoscopies. We performed a histogram normalization on these boxes (see **Figure 1**) and used an ensemble of four network architectures for diagnosing glaucoma:

- Efficientnet-B4 [4]
- EfficientnetV2-M [5]
- Swin Transformer-B [6]
- DeiT-S [7]

Each of these networks was trained using a 5-fold cross validation to arrive at an ensemble of 4x5=20 networks.

For detecting ungradability we used a hybrid approach. Our hypothesis was that ophthalmologists mainly examine the optic disc to detect the presence of glaucoma. Thus, if the optic disc cannot be found by the object detector network, there is a high probability that the image is ungradable. Consequently, we employed the confidence score of the YOLOv5 object detector as one measure. There might be other reasons for ungradability, such as blurred depiction of the optic disc. Therefore, we trained an additional classifier on a manually selected set of optic disc images. The manual selection was performed by screening through the 4000 fundoscopies that exhibited the lowest confidence score of the object detector and manually selecting those images that seemed most affected by blurring or low-quality depiction of the optic disc. This resulted in 600 manually selected images which were likely to be scored as ungradable (even though all of the images in the training set were by definition gradable, we assumed that these images were more on the side of ungradability). For example we found some images, on which the optic disc was missing completely, or on which the optic disc was only partially visible. We combined these two scores into our ungradability score u according to

$$u = (1 - c) + g$$

where c is the confidence score of the YOLOv5 object detector and g is the score given for ungradabiliity by the neural network (EfficientNet-B4) trained on the manually selected cases.

To tune the cut-off point on which we base our binary decision about the gradability of the images, we inspected the ungradability score u for 20.000 images and chose our cut-off point to be at 1:

$$u_{binary} = 1 \; if \; u < 1 \; else \; 0$$

Results & Discussion

At the time of submission (4th Mar 2022) our approach achieves the highest ungradability AUC and Cohens κ score amongst all other submissions in preliminary phase 2 (10% of the test set). Additionally, the partial AUC (pAUC) and the sensitivity at 95% specificity (Sensitivity@95) are on par with the other top ranking approaches. Exemplary cases where our network ensemble fails to correctly predict referable glaucoma are shown in **Figure 2**. Our results on preliminary phase 2 are given in **Table 1**.

Glaucoma pAUC	Glaucoma Sensitivity @95	Ungradability κ	Ungradabilty AUC
0.8884	0.8313	0.8246	0.9853

Table 1: Results of our approach on preliminary phase 2.

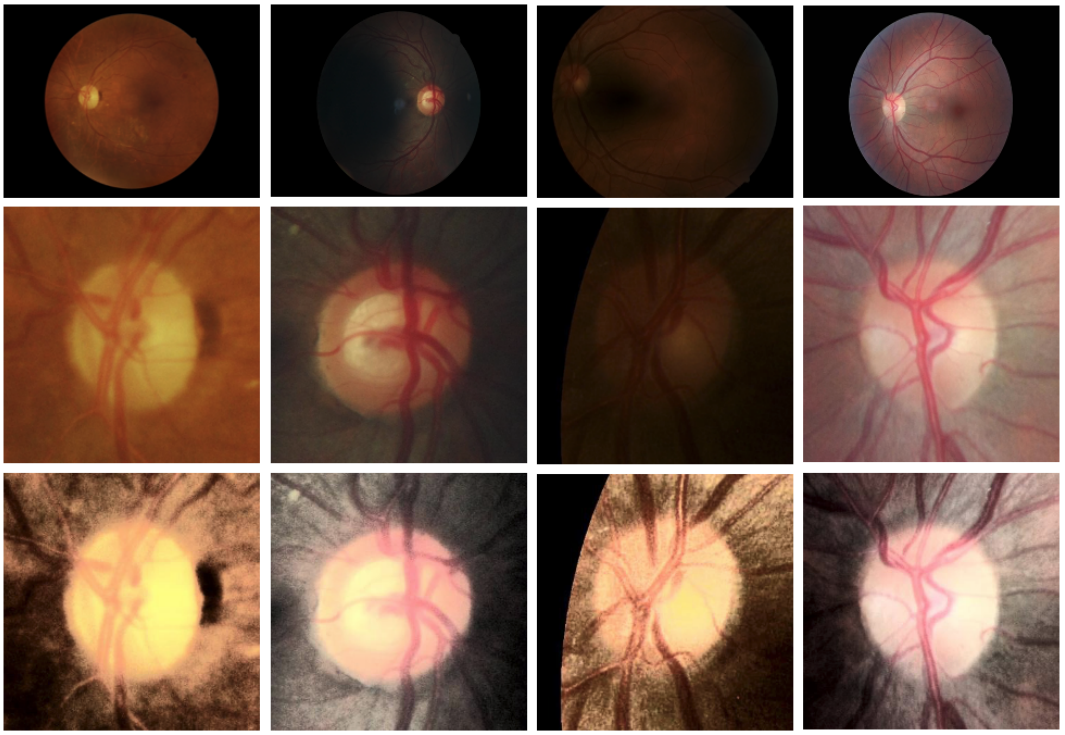

Figure 1: Preprocessing pipeline. In the first step, the whole images (top row) are fed into a YOLOv5 object detection network to find the region around the optic disk (middle row). The images are then further processed using a histogram equalization on the value channel in the HSV color space (bottom row).

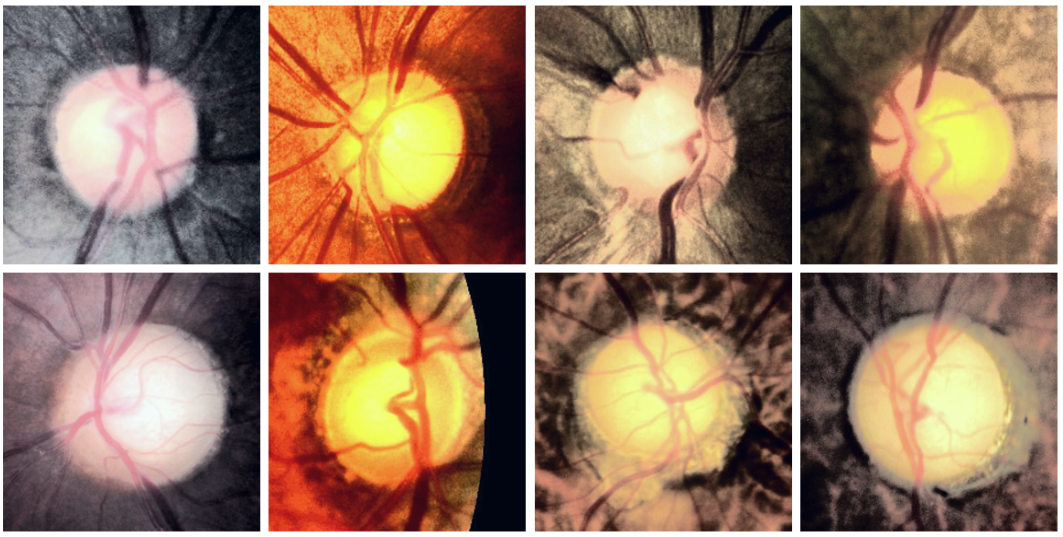

Figure 2: Misclassifications of the neural network ensemble for glaucoma detection. The top row depicts false negative predictions. The bottom row depicts false positive predictions.

References

1. Weinreb, Robert N., Christopher K. S. Leung, Jonathan G. Crowston, Felipe A. Medeiros, David S. Friedman, Janey L. Wiggs, and Keith R. Martin. "Primary Open-Angle Glaucoma." *Nature Reviews. Disease Primers* 2 (September 22, 2016): 16067. https://doi.org/10.1038/nrdp.2016.67.
2. Vente, Coen de, Koenraad A. Vermeer, Nicolas Jaccard, Bram van Ginneken, Hans G. Lemij, and Clara I. Sánchez. "Rotterdam EyePACS AIROGS Train Set." Zenodo, December 1, 2021. https://doi.org/10.5281/zenodo.5793241.
3. Jocher, Glenn, Alex Stoken, Jirka Borovec, NanoCode012, ChristopherSTAN, Liu Changyu, Laughing, et al. *Ultralytics/Yolov5: V3.1 - Bug Fixes and Performance Improvements*. Zenodo, 2020. https://doi.org/10.5281/zenodo.4154370.
4. Tan, Mingxing, and Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." *ArXiv:1905.11946 [Cs, Stat]*, September 11, 2020. http://arxiv.org/abs/1905.11946.
5. Tan, Mingxing, and Quoc V. Le. "EfficientNetV2: Smaller Models and Faster Training." *ArXiv:2104.00298 [Cs]*, June 23, 2021. http://arxiv.org/abs/2104.00298.
6. Liu, Ze, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows." *ArXiv:2103.14030 [Cs]*, August 17, 2021. http://arxiv.org/abs/2103.14030.
7. Touvron, Hugo, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. "Training Data-Efficient Image Transformers & Distillation through Attention." *ArXiv:2012.12877 [Cs]*, January 15, 2021. http://arxiv.org/abs/2012.12877.