Beyond federated learning: fusion strategies for diabetic retinopathy screening algorithms trained from different device types

Sandipan Chakroborty, PhD[1]; Krunalkumar Ramanbhai Patel[1]; and Alexander Freytag, PhD[2]

[1]Center for Applications and Research in India, Carl Zeiss India (Bangalore) Pvt. Ltd., [2]Corporate Research and Technology, Jena, Carl Zeiss AG, Jena, Thuringia, Germany

Poster # 3531741

PURPOSE

- Diabetic retinopathy (DR) is one of the most common causes of blindness, and screening using fundus images helps to detect DR at early stage.
- Various **types of fundus cameras** are available today with different characteristics:
  - Image quality,
  - Portability, and
  - Field of view.
- To screen a large population, it is essential to develop an **automated DR screening system that is device agnostic**.

METHODS

- One model (M_{HH}) was trained exclusively on fundus images captured with a low-cost hand-held fundus camera **(VISUSCOUT® 100; ZEISS, Jena, Germany)**.
- The second model (M_{TT}) was trained on fundus images recorded with different **table-top fundus cameras** manufactured by various companies.
- We evaluated individual models and combination strategies on hold-out portion (20% in respective categories) of two datasets: one from hand-held cameras **(VISUSCOUT®)** and one from table-top cameras (seen in Messidor-2 dataset).
- $M_{HH} \equiv V1$ and $M_{TT} \equiv V2$ (Figure 1)
- <u>Score level combination (Fig. 1a) :</u> For the weighted score level combination, the optimum weights were found experimentally between 0 to 1.
- <u>Decision level combination (Fig. 1b) :</u> In decision level combination, a modified OR logic was used for referable DR.
- <u>System level combination (Fig. 1c):</u> For the system level combination, the metadata of the image was used to find out whether the image was captured using VISUSCOUT or a table-top fundus camera for routing the image to the appropriate algorithm.

CONCLUSIONS

DR screening systems which have been trained on different device data and with different deep learning models can be fused with various combination techniques for increased system performance.

Email: sandipan.chakroborty@zeiss.com, krunal.patel@zeiss.com, and alexander.freytag@zeiss.com

Disclosures: SC(E), KP(E), and AF(E), Carl Zeiss Meditec, Inc.

RESULTS

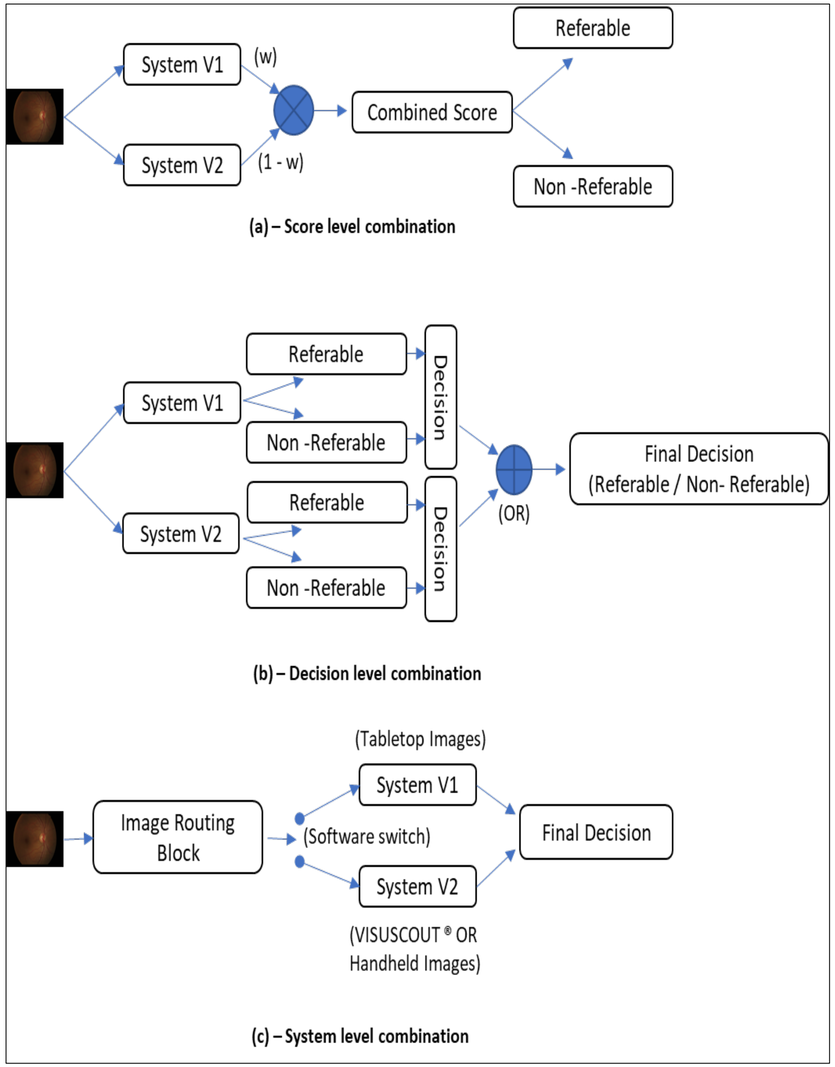

Fig. 1 Different systems for algorithm combinations

- Every single **model performs well in the data regime it was trained on** (M_{HH} on data from hand-held cameras, (M_{TT} on table-top camera data)
- Accuracy **drops significantly when tested with** data coming from **different devices**.
- For fusing both models into a single system, we observe **best screening results with score level fusion**, being clearly superior to decision level or system level fusion and insensitive to the data type.

Strategy \ Dataset	Hand-held (48,710 samples)	Table-Top (874 samples)	Combined (49,584 samples)
Only M_{TT}	89.43%	96.21%	91.49%
Only M_{HH}	96.48%	60.71%	93.49%
Score Level Fusion	96.48% (w = 1)	96.21% (w = 0)	95.74% (w = 0.4)
Decision Level Fusion	90.90%	75.65%	89.05%
System Level Fusion	Same as M_{HH}	Same as M_{TT}	90.90%

Fig 2. AUC scores obtained with different fusion strategies