

Classification for referable glaucoma with fundus photographs using multimodal deep learning

Yeong Chan Lee¹, Hyun Bin Cho², and Yoon Ho Choi²

¹Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea

²Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea

* All authors contributed equally to this work.

Glaucoma is a silent eye disease leading to blindness. Artificial intelligence models for predicting glaucoma using fundus photographs have been developed, however, it is important to evaluate robustness for outliers with high performance to classify glaucoma. We joined The AIROGS challenge, developed, and evaluated a multi-modal deep learning model to predict glaucoma and assess ungradability which is uncertainty for classifying certain class given images. We achieved 0.7635 for the partial AUROC, 0.6125 for sensitivity at 95% specificity, 0.5316 for ungradability kappa, and 0.8057 for ungradability AUROC in preliminary test phase 2.

Correspondence: Yeong Chan Lee, conan_8th@naver.com

Introduction

Glaucoma is a progressive eye disease leading to blindness due to chronic damage to the optic nerve. It may be asymptomatic in early glaucoma, therefore, the key to preventing loss of vision from glaucoma is early detection and treatment by ophthalmologists [1]. Artificial intelligence models have been widely used to predict various eye diseases from fundus images [2]. However, the models were not evaluated with robustness for dealing with out-of-distribution data when maintaining high performance to predict glaucoma. To be used in real-world scenarios, it is important to determine whether a fundus image has enough information to diagnose diseases.

Challenge objectives and constraints. The AIROGS challenge was organized to predict glaucoma with a real-world fundus photograph using computational algorithms. The algorithms will be evaluated for screening performance and robustness. The participants could not access the test set including fundus photographs and ungradable images which cannot be decided as glaucoma. Glaucoma in the test set should be predicted within 10 seconds per single image. In the final test, the participants submitted the predicted results and a short paper only once.

Methods

Dataset. We obtained a total of 101,442 fundus photographs including 98,172 non-referable glaucoma (NRG) and 3,270 referable glaucoma (RG) from The Rotterdam EyePACS

Table 1. Dataset configuration. In phases 1–3, we selected all different images for non-referable glaucoma.

	Phase 1	Phase 2	Phase 3
Training set			
Referable glaucoma	2,588	2,588	2,614
Non-referable glaucoma	2,550	2,575	2,605
Validation			
Referable glaucoma	631	631	640
Non-referable glaucoma	660	644	649

AIROGS dataset [3] RG accounts for only 3.2% in the imbalanced dataset. Thus, we chose an undersampling strategy for preventing overfitting to NRG during the all three phases (Table 1). A phase means a process to train and validate a model. The photographs of NRG were randomly selected same as the number of images of RG (3,270). We constructed three datasets with undersampled NRG images and RG images. Few images were excluded if optic disc had not been detected in the image.

Image preprocessing. Overall, we converted images to grayscale, resized the images with 608 pixels of width and 608 pixels of height, and applied the contrast limited adaptive histogram equalization (CLAHE) to fundus photographs for emphasizing the features of the fundus. We rotated the images within 30 degrees, enlarged within range of 80% and 120%, flipped horizontally and vertically, shifted within 100 pixels, and brightened within range of 80% and 120%. We added Gaussian noise in the images. We augmented the images brighter or darker because the brightness of fundus photographs was diverse. If image preprocessing failed during test, we used a same sized image filled with zeros.

Segmentation of an optic disc using weakly supervised learning. It is essential to identify the glaucomatous changes at the optic disc. So, we decided to develop a model for detecting and segmenting the optic disc in the fundus photograph. We annotated the position of the optic disc in the 101 images (TRAIN000000 to TRAIN000100) roughly. Then we did not label the position with white (RGB, [255, 255, 255]), but label the position with gradient through multivariate normal distribution. The probability density func-

