Proliferative Diabetic Retinopathy Detection with Multimodal Deep Learning

Yanis Schärer

Supervisor: Prof. Dr. Raphael Sznitman
Institutions: University of Bern, ARTORG Center for Biomedical Engineering Research
University Hospital Bern (Inselspital), Department of Ophthalmology

Introduction

Diabetic retinopathy (DR) is a prevalent and potentially blinding eye disease that affects individuals with diabetes. Early and accurate detection of DR is crucial for timely intervention to prevent irreversible vision loss. However, manual diagnosis of DR can be tedious and error-prone, underlining the need for automated and reliable detection systems. The first aim of this thesis is to classify infrared (IR) fundus images and optical coherence tomography (OCT) scans into non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). Additionally, the thesis explores the fusion of different imaging modalities to investigate the potential benefits of combining IR fundus images and OCT scans.

Methods

The dataset used in this study was acquired from the University Hospital of Bern. The models are trained and validated using triplets consisting of an IR fundus image, a horizontal OCT scan and a vertical OCT scan from the same eye.

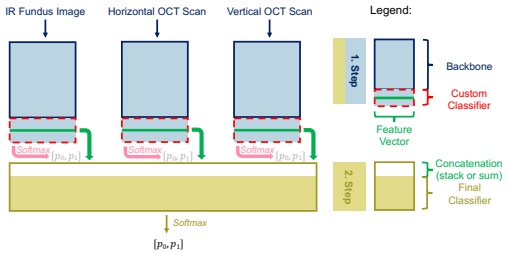

Figure 1: Architecture of the model. p_0 and p_1 are the probabilities of class 0 (NPDR) and class 1 (PDR), respectively.

The approach to fuse modalities involves pre-training unimodal backbone models in combination with a customized classifier, one for each modality. During the fusion process, all three images are passed through their respective unimodal model, and the resulting feature vectors, that is, the last linear layer of each model before classification, are extracted and subsequently fused together using a concatenation operation. Finally, the fused representation is fed into the final classifier, which generates the class prediction as the output.
In order to provide reliable visual interpretability, two established but distinct methods, HiResCAM and occlusion maps, were chosen and implemented.

Results

Training and evaluation was repeated three times to consider statistical variability. The AUC of the final multimodal model is 0.84 ± 0.01 and the AP is 0.68 ± 0.02. The narrow standard deviation indicates minimal variability across training runs. Among the unimodal backbones, the model trained on horizontal OCT scans consistently exhibits the highest AUC and AP (0.83 ± 0.03 and 0.63 ± 0.02, respectively).

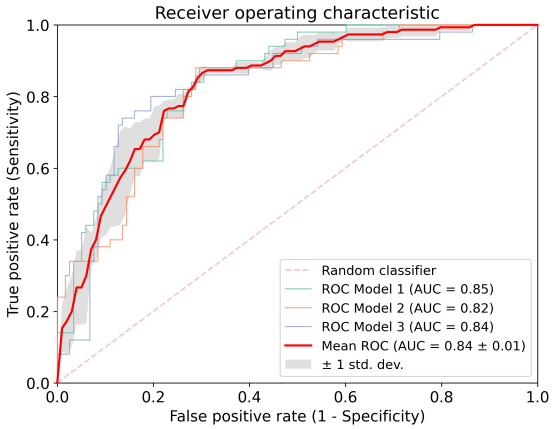

Figure 2: Receiver operating characteristic (ROC) curve for the multimodal model.

Fig. 3 representatively displays the heat maps of a correctly classified PDR (class 1) sample. The focus of the model on the IR fundus images is on areas known to be associated with the presence of neovascularization, a characteristic feature of PDR. In the OCT scans, the region of and around the macula is consistently highlighted, showing its significance in the model's decision-making process.

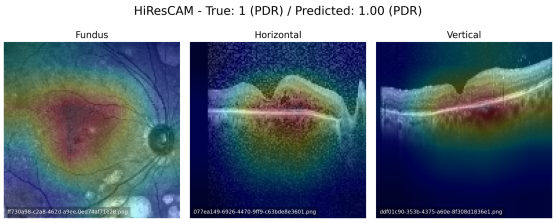

Figure 3: HiResCAM applied to a correctly classified PDR (class 1) triplet.

Discussion

The results are promising, particularly when considering the small size and the low quality of the dataset. The findings of this thesis demonstrate the potential benefits of multimodal fusion, as it enhances the performance of PDR detection compared to unimodal classification. It is important to note that the absence of a benchmark, such as the evaluation by medical specialists, limits the ability to contextualize the performance of the model in a clinical setting.