

Appendix for "Generalizing to Unseen Domains in Diabetic Retinopathy with Disentangled Representations"

Peng Xia^{1*}, Ming Hu^{1*}, Feilong Tang¹, Wenxue Li¹, Wenhao Zheng²,
Lie Ju¹, Peibo Duan¹, Huaxiu Yao²✉, and Zongyuan Ge¹✉

¹ Monash University, Melbourne, Victoria, Australia

² UNC-Chapel Hill, Chapel Hill, NC, USA

richard.peng.xia@gmail.com, huaxiu@cs.unc.edu, zongyuan.ge@monash.edu

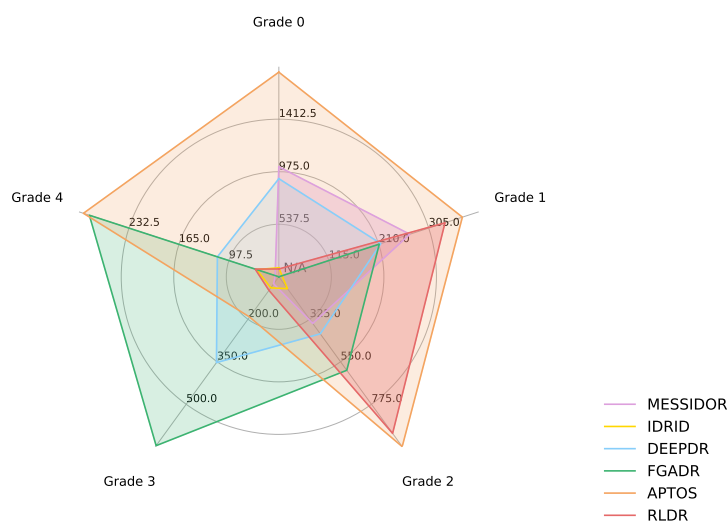


Fig. 1: Data distribution for each category in 6 datasets.

* Equal Contribution

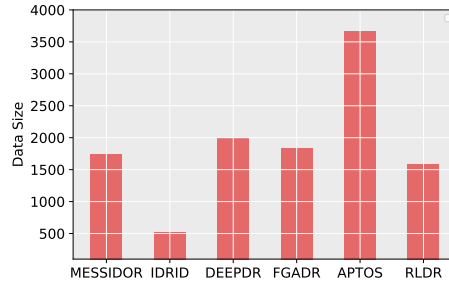


Fig. 2: Distribution of individual datasets.

Hyperparameters	DECO (DG)	DECO (ESDG)	Comparison methods
Epochs	100	100	100
Learning Rate	5×10^{-4}	5×10^{-4}	5×10^{-4}
Weight Decay	10^{-4}	10^{-4}	10^{-4}
Batch Size	32	32	32
Warm Start Epochs [†]	30	35	-
Class Prototype Mixup Parameter α_c	0.5	0.55	-
Domain Prototype Mixup Parameter α_d	0.5	0.4	-
γ_c	0.2	0.2	-
γ_d	0.2	-	-

Table 1: Hyperparameters for experiments. [†] denotes we adopt a warm start strategy of running vanilla ERM for the first few epochs to ensure reliable disentanglement. Interpolation coefficient $\lambda_c \sim \text{Beta}(\alpha_c, \alpha_c)$ and $\lambda_d \sim \text{Beta}(\alpha_d, \alpha_d)$.