Abnormality prediction from fundus images using Deep Learning and large amounts of data

Krunalkumar Ramanbhai Patel (ME)[1]; Alexander Freytag (PhD)[2]; Keyur Ranipa (M.Tech)[1]; Nathalia Spier (MSc)[3]; Alexander Urich (PhD)[4]

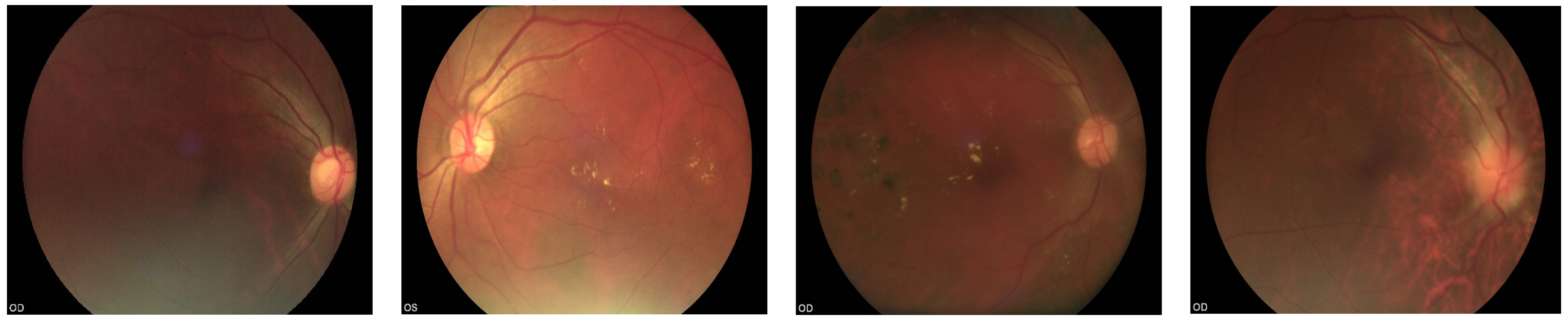
[1]CARIn, Carl Zeiss India (Bangalore) Pvt. Ltd., Bangalore, India; [2]Carl Zeiss AG, Jena, Germany; [3]Carl Zeiss AG, Oberkochen, Germany; [4]Carl Zeiss AG, Munich, Germany

Poster # PB0102

PURPOSE

Motivation

- Recent screening solutions for eye-related diseases mostly **focus on individual diseases** like DR, AMD, or Glaucoma.
- In contrast, **identifying any abnormality** in fundus images adds more value for the screening in general practitioner or non-ophthalmic segments such as corporate screening programs.
- Therefore, an **automated solution** to differentiate between any abnormal and normal fundus images is highly desirable.



(a) (b) (c) (d)

Figure 1: (a) a normal fundus image. (b-d) fundus images with different signs of abnormality.

Goal

- Develop an algorithm based on **Deep Neural Networks** (DNN) for reliable abnormality predictions with images acquired from Visuscout® 100 (ZEISS, Jena, Germany) fundus camera.
- Investigate the effect of **increasing amounts of training data**, especially for the comparison between a pre-trained DNN and a DNN initialized from scratch.

METHODS

- We conducted experiments on a 2D fundus image dataset obtained with a low-cost, hand-held fundus camera, Visuscout 100 (ZEISS, Jena, Germany). We collected data with **VISUHEALTH,** a **cloud based platform**, through which a referred patient is **remotely screened by a retina specialist**.
- The dataset is comprised of **148k images** (135k images of normal patients and 13k of patients with signs of abnormality). A medical expert annotated all images.
- The dataset was split into 80% for training and 20% for testing by randomly assigning **patients exclusively to either train or test sets**.
- We trained state-of-the-art DNNs **from scratch** or **pre-trained from ImageNet** with **varying data sizes** and repeated experiments four times for statistical relevance.

E-mail: krunal.patel@zeiss.com

Disclosures: KP (E), KR (E): Carl Zeiss India, Bangalore; AF(E): Carl Zeiss AG, Jena, Germany; NS(E): Carl Zeiss AG, Oberkochen, Germany; AU (E) : Carl Zeiss AG, Munich, Germany;

RESULTS

Overall accuracy (Figure 2)

- The best DNN (trained with all available training data and using pre-trained weights) **surpasses an AUC of 92%.**
- The resulting equal error rate is **86% for sensitivity and specificity**.

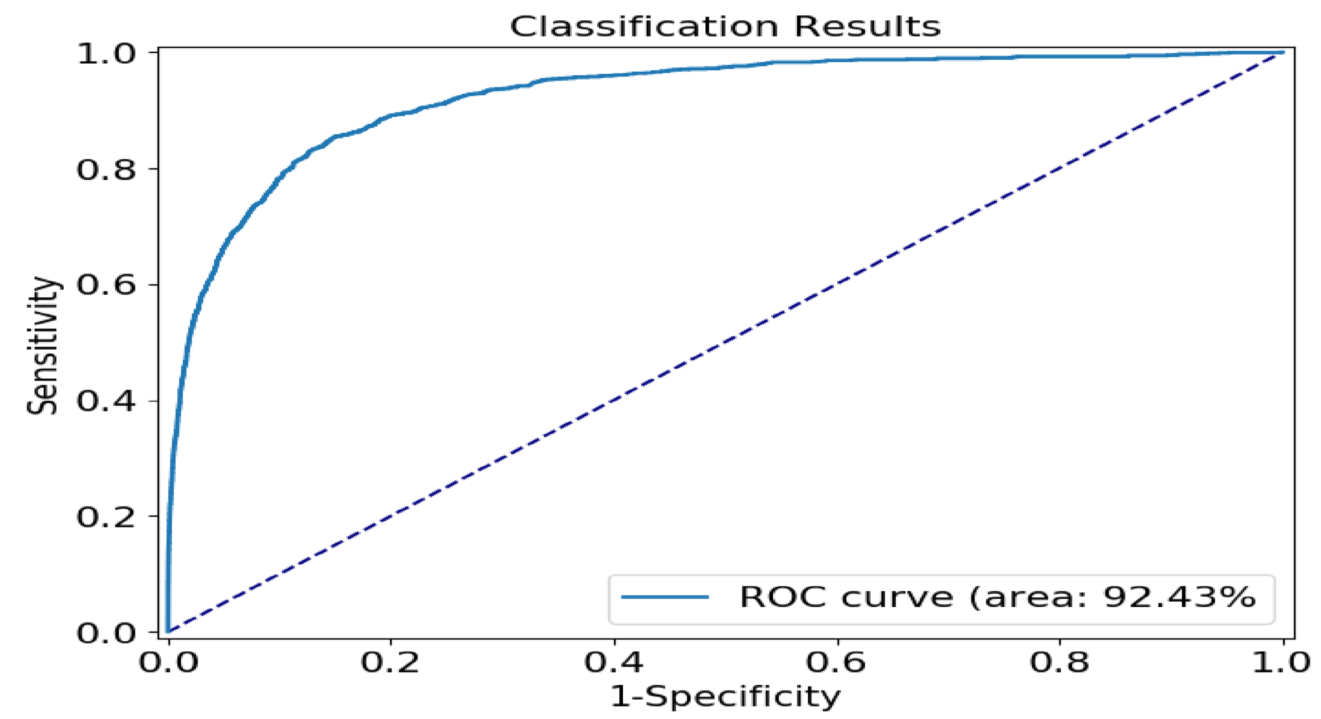


Figure 2: Results of predicting abnormality from 2D fundus images

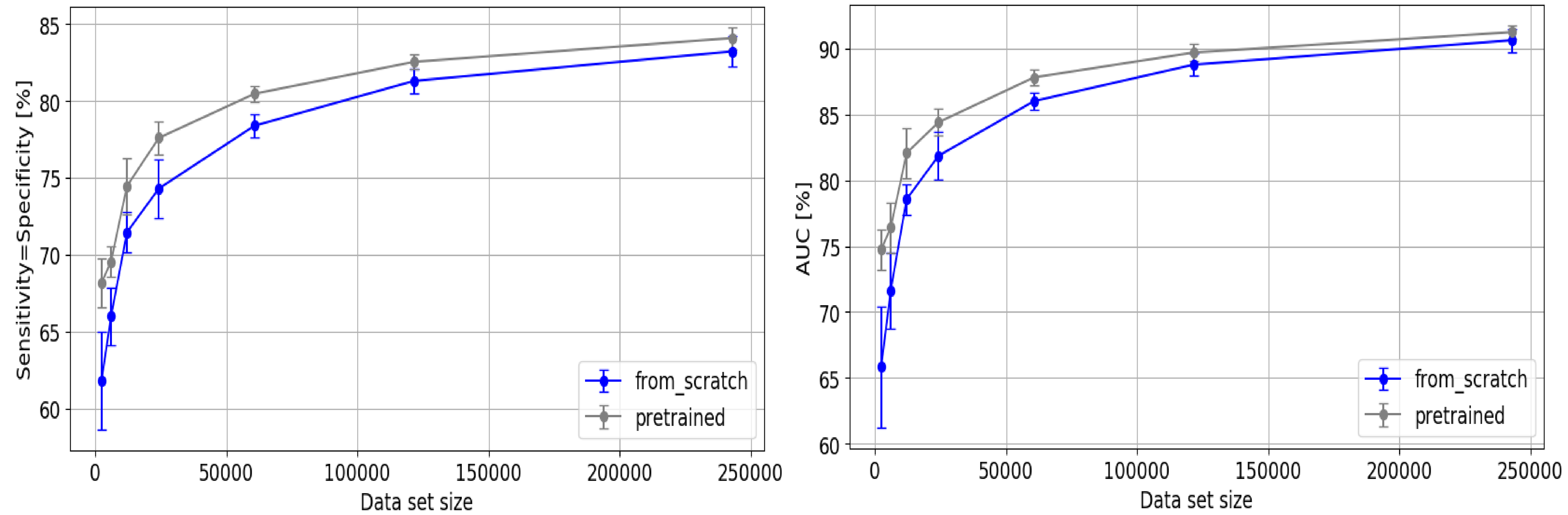


Figure 3: Dependency of the sensitivity-equals-specificity working points (left) and the AUC (right) on the amount of training data. Average of four repetitions of each experiment.

Accuracy vs. Dataset size (Figure 3)

- AUC and Sensitivity-equals-Specificity points (S=S) both clearly **benefit from large datasets**.
- Small datasets lead to large variations in model accuracy. We conclude that **models trained from small datasets are unreliable** and should not be used in medical applications.
- For **small** datasets (<100k images), **pre-training** provides a clear **benefit**.
- For **large** datasets (>100k images), **pre-training** can be **neglected**.

CONCLUSIONS

- **"Abnormality prediction in fundus images is possible"**: we evaluated abnormality prediction from 2D fundus images with deep neural networks. Leveraged by large amounts of data, our trained models surpassed 92% AUC on a real-world held-out test set of 14K images.
- **"More data gives better and more reliable models"**: we evaluated that increasing amounts of data help to improve the Sensitivity, Specificity, and AUC values of Deep Neural Networks.