Stable classification of diabetic structures from incorrectly labeled OCTA enface images using multiple instance learning

Philipp Matten, MSc; Julius Scherer, MSc; Thomas Schlegl, PhD; Jonas Nienhaus, MSc; Heiko Stina, MD; Andreas Polleitz, MD; Wolfgang Drexler, PhD; Rainer A. Leitgeb, PhD; Tilman Schmoll*; PhD

1Center for Medical Physics and Biomedical Engineering, Medical University of Vienna; 2Dept. of Ophthalmology, Medical University of Vienna; 3Carl Zeiss Meditec, Inc., Dublin, CA, USA

PURPOSE

- Diabetic retinopathy (DR) is one of the leading vision impairments in working-aged adults
- DR progression can be controlled when diagnosed early
- Early signs of lesions start occurring in the periphery of the retina – widefield OCT angiography (wOCTA) en face images provide this information
- We propose a novel multiple instance learning (MIL)-based CNN classifier for classifying DR in wOCTA en face images with weakly (binary) labeled data

METHODS

- wOCTA images cover a field of view (FoV) of 65°
- Used a database consisting of 354 wOCTA en face images
- 257 en face images of diabetic patients and 97 images of healthy volunteers, split into:
  - Training (211/64)
  - Validation (24/24)
  - Testing (22/8)
- Image dimensions are 1536 x 2048 x 2048 samples for every volume, covering 6 x 18 x 18 mm³ (Figure 1 (c))
- Figure 1 shows the concept of MIL:
  - A bag is a collection of sub-structures/features, so-called instances (Figure 1 (a))
  - Information is assumed to be held at the instance level
  - Only binary bag labels available for the entire dataset – whether it’s from a diabetic or not (Figure 1 (b))
  - MIL-processing requires image normalization and creation of 10 x 10 instances per bag (Figure 2)
- We benchmarked our network, MIL-ResNet14, against proven capable DR-classifiers: ResNet14 & VGG16

RESULTS

- Figure 3 shows the classification results of carefully put together test dataset, containing the entire spectrum of severity of DR in our dataset
- Table 1 shows MIL-ResNet14 reached an F1-score (harmonic mean of precision and recall) of 0.95 and outperformed ResNet14 and VGG16 (Table 1)
- MIL-ResNet14 was trained until the AUC-accuracy didn’t improve anymore
- MIL-ResNet14 generalized better during training while ResNet14 and VGG16 required careful hyperparameter tuning
- We created Grad-CAM overlays for all corresponding original wOCTA en face – MIL-ResNet14 “paid closer attention” to all relevant biomarkers/lesions

CONCLUSIONS

- We developed a multiple instance learning-based classifier which outperformed state-of-the-art DR-Classifiers, ResNet14 and VGG16
- Grad-CAM images give a good idea of which parts of the image were deemed important but should not be mistaken with semantic annotations/segmentations
- MIL-ResNet14 has potential to be used as a clinical support tool for decision making and early detection of DR

• We created Grad-CAM overlays to show feature activation heatmaps of all different networks (Figure 4 (a)-(c) left & right)

Table 1: Classification metrics of the benchmarked classifiers on test dataset.

CONFLICTS OF INTEREST

None declared.

Pharmaceutical/Device Support

Bayer, Roche, Novartis, Carl Zeiss Meditec, Inc., Medical University of Vienna, Carl Zeiss Meditec, Inc., Inc.

Electrical Support

Bayer, Roche, Novartis, Carl Zeiss Meditec, Inc., Medical University of Vienna, Carl Zeiss Meditec, Inc., Inc.

Clinical Support

Bayer, Roche, Novartis, Carl Zeiss Meditec, Inc., Medical University of Vienna, Carl Zeiss Meditec, Inc., Inc.

Funding

Bayer, Roche, Novartis, Carl Zeiss Meditec, Inc., Medical University of Vienna, Carl Zeiss Meditec, Inc., Inc.

Poster # 1087 - C0181

Email: philipp.matten@meduniwien.ac.at