Clinically applicable deep learning for diagnosis and referral in retinal disease

Jeffrey De Fauw, Joseph R. Ledsam, Bernardino Romera-Paredes, Stanislav Nikolov, Nenad Tomasev, Sam Blackwell, Harry Askham, Xavier Glorot, Brendan O'Donoghue, Daniel Visentin, George van den Driessche, Balaji Lakshminarayanan, Clemens Meyer, Faith Mackinder, Simon Bouton, Kareem Ayoub, Reena Chopra, Dominic King, Alan Karthikesalingam, Cían O. Hughes, Rosalind Raine, Julian Hughes, Dawn A. Sim, Catherine Egan, Adnan Tufail, Hugh Montgomery, Demis Hassabis, Geraint Rees, Trevor Back, Peng T. Khaw, Mustafa Suleyman, Julien Cornebise, Pearse A. Keane & Olaf Ronneberger

Presented by Ming
27-08-2018

Outline

- Background of the study
- Study design / methods
- Results
- Summary

Background - goal

Article | Published: 13 August 2018

Clinically applicable deep learning for diagnosis and referral in retinal disease

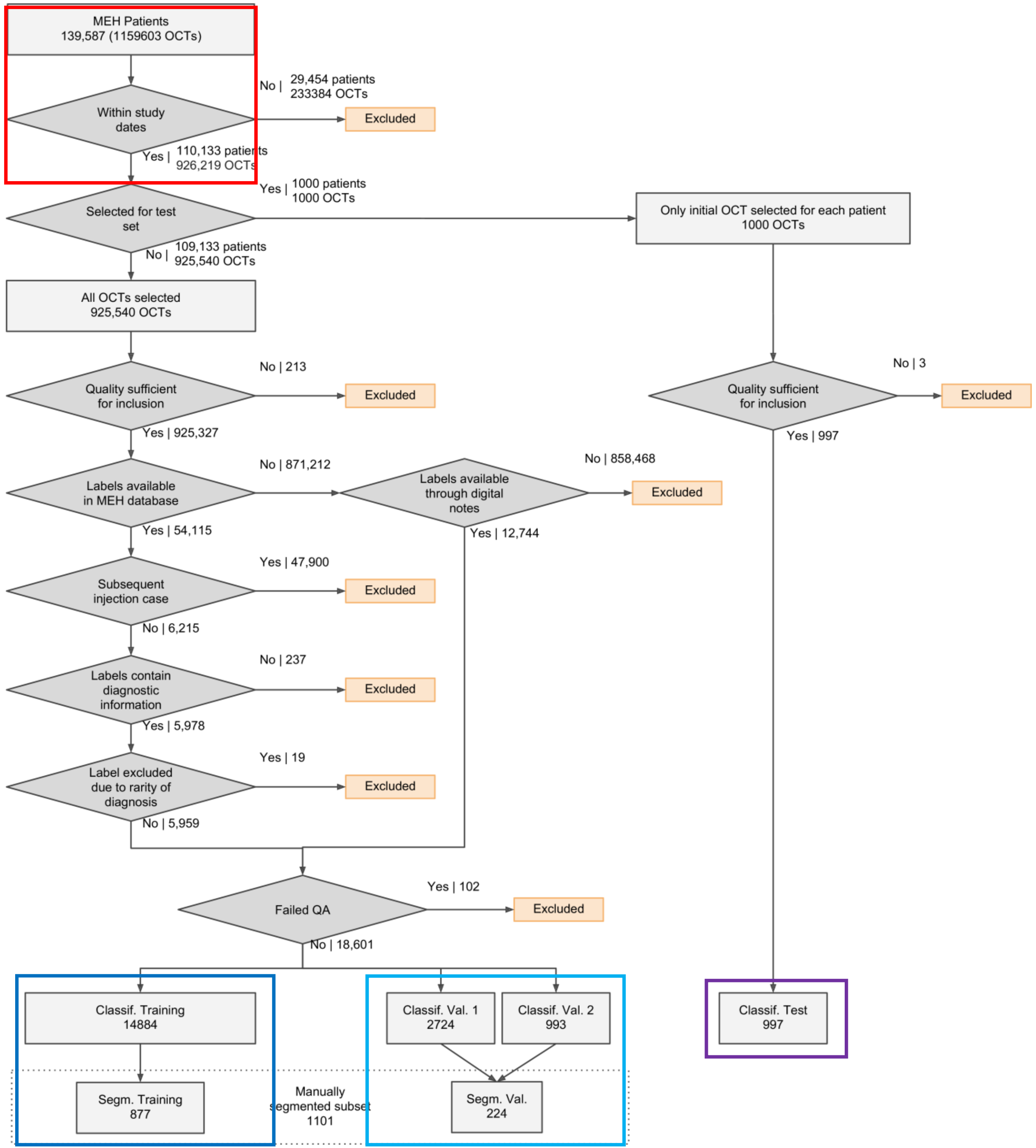
Jeffrey De Fauw, Joseph R. Ledsam, [...] Olaf Ronneberger

Nature Medicine (2018) | Download Citation

- to **produce automated diagnosis** that reach **performance of expert clinicians**
  - Referral decisions
  - Classification of 9 retinal pathologies
- on three-dimensional diagnostic scans of the retina

Background - dataset

- A retrospective cohort of patients who
  - attended Moorfields Eye Hospital between 1 June 2012 and 31 January 2017,
  - received OCT imaging as part of their routine clinical care
- Training set (n= 14884 /877)
  - Train the networks in the framework
- Validation set (993/224)
  - Fine-tune hyperparameters
- Test set (997)
  - Compare performance of the framework with human experts



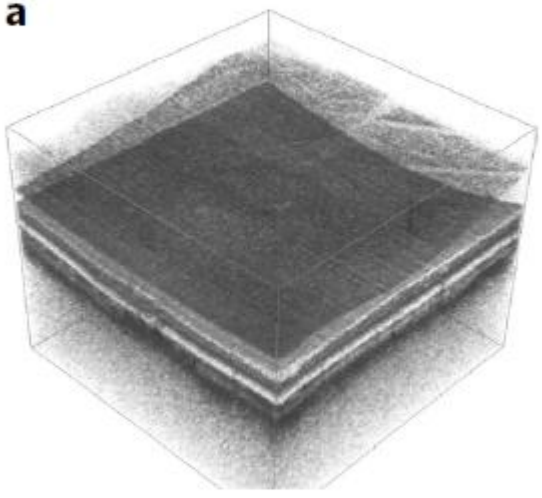
Supp. Fig. 11, Fauw *et al.* (2018) *Nat. Med.*

Outline

- Background of the study
- Study design / methods
- Results
- Summary

Optical coherence tomography (OCT) image

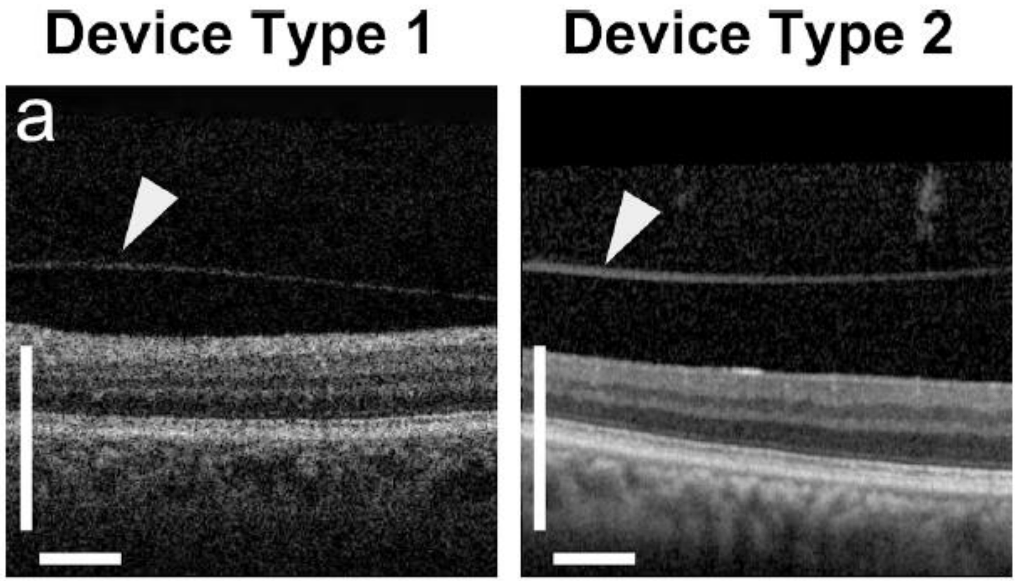
- 3-D medical imaging technique
- Analogous to 3-D ultrasonography (near-infrared light instead of sound waves)
- Resolution: ~5 μ m



Digital OCT scan

Image credit : Fauw *et al.* (2018) *Nat. Med.* Midwest City Eye Care Associates

Challenge I - technical variability



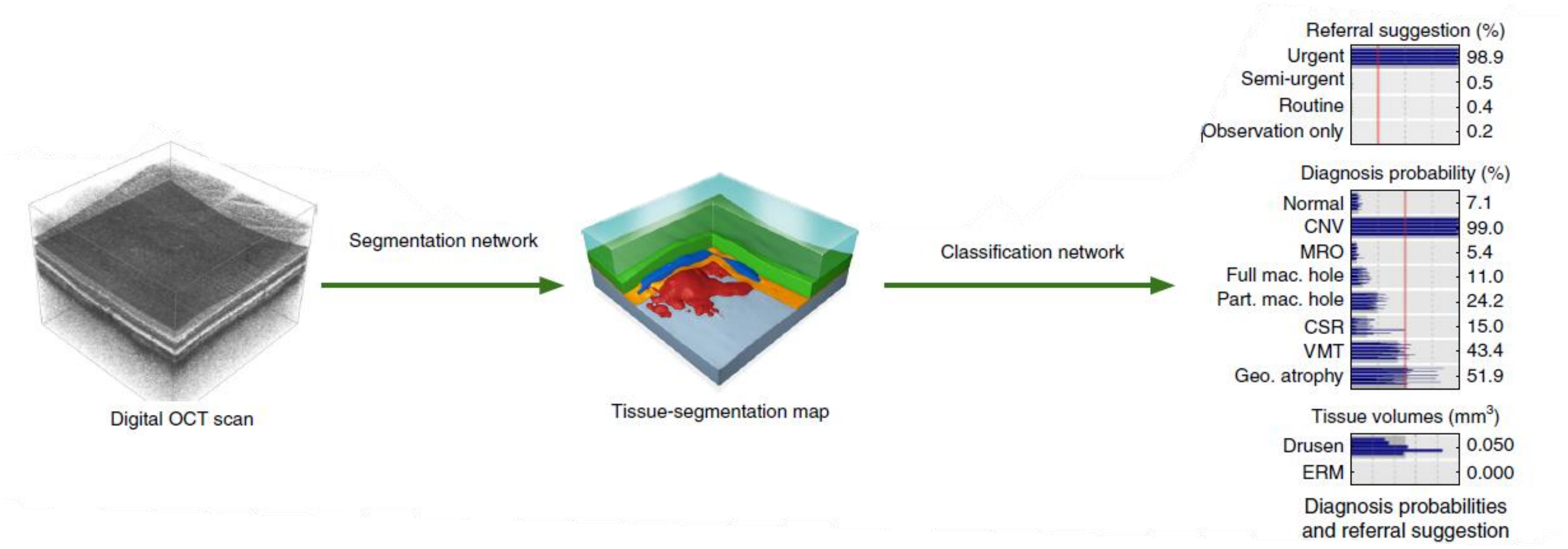
Topcon 3D OCT

Heidelberg Spectralis

- Model trained on images from one device may perform poorly on images from the other

Modified from Supp. Fig. 9, Fauw *et al.* (2018) *Nat. Med.*

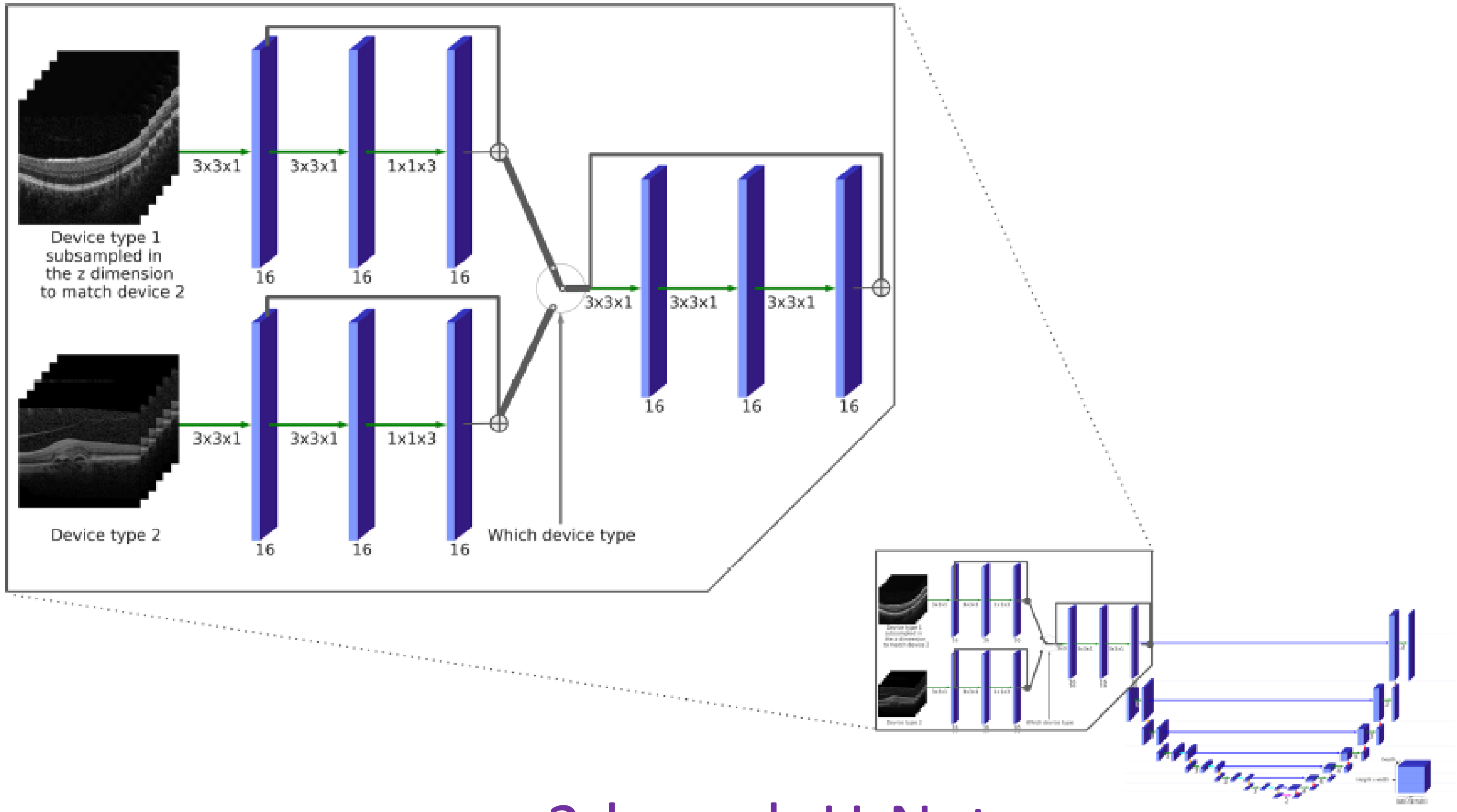
Method - proposed 2-stage framework



1. Creates device independent segmentation map
2. Prediction of referral decision and retinal morphologies

Modified from Fig. 1, Fauw *et al.* (2018) *Nat. Med.*

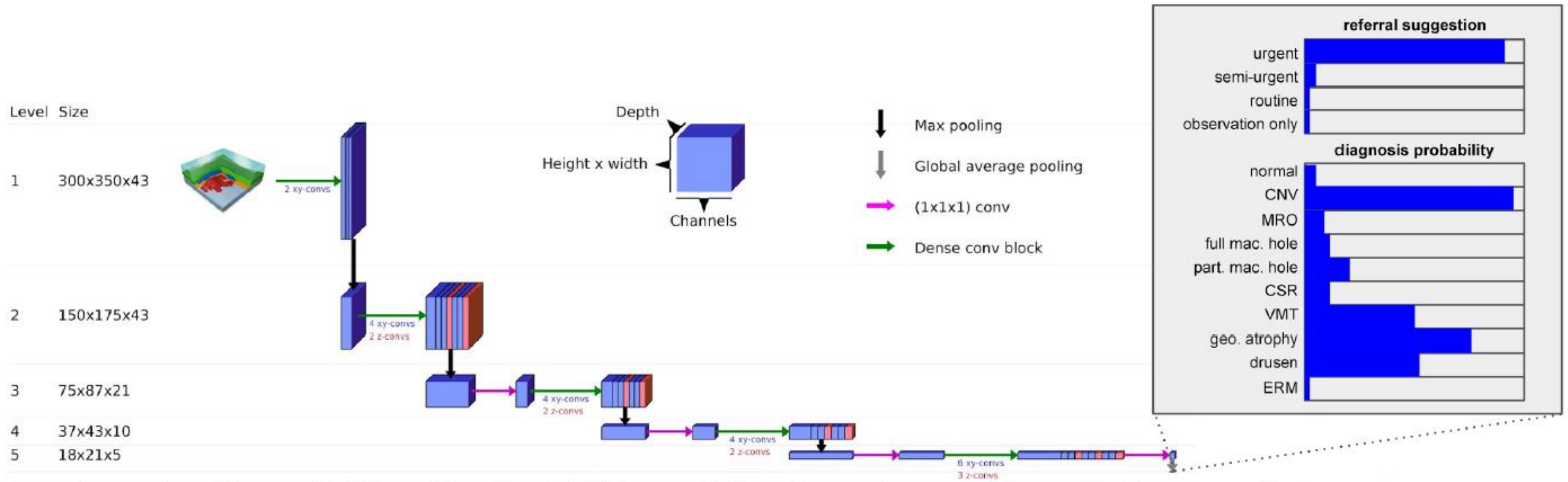
Architecture of the segmentation network



2-branch U-Net

Supp. Fig. 15, Fauw *et al.* (2018) *Nat. Med.*

Architecture of the classification network

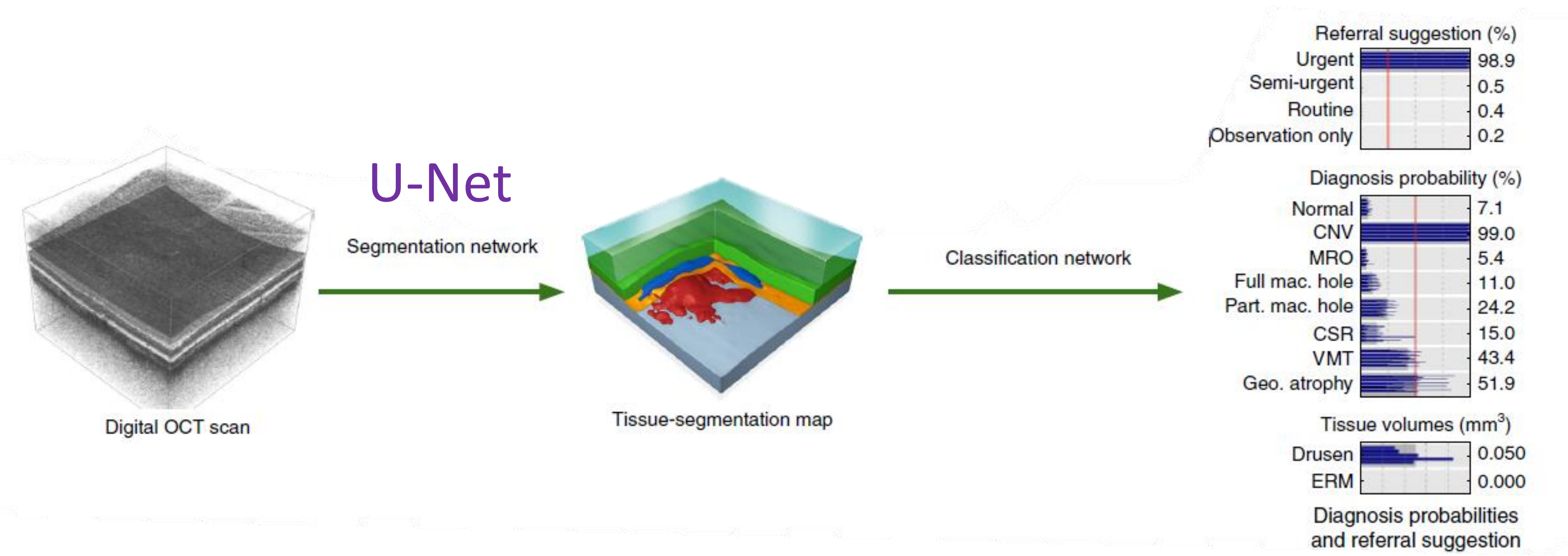


Supplementary Figure 16 | Classification CNN (convolutional neural network) used in the second stage of our approach. Blue and red boxes illustrate the 4D activation maps. Blue boxes are the result of a (3x3x1) convolution, while red boxes are the result of a (1x1x3) convolution.

Supp. Fig. 16, Fauw *et al.* (2018) *Nat. Med.*

Challenge II - Annotation

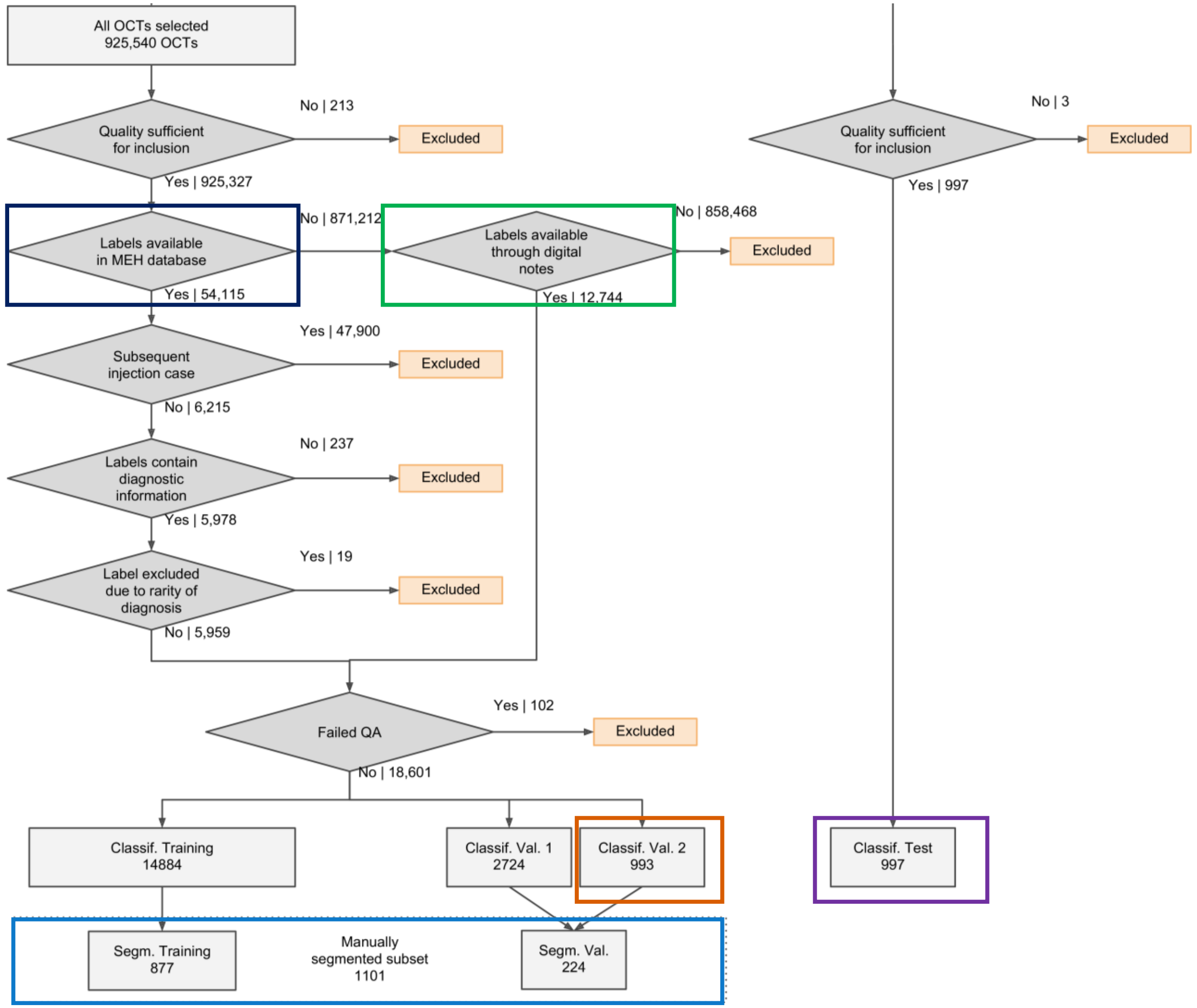
- Deep learning typically requires a large amount of data
- Many data are without label
- Unrealistic to annotate all images



Modified from Fig. 1, Fauw *et al.* (2018) *Nat. Med.*

Method - annotations scheme

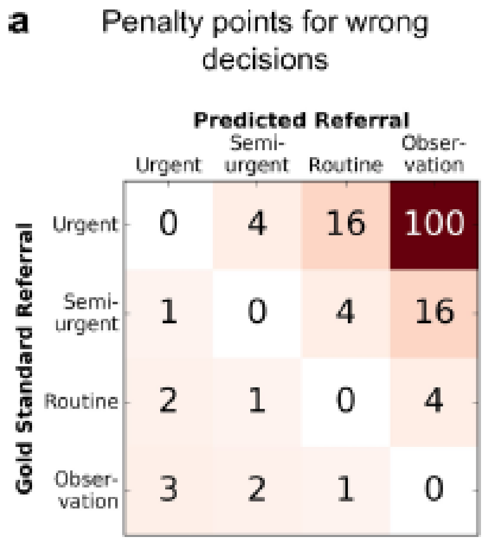
1. Referral decision
   - Electronic health record
   - 4 junior + 4 senior experts
2. Retinal pathologies
   - Automated notes search + manual review of the OCT scans
   - Manual annotation by 3 junior graders by 1 senior
   - 4 junior + 4 senior experts
3. Segmentation map
   - Manual annotations **on 3-5 out of 128 slices per OCT** with review by 1 senior



Modified from Supp. Fig. 11, Fauw *et al.* (2018) *Nat. Med.*

Penalty score

- Scale the penalty to the severity of adverse consequences for each misdiagnosis



Modified from Supp. Fig. 6, Fauw *et al.* (2018) *Nat. Med.*

The framework

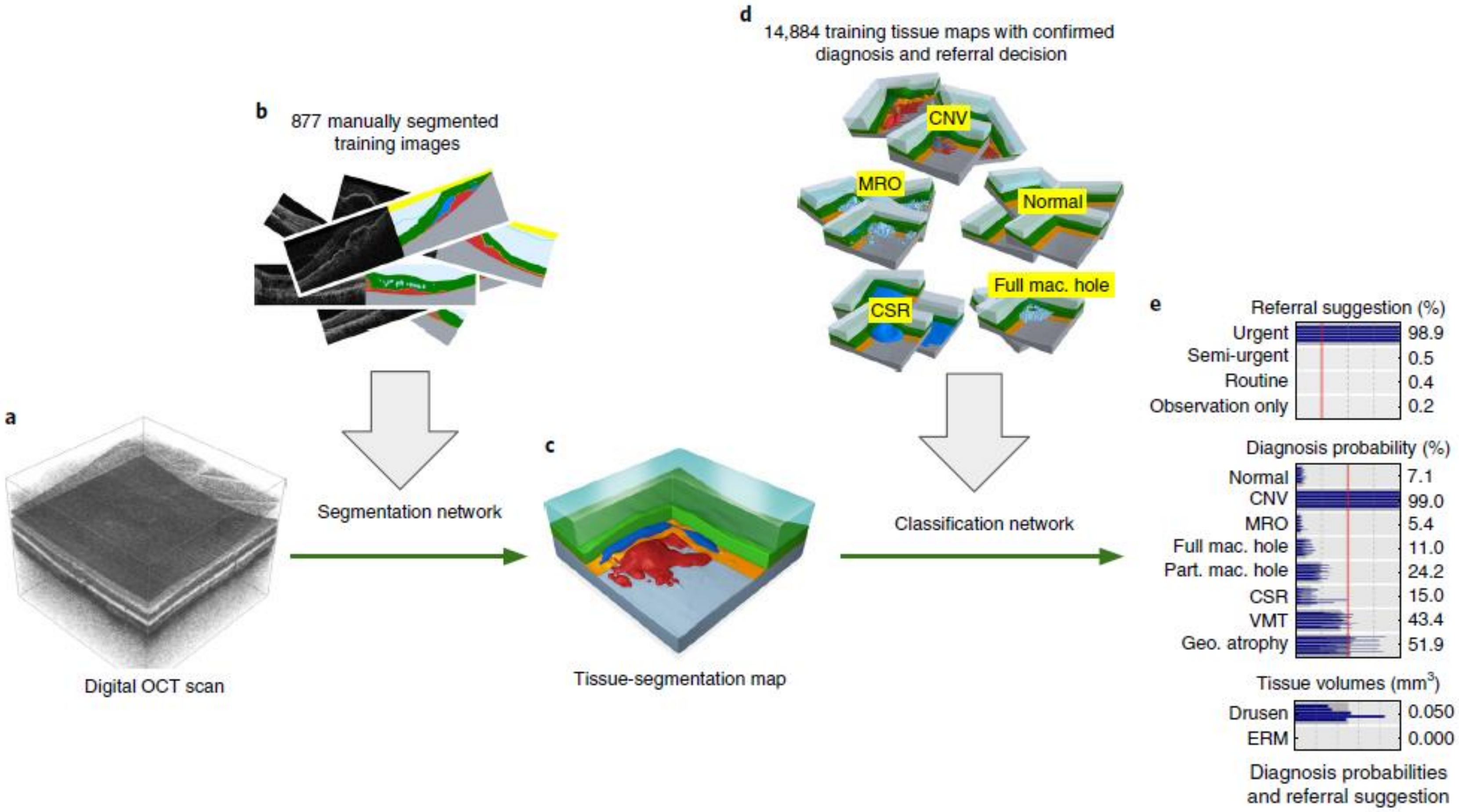
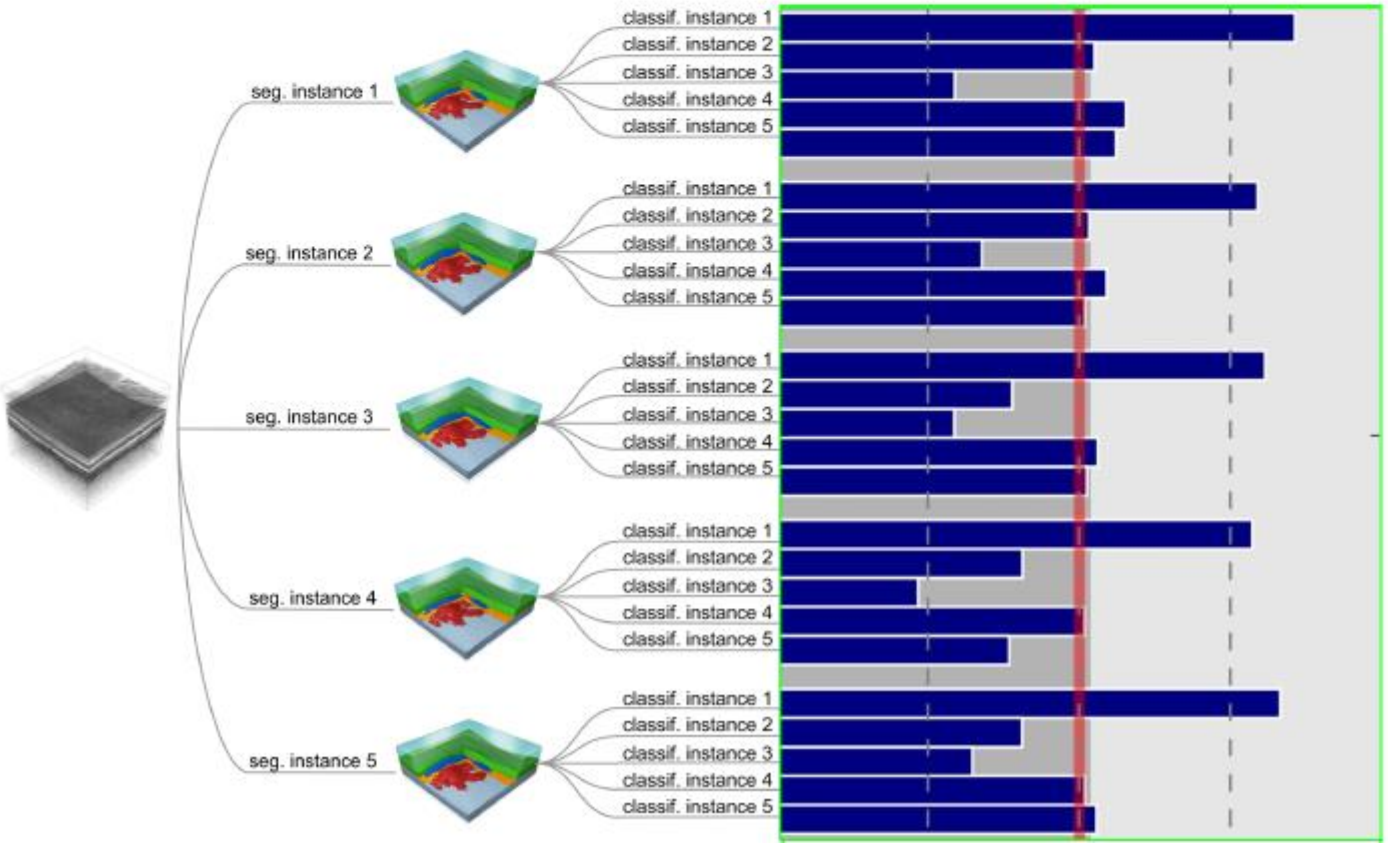


Fig. 1 | Our proposed AI framework. a, Raw retinal OCT scan ($6 \times 6 \times 2.3\,mm^3$ around the macula). **b**, Deep segmentation network, trained with manually segmented OCT scans. **c**, Resulting tissue segmentation map. **d**, Deep classification network, trained with tissue maps with confirmed diagnoses and optimal referral decisions. **e**, Predicted diagnosis probabilities and referral suggestions.

Fauw *et al.* (2018) *Nat. Med.*

Method – ensemble output

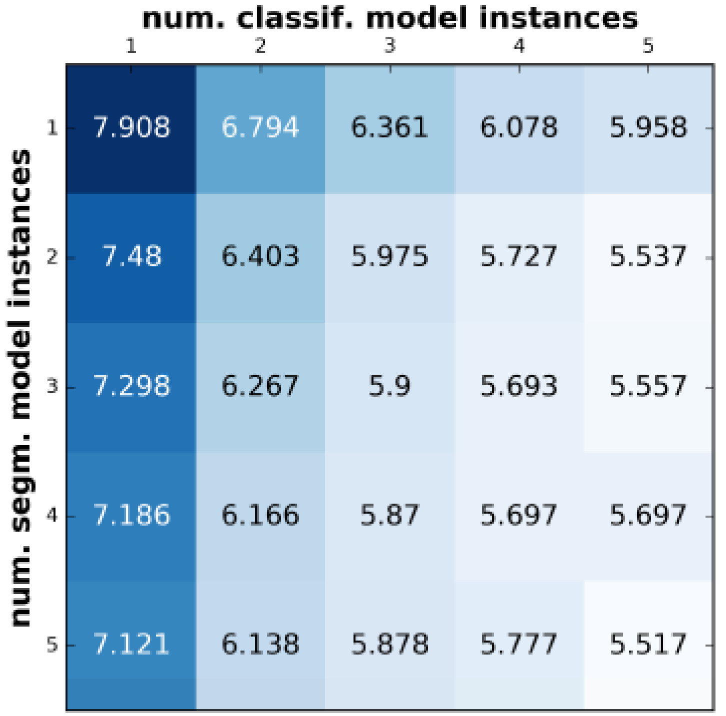
- 5 segmentation network instance x 5 classification network instances
- Averaging the probabilities of 25 outputs



Modified from Supp. Fig. 1 , Fauw *et al.* (2018) *Nat. Med.*

Error rate with different number of instances

Error rate [percent]



Outline

- Background of the study
- Study design / methods
- Results
- Summary

Result - segmentation

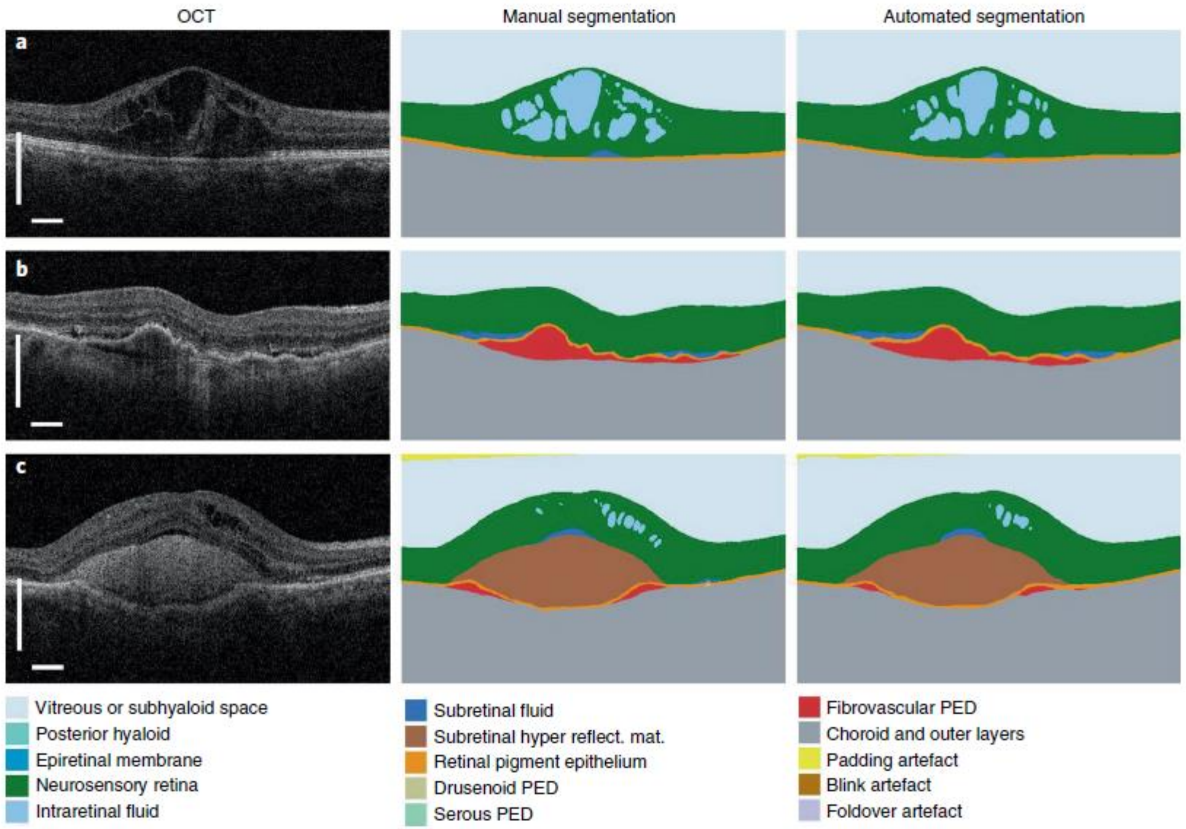


Fig. 2 | Results of the segmentation network. Three selected two-dimensional slices from the $n = 224$ OCT scans in the segmentation test set (left) with manual segmentation (middle) and automated segmentation (right; detailed color legend in Supplementary Table 2). **a**, A patient with diabetic

Fauw *et al.* (2018) *Nat. Med.*

Result – performance on referral decision I

- Framework vs 8 clinical experts
- Empty markers: expert performance using OCT, fundus image and summary notes
- Filled: using only OCT

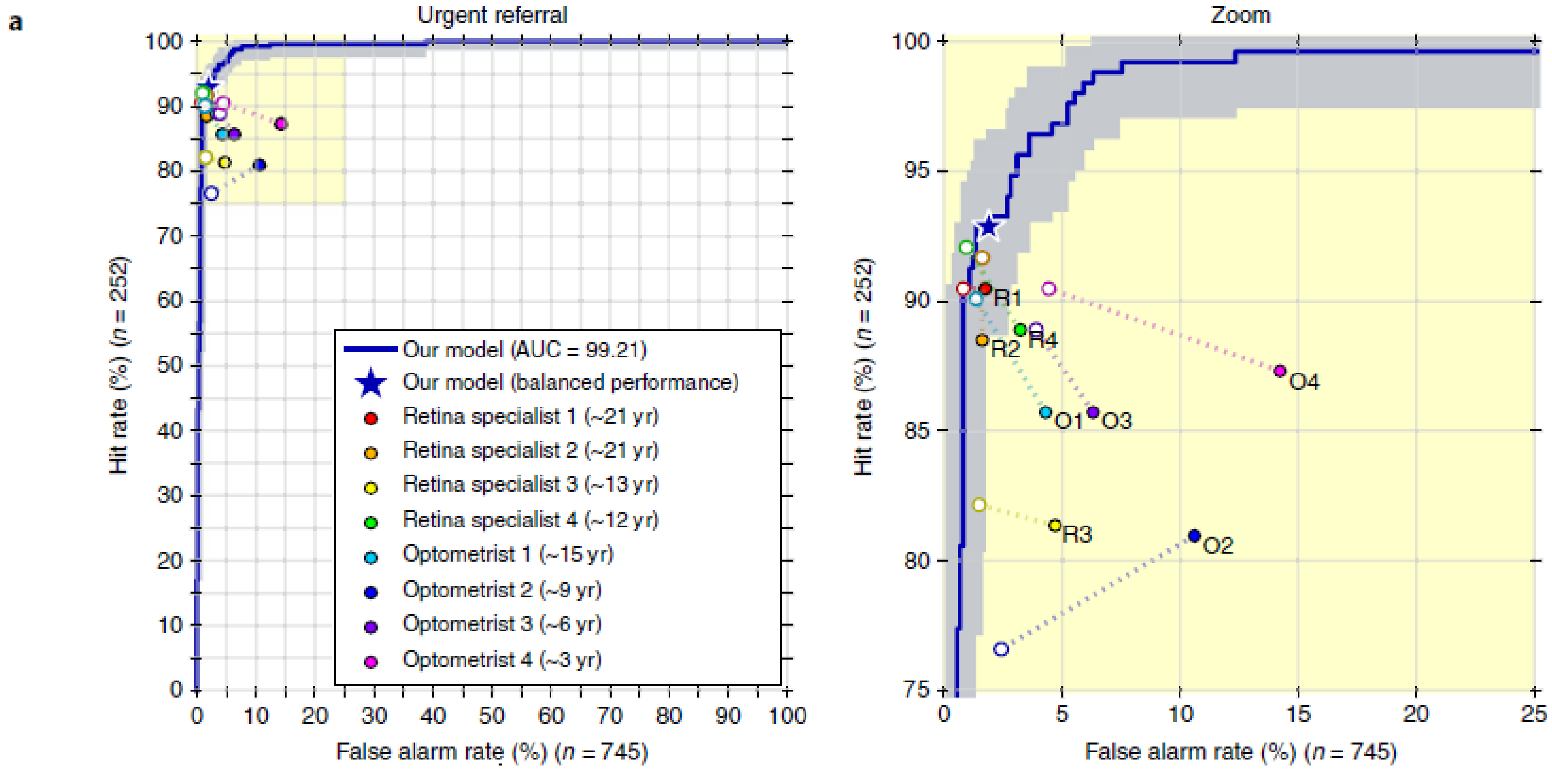


Fig. 3, Fauw *et al.* (2018) *Nat. Med.*

Result – performance on referral decision II

- Framework vs 2 senior experts

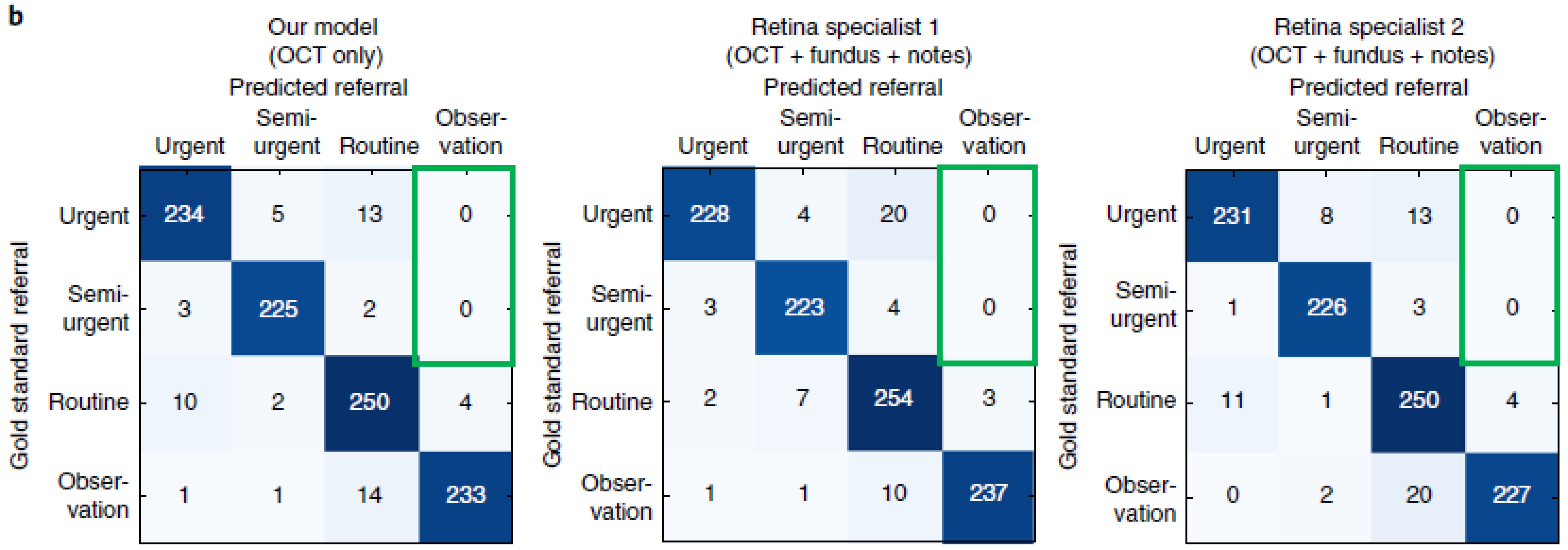
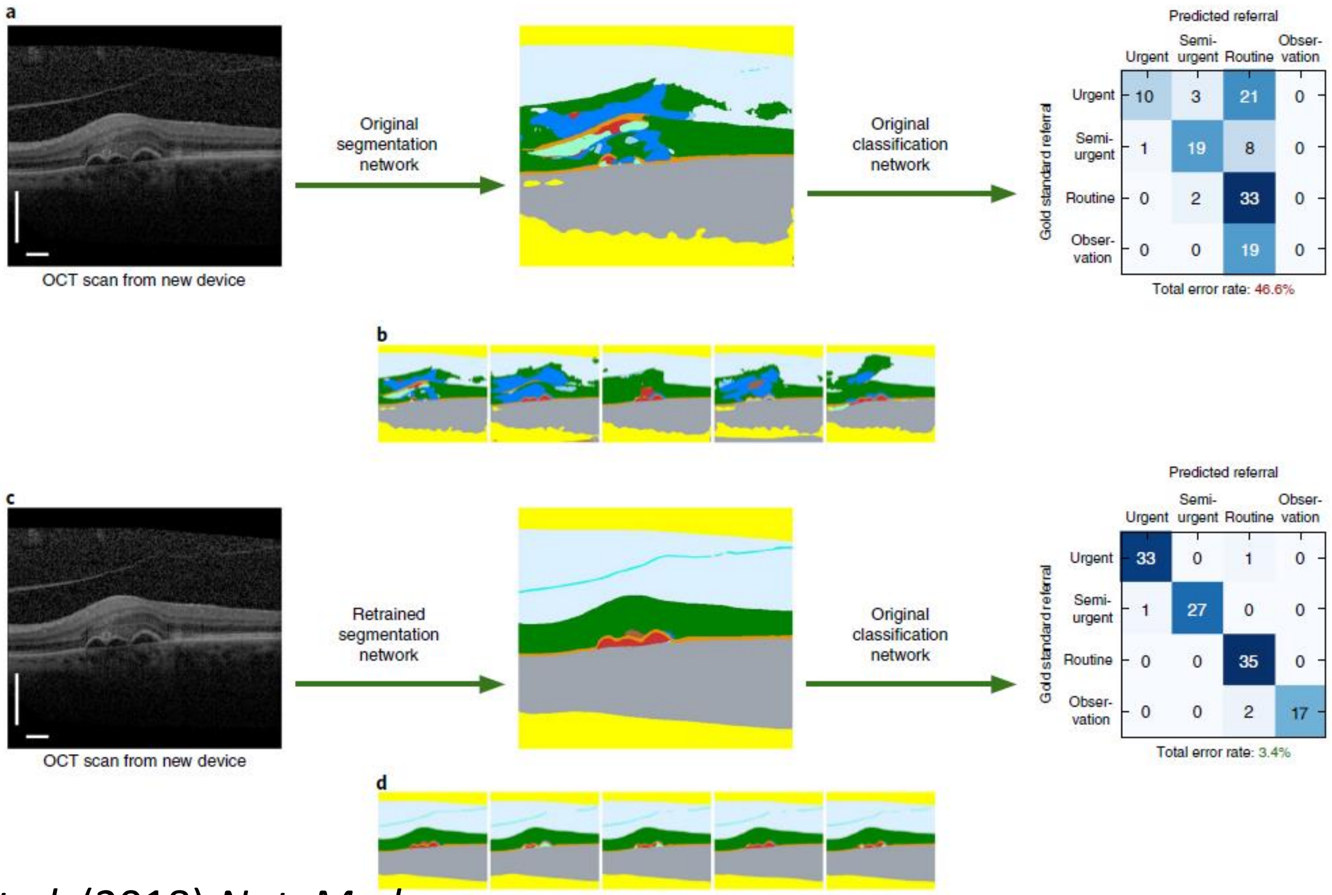


Fig. 3, Fauw *et al.* (2018) *Nat. Med.*

Result – generalization to a new device type

- High error rate likely due to erroneous segmentation



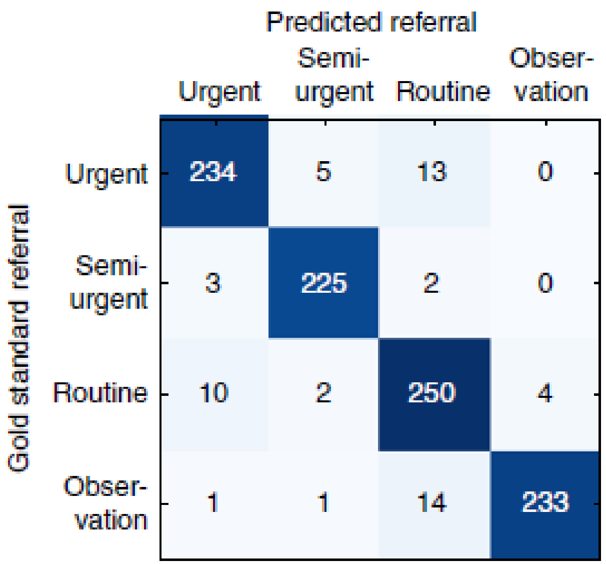
With original U-Net
for segmentation
Error rate: 46.6 %

With 2-branch U-Net
Error rate: 3.4 %

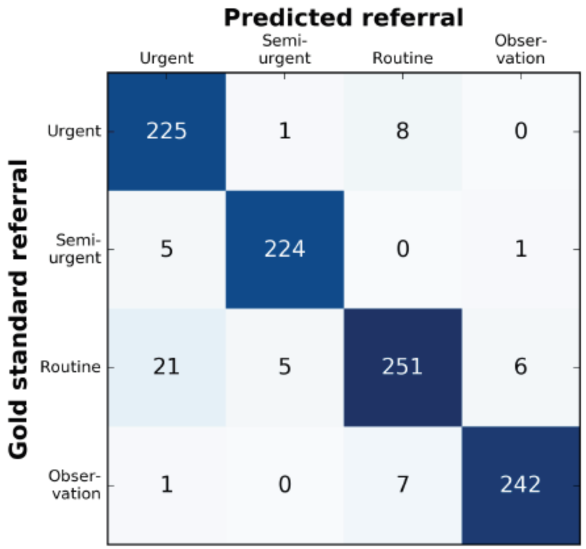
Fig. 4, Fauw *et al.* (2018) *Nat. Med.*

Result – 2-stage vs end-to-end approach

- The same classification network architecture trained on 14884 raw images instead of segmentation map
- Error rate: 5.5% both



2-stage approach



end-to-end approach

Note:
All images are from device type 1 !
Performance on device type 2 unknown.

Fig. 3 & Supp. Fig. 7 , Fauw *et al.* (2018) *Nat. Med.*

Summary

- New architecture
  - 2-stage approach
    - Reduced training data requirement
- Penalty customized to reflect consequence of error
- Generalizability
  - Wide coverage on pathologies
  - Images collected using 37 devices (2 types) at 32 different clinical sites
- The ensemble output had **performance comparable to top clinical expert**