Enhancing Model Generalisability through Sampling Diverse and Balanced Retinal Images

Tianfeng Zhou[1] and Yukun Zhou[2,3]

[1] School of Electronic Information, Central South University, China
[2] UCL Institute of Ophthalmology, University College London, UK
[3] Centre for Medical Image Computing, University College London, UK
yukun.zhou.19@ucl.ac.uk

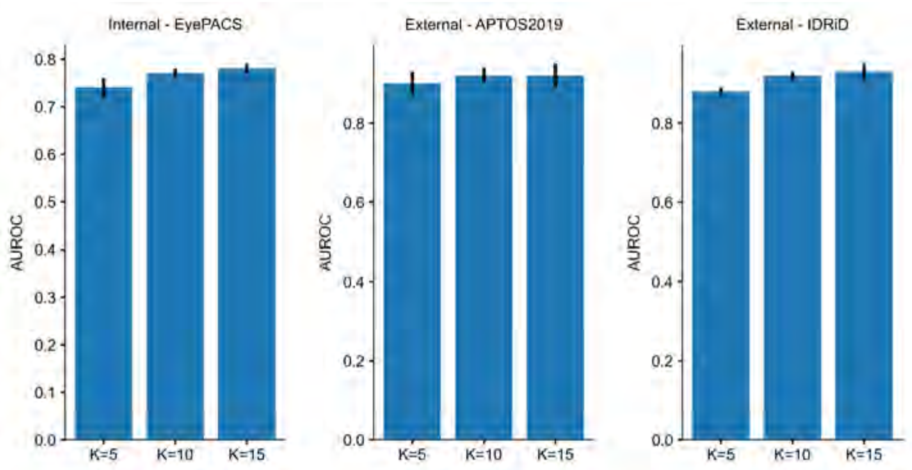

Fig. 1. Effects of hyperparameter cluster number K. We investigated the performance with $K = 5, 10, 15$ and found that larger K (10,15) contributed to slightly better performance.

Table 1. Sample categorical balance for glaucoma detection (AIROGS). The unbalanced issue has been alleviated by all sampling strategies, in particular with DataDIVA.

Method	Total size	Non-glaucoma	Glaucoma	Disease proportion
Image pool D_u	71,009	2,268	68,741	3.2%
Random	1,200	32	1,168	2.7%
CorSet	1,200	93	1,107	7.8%
ALFA-Mix	1,200	146	1,054	12.2%
DataDIVA	1,200	164	1,036	13.7%

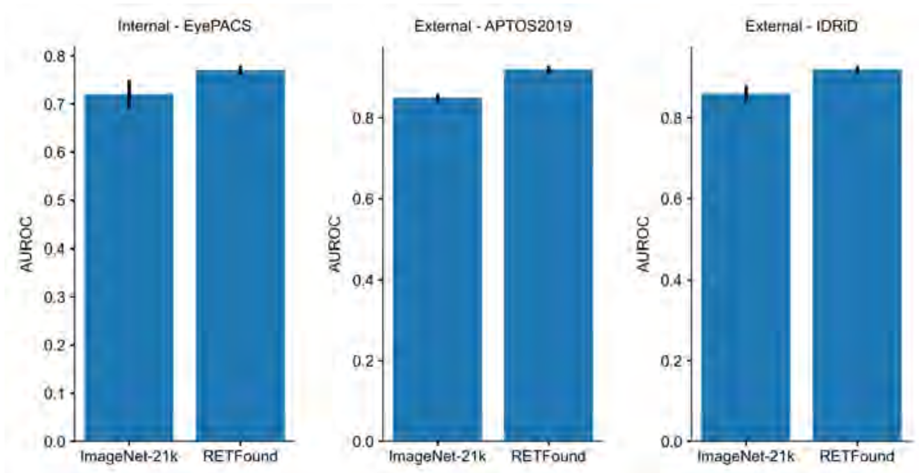

Fig. 2. Model performance using a foundation model in extracting features, compared with a model loading ImageNet-21k weights. All models use the same network backbone ViT-large.

Table 2. Sample categorical balance for referable diabetic retinopathy detection (EyePACS). The unbalanced issue has been alleviated by all sampling strategies, in particular with DataDIVA.

Method	Total size	Non-referable DR	DR	Disease proportion
Image pool D_u	35,126	6,873	28,253	19.6%
Random	600	115	428	19.2%
CorSet	600	142	428	23.7%
ALFA-Mix	600	153	447	25.5%
DataDIVA	600	166	434	27.8%