# Deep Learning For Referable Glaucoma Screening and Out-of-Distribution Detection

**Zekang Yang**[1], **Hong Liu**[1], **and Zihao Shang**[1]

[1]Institute of Computing Technology

**This paper presents a solution for Artificial Intelligence for RObust Glaucoma Screening (AIROGS) Challenge. We utilized ResNet50 to train a model that not only can diagnose the referable glaucoma using color fundus images, but also can give the ungradable probability when the images' quality is poor.**

## Introduction

Glaucoma is the second leading cause of blindness in the world. Early detection of glaucoma can lead to timely treatment. Glaucoma patients are mainly diagnosed by specialist looking at the fundus images. However, the specialist's energy is limited. Once the fundus examination increases, the workload of specialists also increases, which may make specialists misdiagnosis or missed diagnosis. Automatic screening for glaucoma not only can reduce the workload of specialists, but also have great significance to promote large-scale fundus census. Recent years, artificial intelligence technology has been applied to aid in the diagnosis of glaucoma and achieves good performance at-the-lab. However, in real-world settings,the performance of artificial intelligence will deteriorate due to the existence of out-of-distribution data(eg:bad quality images). With this in mind, we propose a model that can not only diagnose glaucoma, but also give the confidence of the diagnosis.

## The AIROS challenge

In this section, we provide a brief description of AIROS challenge(Artificial Intelligence for Robust Glaucoma Screening Challenge).[1] The Rotterdam EyePACS AIROGS datasets contains 113,893 color fundus images from 60,357 subjects. The training data set contains approximately 102,000 gradable images(3,270 referable glaucoma, and 98,172 no referable glaucoma). The test data set contains about 11,000 gradable and ungradable images.

For each input image during evaluation, the desired output is a likelihood score for referable glaucoma (O1), a binary decision on referable glaucoma presence (O2), a binary decision on whether an image is ungradable (O3, true if ungradable, false if gradable), and a non-thresholded scalar value that is positively correlated with the likelihood for ungradability (O4).

The evaluation will be based on two aspects: screening performance and robustness. The screening performance will be evaluated using the partial area under the receiver operator characteristic curve (90-100% specificity) for referable glaucoma (α) and sensitivity at 95% specificity (β). Using Cohen's kappa score, the agreement between the reference and the decisions provided by the challenge participants on image gradability, O3, is calculated (γ). Furthermore, the area under the receiver operator characteristic curve will be determined using the human reference for ungradability as the true labels and the ungradability scalar values provided by the participants, O4, as the target scores (δ).Finally, all participants will be ranked on the individual metrics α, β, γ and δ, resulting in rankings Rα, Rβ, Rγ and Rδ, respectively. The final score will be calculated as follows: Sfinal = (Rα + Rβ + Rγ + Rδ)/4. The final ranking will subsequently be based on Sfinal, where a lower value for Sfinal will result in a higher ranking.

## Solution

Our solution contains three components: data preprocessing, screening referable glaucoma and out-of-distribution detection.

**Data preprocessing.** Because the image resolution is large and contains redundant information, we preprocess the input image first. We discard the black edges to crop out the eye position and reduce the width to 512 in proportion to the length and width.

**Data Augmentation..** During the training phase, we use RandomAugment[2] for data augmentation. During the inference phase, we random crop five 512x512 crop from input image, then input all crops separately into the model and get five scores. If the maximum of five scores is greater than 0.9, we let it be the output of the model, otherwise we take the mean of the five scores as the output of the model.

**Screening referable glaucoma.** We trained a dichotomous model utilizing pre-trained ResNet50[3]. During inference, the model will give a score $\in [0, 1]$ for the probability of glaucoma. When the score is greater than 0.5, we diagnose glaucoma, otherwise it is non-glaucoma.

**Out-of-distribution detection.** Due to the existence of out-of-distribution data is difficult to diagnose glaucoma or non-glaucoma. The score of out-of-distribution data is far from 0 and 1. We use the difference between the score and 0 (when the score is lower than 0.5) or 1 (when the score is greater than 0.5) to represent the probability of ungradable. And

**Table 1.** Performance of the our method on our test data and the preliminary test phase.

| Test Data | pAUC | TPR@95 | kappa | gAUC |
|---|---|---|---|---|
| our test data | 0.9043 | 0.8628 | - | - |
| preliminary test phase | 0.8542 | 0.7875 | 0.5073 | 0.8994 |

when the probability of ungradable is greater than 0.1, we think it is ungradable, otherwise gradable.

## Results

We divided all training data into training set, validation set and test set according to 7:1:2. We trained our model on the training set, selected the model with the lowest loss value in the validation set, and finally tested the generalization ability of our model on the test set. The results of our model is summarized in the Table 1 .

## References

1. Coen de Vente, Koenraad A. Vermeer, Nicolas Jaccard, Bram van Ginneken, Hans G. Lemij, and Clara I. Sánchez. Rotterdam eyepacs airogs train set, December 2021. The previous version was split into two records. This new version contains all data and the second record is deprecated.
2. Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020.
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.