Enhance Eye Disease Detection using Learnable Probabilistic Discrete Latents in Machine Learning Architectures

Anirudh Prabhakaran* *ophv79@visitor.nus.edu.sg*
National University Singapore

YeKun Xiao* *yekun.xiao@u.nus.edu*
Asian Institute of Digital Finance
National University Singapore

Ching-Yu Cheng *chingyu.cheng@nus.edu.sg*
School of Medicine
National University Singapore

Dianbo Liu *dianbo@nus.edu.sg*
School of Medicine and College of Design & Engineering
National University Singapore

1 Supplementary Information

A Datasets

A.1 ODIR

The ODIR-5K dataset, released by a Chinese team, is a multi-label classification dataset of fundus images. It was made available during the "Intelligent Eye" competition hosted by Peking University in 2019 and contains paired fundus images of the left and right eyes from 5000 patients, with labeled data for 3500 cases released for training.

Unlike other fundus datasets (like CHASE and DRIVE), ODIR-5K's key distinction lies in providing paired data for both left and right eyes along with relevant descriptions. This setup allows for the data to be used both as 7000 individual cases or as 3500 paired cases for exploring consistencies or other aspects. ODIR-5K provides 8 categories of labels: Normal (N), Diabetes (D), Glaucoma (G), Cataract (C), Age-related Macular Degeneration (AMD), Hypertension (H), Myopia (M), and Other diseases/abnormalities (O). These categories encompass many common ophthalmic diseases and present an opportunity to explore the long-tail problem of datasets where some categories have much less data than others.

The significance of developing local datasets like ODIR-5K is profound for constructing medical diagnostic AI that is more suitable for the Chinese population, addressing biases that datasets can bring into AI research. By tailoring datasets to reflect the demographic and clinical profiles of the local population, AI systems can deliver more accurate and reliable diagnoses, ultimately enhancing patient care.

A.2 RFMiD

The RFMiD 2.0 dataset is a multi-label classification dataset for fundus images. It is an updated version of the RFMiD (1.0 version), which was used in the RIADD Challenge at ISBI 2021. Released in 2023, the update primarily includes modifications to the label categories, refining the previously general 'Other' category into specific rare diseases, thereby enhancing the quality of the labels. The dataset contains a total of 3,200 cases, and as of now, images and labels for all cases have been provided. With annotations for

*These authors contributed equally to this work.

Table 1: Details about ODIR-5K dataset

Tag	Training Set	Off-Site Test Samples	On-Site Test Samples	Total
N	1135	161	324	1620
D	1131	162	323	1616
G	207	30	58	307
C	211	32	64	243
A	171	25	47	295
H	94	14	30	138
M	177	23	49	249
O	944	134	268	1346

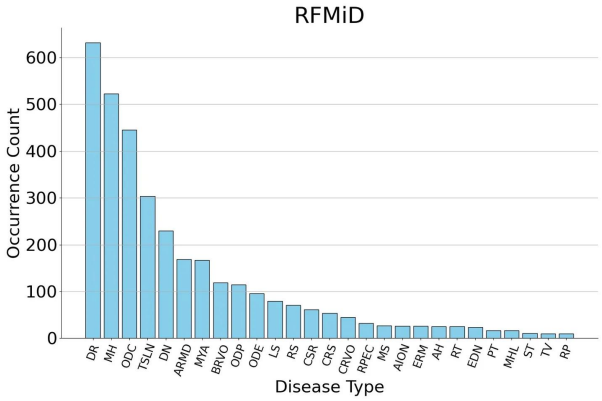

Figure 1: Class count of data points in the RFMiD Dataset

45 different ocular diseases, RFMiD 2.0 holds the distinction of having the most disease categories among publicly available fundus datasets.

Fundus images, due to their similarity to natural images, have been used as a benchmark for the generalizability of methods in the medical field by researchers of natural images. This includes datasets such as CHASE, DRIVE, etc., which have been discussed in our previous articles. As the fundus dataset with the most annotated categories available publicly, RFMiD 2.0 is of value to both general multi-label classification researchers and those working on computer-aided diagnostic research related to vision. Given the large number of annotation categories in the dataset, some categories inevitably have very few samples, making this dataset also relevant for research into multi-label long-tail problems and issues related to limited samples in the medical field.

A.3 JSIEC

The JSIEC dataset consists of a total of 209,494 fundus images, covering 39 categories. This article introduces a subset of the JSIEC dataset, which includes 1,000 fundus images distributed across 39 categories. This dataset collects fundus images from 7 different data sources for the development and validation of deep learning algorithms. The primary datasets for training, validation, and testing come from the Picture Archiving and Communication System (PACS) of the Joint Shantou International Eye Center (JSIEC) in China, China's Lifeline Express Diabetic Retinopathy Screening System (LEDRS), and the Eye Picture Archiving and Communication System (EyePACS) in the USA.

Millions of people worldwide are affected by fundus diseases such as Diabetic Retinopathy (DR), Age-related Macular Degeneration (AMD), Retinal Vein Occlusion (RVO), Retinal Artery Occlusion (RAO), Glaucoma, Retinal Detachment (RD), and fundus tumors. Among these, DR, AMD, and Glaucoma are

the most common causes of vision impairment in most populations. Without accurate diagnosis and timely appropriate treatment, these fundus diseases can lead to irreversible blurring of vision, visual distortion, field defects, and even blindness. However, in rural and remote areas, especially in developing countries, there is a lack of ophthalmic services and ophthalmologists, making early detection and timely referral for treatment often inaccessible. Notably, fundus photography provides basic detection of these diseases and is available and affordable in most parts of the world. Non-professionals can handle fundus photographs and send them online to major ophthalmic institutions for follow-up. Artificial intelligence technology can be used to assist in diagnosis.

Table 2: Details about the JSIEC dataset

Disease	**No. of Images**	**Disease**	**No. of Images**
Normal	38	CRVO	22
Tessellated	13	Yellow-white spots-flecks	29
Large Optic Cup	50	Cotton-wool spots	10
DR1	18	Vessel tortuosity	14
Possible glaucoma	13	Chorioretinal atrophy-coloboma	15
Optic atrophy	12	Preretinal hemorrhage	10
DR2	49	Fibrosis	10
DR3	39	Laser Spots	20
Sever hypertensive	15	Silicon oil in eye	19
Disc swelling and elevation	13	Blur fundus without PDR	111
Dragged Disc	10	Blur fundus with suspected PDR	45
Congenital disc abnormality	10	RAO	16
Retinitis pigmentosa	22	Rhegmatogenous RD	57
Bietti Crystalline Dystrophy	8	CSCR	14
Peripheral Retinal Degeneration and Break	14	VKH Disease	14
Myelinated Nerve Fiber	11	Maculopathy	74
Vitreous Particles	14	ERM	26
Fundus Neoplasm	8	MH	23
BRVO	44	Pathological Myopia	54
Massive Hard Exudates	13		

B Images

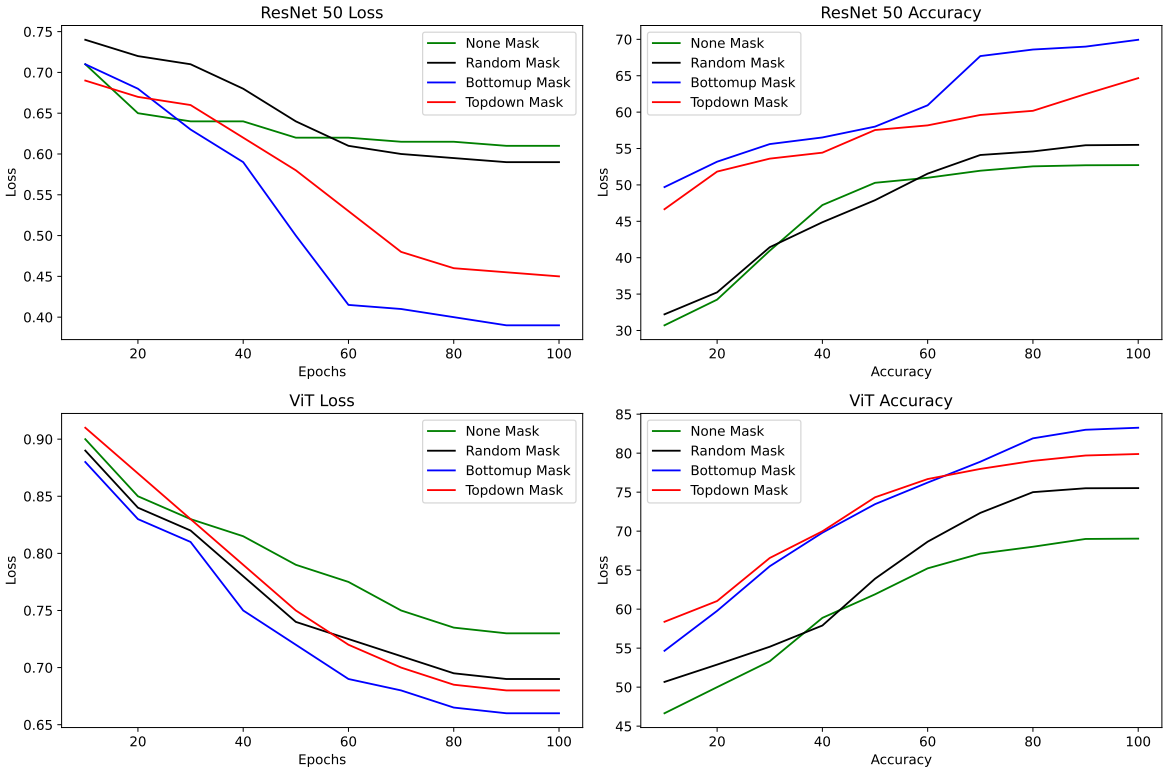

Figure 2: These plots show the loss curves and accuracy curves for the different models used. The top row has the metrics for ResNet18 model, and the bottom row has the metrics for the Vision Transformer model. We also plot metrics for each of the masks evaluated: `none`, `random`, `topdown` and `bottomup`.

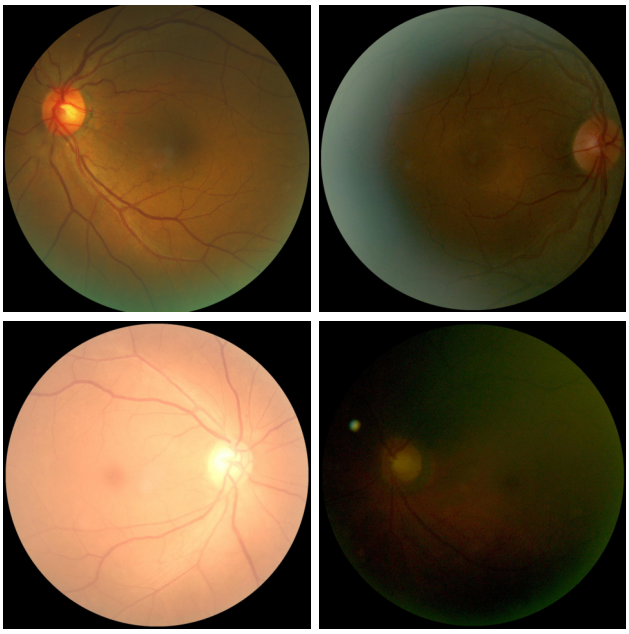

Figure 3: Fundus images from datasets with the minimum and maximum entropy. The top row consists of diabetic and normal fundus images, respectively, which have the minimum entropy. The bottom row consists of diabetic and normal fundus images, respectively, which has maximum entropy. We note that the model has highest confidence in its predictions when the image is clear, and the least confidence when the image is under or over-exposed.

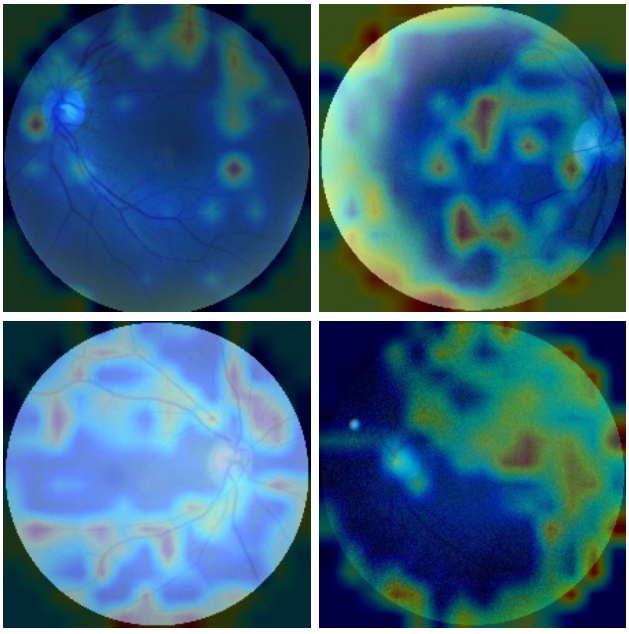

Figure 4: GradCAM analysis of the attention maps of the Vision Transformer. The top row consists of fundus images of diabetic and normal patients with minimum entropy. The bottom row consists of fundus images of diabetic and normal patients with maximum entropy. On top of these images, we apply the attention map computed using GradCAM to understand which parts are considered important by the model.