







In the format provided by the authors and unedited.

Clinically applicable deep learning for diagnosis and referral in retinal disease

Jeffrey De Fauw¹, Joseph R. Ledsam¹, Bernardino Romera-Paredes¹, Stanislav Nikolov¹, Nenad Tomasev¹, Sam Blackwell¹, Harry Askham¹, Xavier Glorot¹, Brendan O'Donoghue¹, Daniel Visentin¹, George van den Driessche¹, Balaji Lakshminarayanan¹, Clemens Meyer¹, Faith Mackinder¹, Simon Bouton¹, Kareem Ayoub¹, Reena Chopra ², Dominic King¹, Alan Karthikesalingam¹, Cían O. Hughes ^{1,3}, Rosalind Raine³, Julian Hughes², Dawn A. Sim², Catherine Egan², Adnan Tufail², Hugh Montgomery ³, Demis Hassabis¹, Geraint Rees ³, Trevor Back¹, Peng T. Khaw², Mustafa Suleyman¹, Julien Cornebise^{1,3,4}, Pearse A. Keane ^{2,4*} and Olaf Ronneberger ^{1,4*}

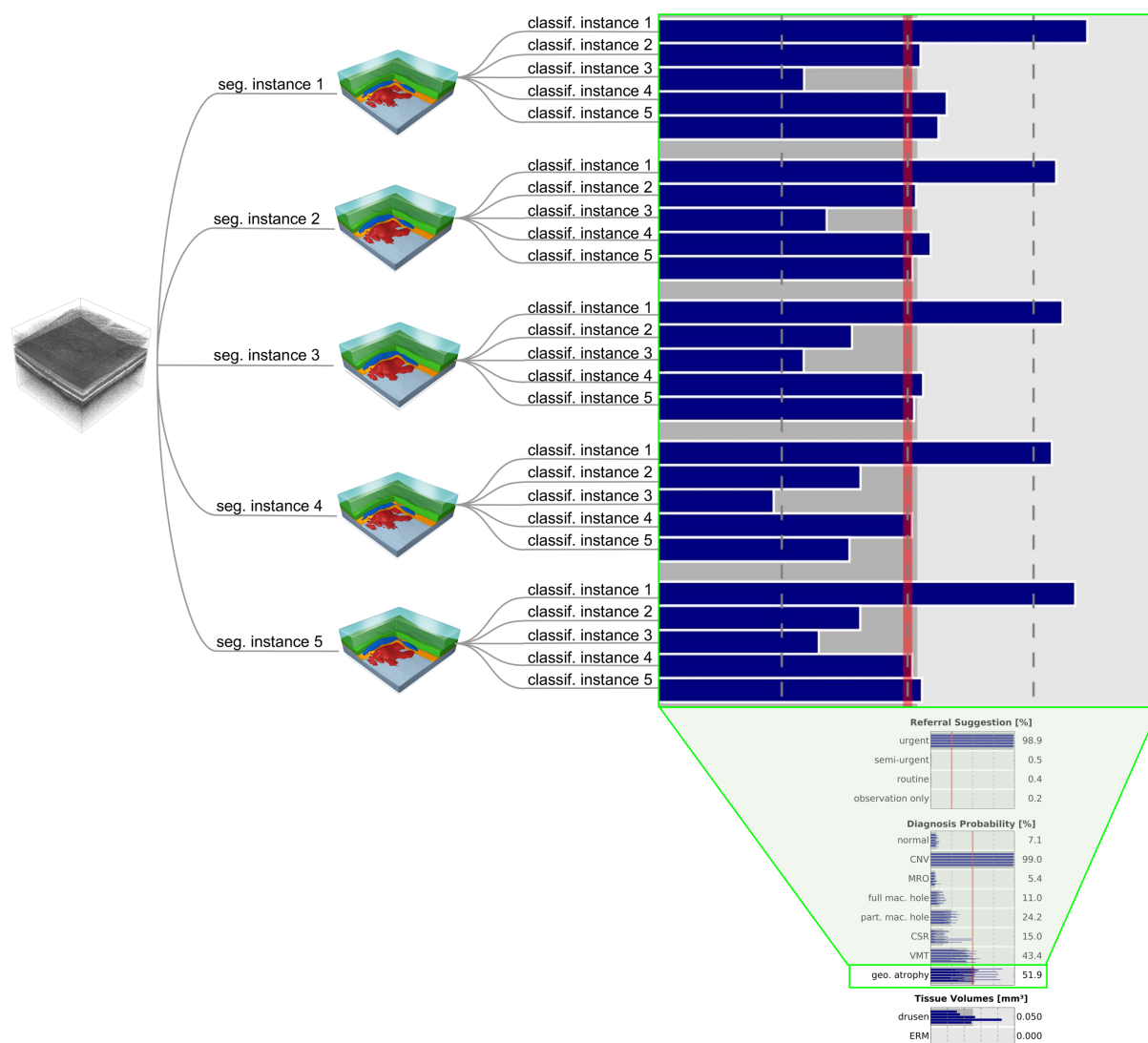
¹DeepMind, London, UK. ²NIHR Biomedical Research Centre at Moorfields Eye Hospital and UCL Institute of Ophthalmology, London, UK.

³Present address: University College London, London, UK. ⁴These authors contributed equally: Julien Cornebise, Pearse A. Keane, Olaf Ronneberger.

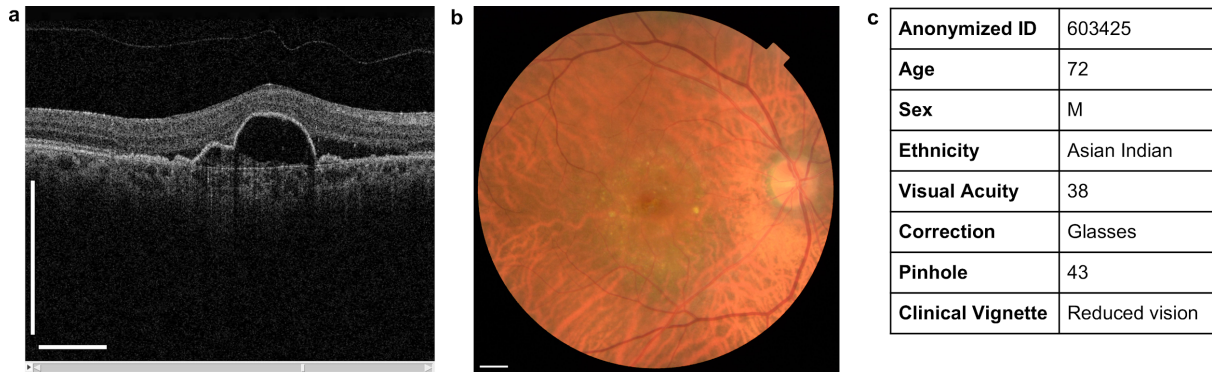
*e-mail: pearse.keane@ Moorfields.nhs.uk; olافر@deepmind.com

Supplementary Information for *Clinically applicable deep learning for diagnosis and referral in retinal disease*

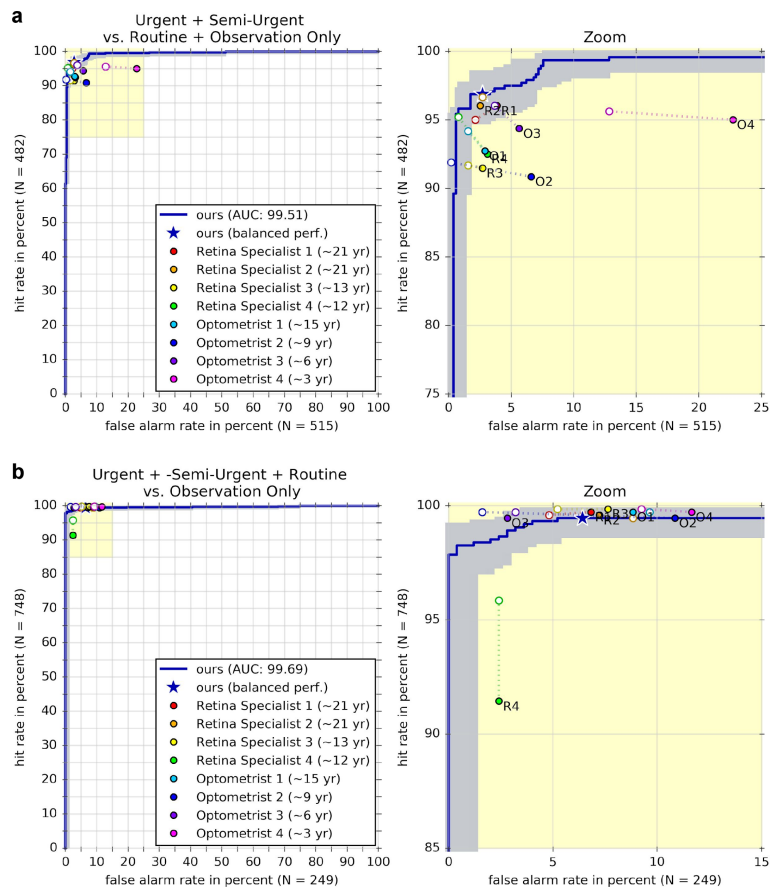
Supplementary Figures



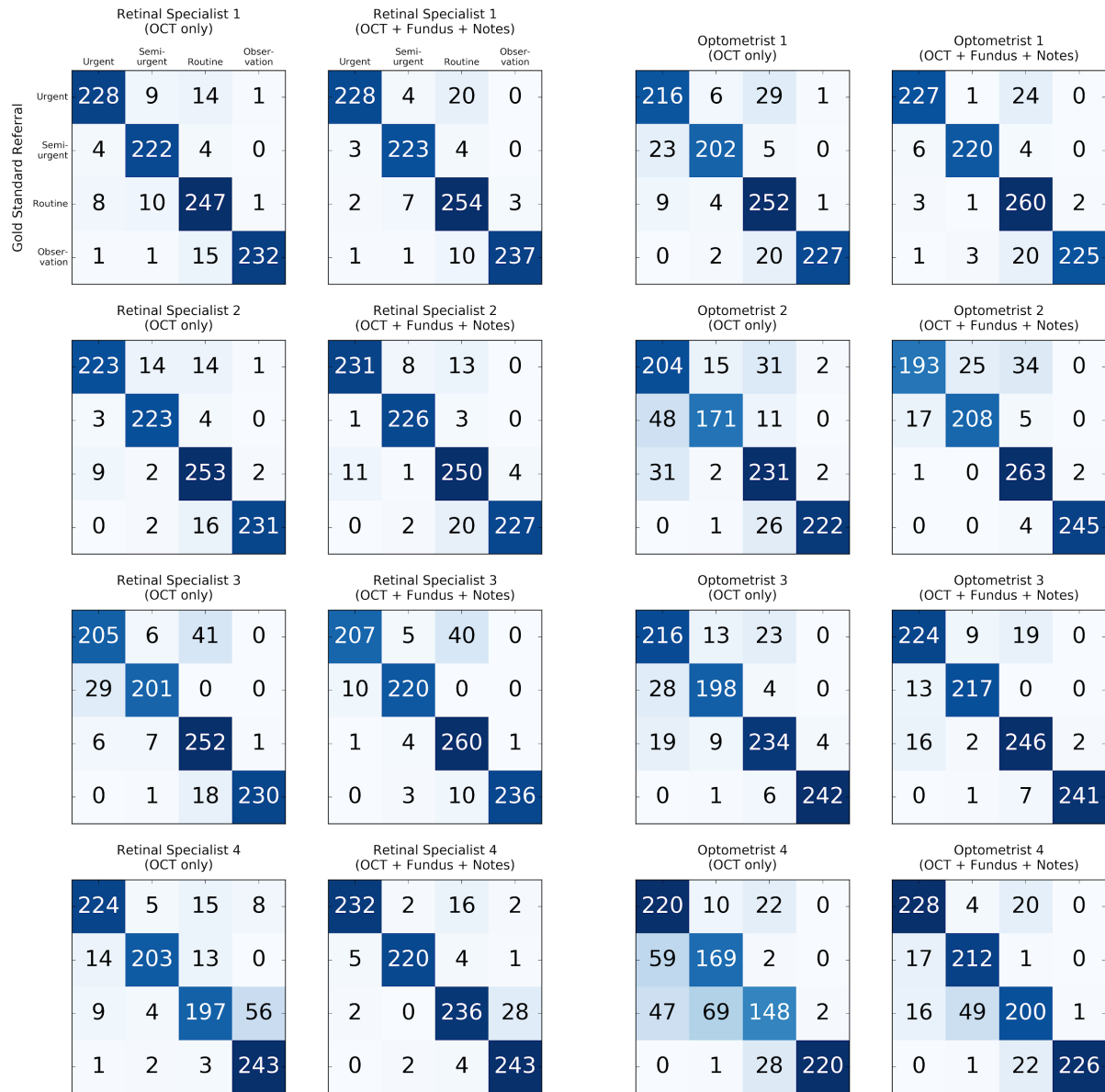
Supplementary Figure 1 | Generating predictions with an ensemble of segmentation and classification networks. Illustration showing how the ensemble of 5 segmentation network instances and 5 classification network instances are jointly used to generate 25 predictions for one scan. Each segmentation network instance first provides a segmentation map hypothesis based on the input OCT. For each of these 5 segmentation map hypotheses every classification network instance provides a probability for each label, here shown in detail for the geographic atrophy label.



Supplementary Figure 2 | Example information available to the experts to make the referral decision. (a) Full OCT scan with a slider to scroll through the slices. **(b)** Fundus image. **(c)** Patient summary notes. Scale bars: 1mm



Supplementary Figure 3 | Receiver operating characteristic (ROC) diagrams for referral decisions. (a) Urgent and semi-urgent referral versus routine referral and observation only (n=997 patients). The blue ROC curve is created by sweeping a threshold over the predicted probability (or the measured segmentation volume in case of Drusen and epiretinal membrane (ERM)). Points outside the light blue area correspond to a significantly different performance (95% confidence level, using a two-sided exact binomial test). Filled markers denote expert's performance using OCT only; outlined markers denote their performance using OCT, fundus image and summary notes. Dashed lines connect the two performance points of each expert. **(b)** Urgent, semi-urgent and routine referral versus observation only (n=997 patients)



Supplementary Figure 4 | Confusion matrices for the referral decision for all 8 experts. n=997 patients.

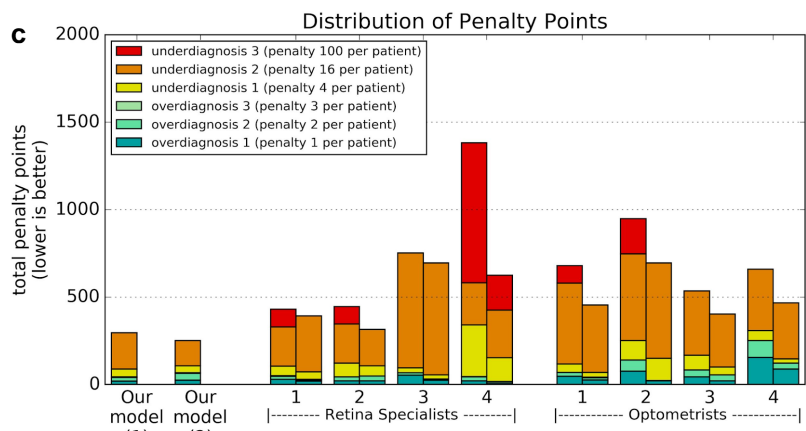
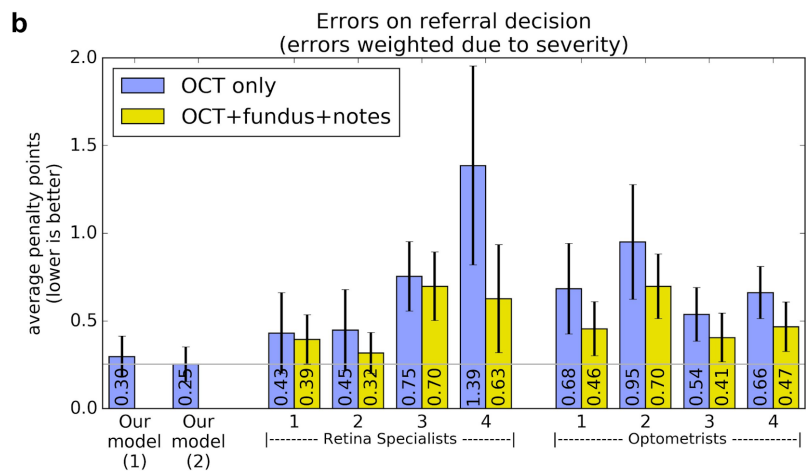
Error rate [percent]

		num. classif. model instances				
		1	2	3	4	5
num. segm. model instances	1	7.908	6.794	6.361	6.078	5.958
	2	7.48	6.403	5.975	5.727	5.537
	3	7.298	6.267	5.9	5.693	5.557
	4	7.186	6.166	5.87	5.697	5.697
	5	7.121	6.138	5.878	5.777	5.517

Supplementary Figure 5 | Error rate of the referral decision on device type 1. The error rate is shown for the device type 1 test set (dataset #5 in **Supplementary Table 3**) for different numbers of model instances in the ensemble. The error rate was computed as the average over all possible combinations of N x M segmentation and classification model instances out of the 5 x 5 instances that are used in the rest of this study. The performance differences between 4 x 4 instances and 5 x 5 instances are only marginal (n=997 patients)

a Penalty points for wrong decisions

Gold Standard Referral	Predicted Referral			
	Urgent	Semi-urgent	Routine	Observation
Urgent	0	4	16	100
Semi-urgent	1	0	4	16
Routine	2	1	0	4
Observation	3	2	1	0

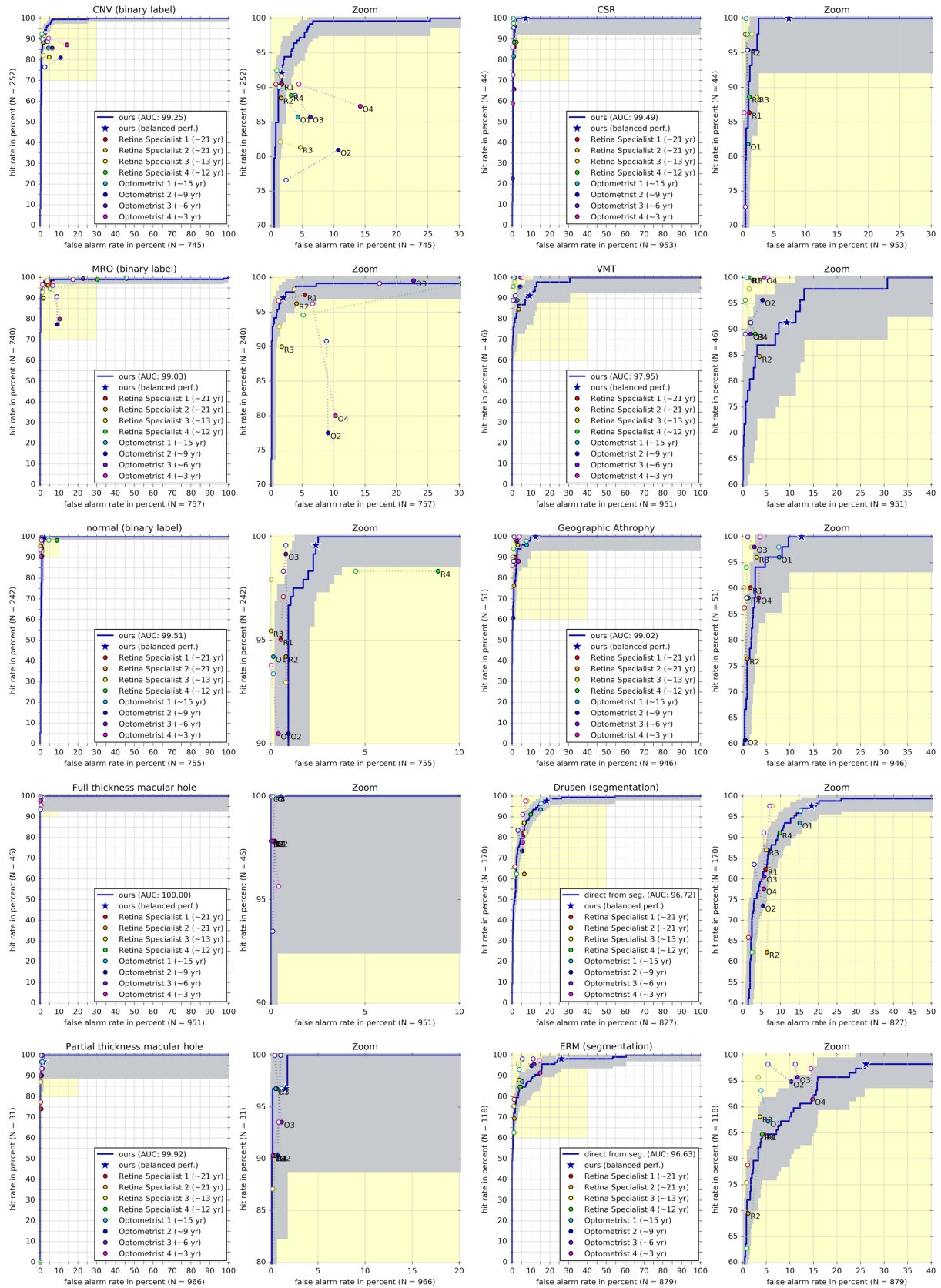


Supplementary Figure 6 | Adverse consequences of wrong referral decisions. (a) Our proposed penalty points for a wrong referral decision. In the first row the penalty points correspond approximately to the number of weeks that a CNV patient might lose in referral time before treatment (with 100 as a maximum for a patient triaged as ‘observation’ and would not be called back for assessment). The penalty points in the other rows are selected relative to this, with the additional constraint that an overdiagnosis (lower left triangle) is considered less harmful to an individual than an underdiagnosis. (b) Average penalty points per patient according to our proposed penalty metrics. Framework (1) is optimized for balanced performance; framework (2) is optimized for a better penalty score (n=997 patients, error bars indicate 95% confidence interval, computed from the standard error of the sample mean) (c) Distribution of the collected penalty points for our models and the experts (n=997 patients) in the same layout as above. The colored parts of each bar indicate the amount of penalty points collected in each category.

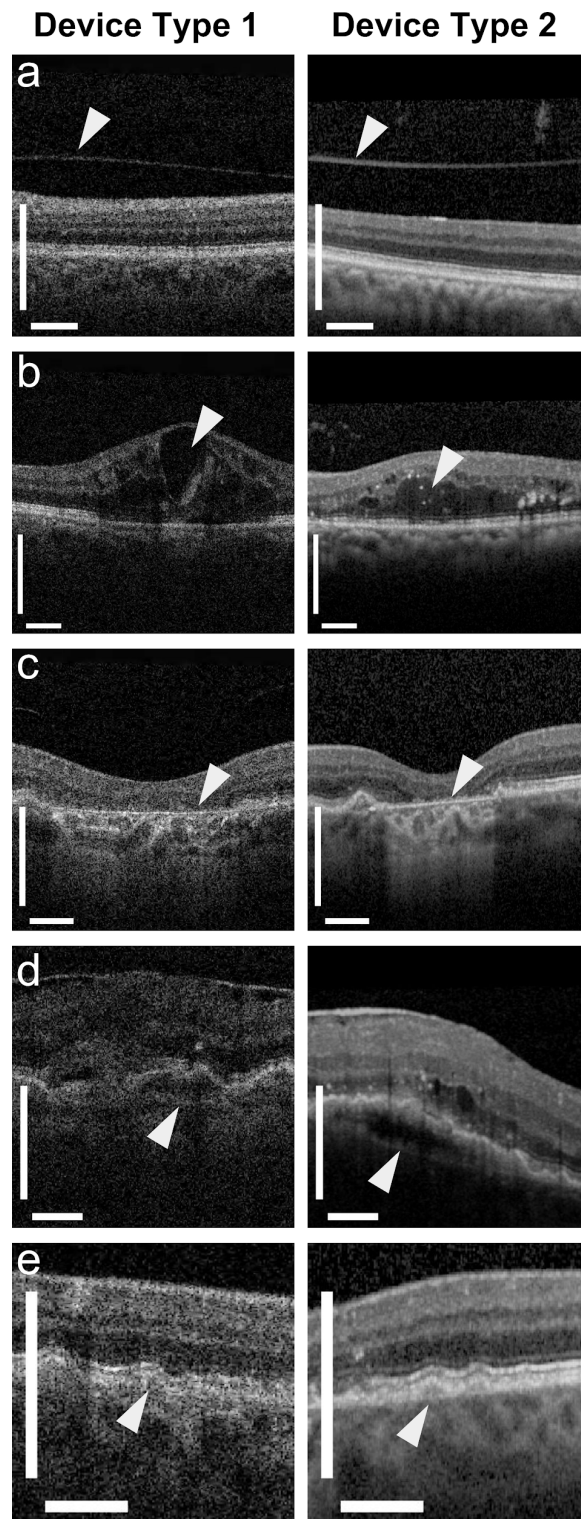
Confusion matrix

Gold standard referral	Predicted referral			
	Urgent	Semi-urgent	Routine	Observation
Urgent	225	1	8	0
Semi-urgent	5	224	0	1
Routine	21	5	251	6
Observation	1	0	7	242

Supplementary Figure 7 | Confusion matrix of an end-to-end classification network applied to the test set (dataset #5 in Supplementary Table 3). The results were obtained by an ensemble of 5 model instances (n=997 patients).

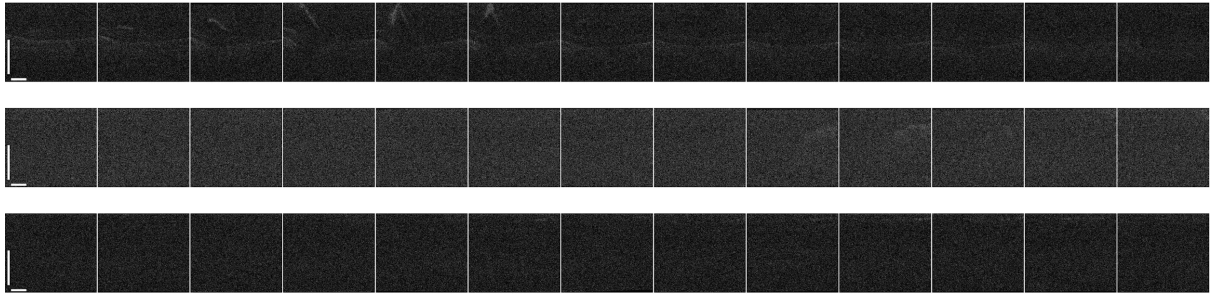


Supplementary Figure 8 | Receiver operating characteristic (ROC) diagrams for the additional pathologies. The blue ROC curve is created by sweeping a threshold over the predicted probability (or the measured segmentation volume in case of Drusen and ERM). $n=997$ patients. Points outside the light blue area correspond to a significantly different performance (95% confidence level, using a two-sided exact binomial test). Filled markers denote expert's performance using OCT only; outlined markers denote their performance using OCT, fundus image and summary notes. Dashed lines connect the two performance points of each expert.

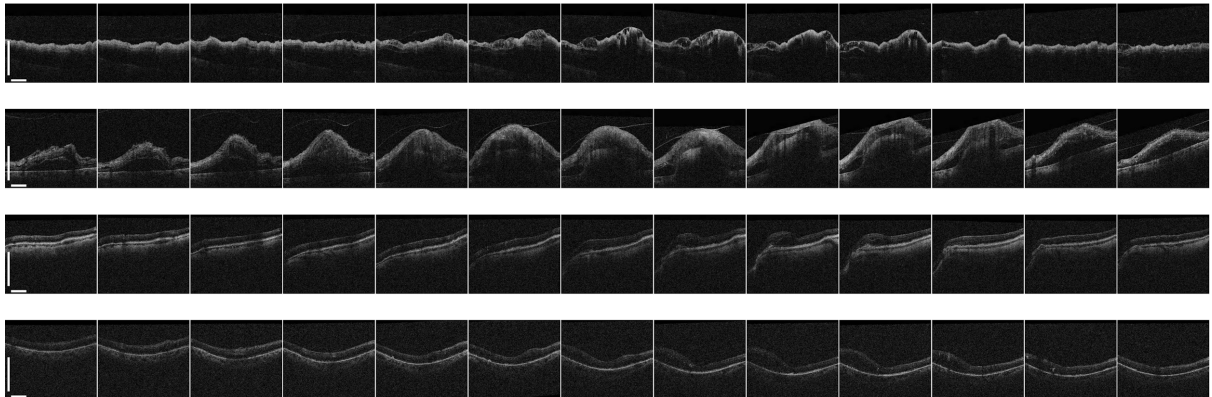


Supplementary Figure 9 | Differences between the two device types. The scans taken by device type 2 (Heidelberg Spectralis) have obvious differences in appearance compared to those taken by device type 1 (Topcon 3D OCT). The higher contrast in device type 2 results in better feature definition which could mislead segmentation models trained only on device type 1. Arrowheads in each image show the differences between device types 1 & 2 respectively. (a) Posterior hyaloid. (b) Intraretinal fluid. (c) Geographic atrophy. (d) Fibrovascular PED. (e) Drusen. In addition, in all images the choroid - the area below the retina at the bottom of the images - is better defined with device type 2. While this has benefits for diagnosis the differences can confuse models which may mistake the differences for additional retinal layers. Scale bars: 0.5mm

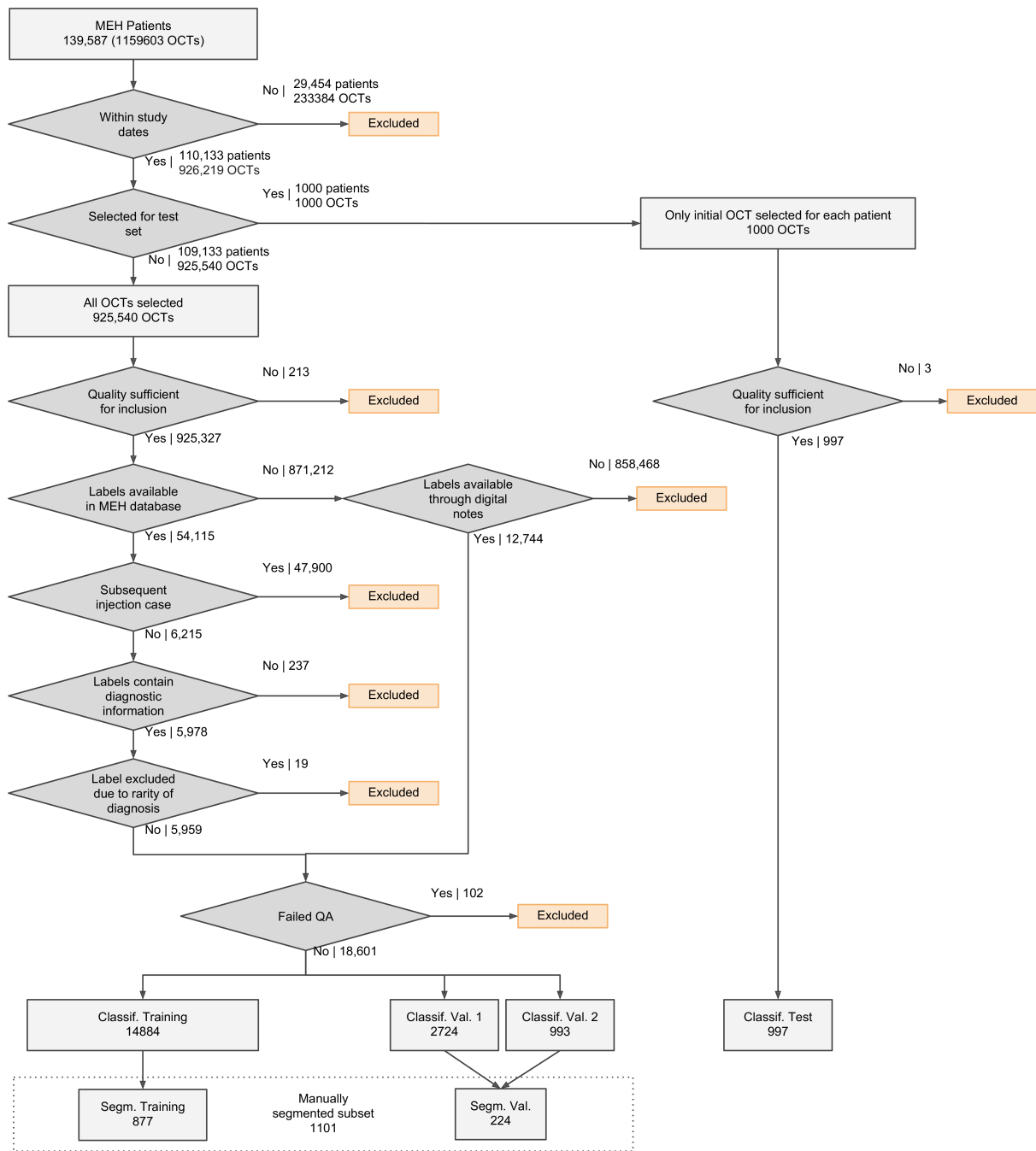
a Excluded from the test set



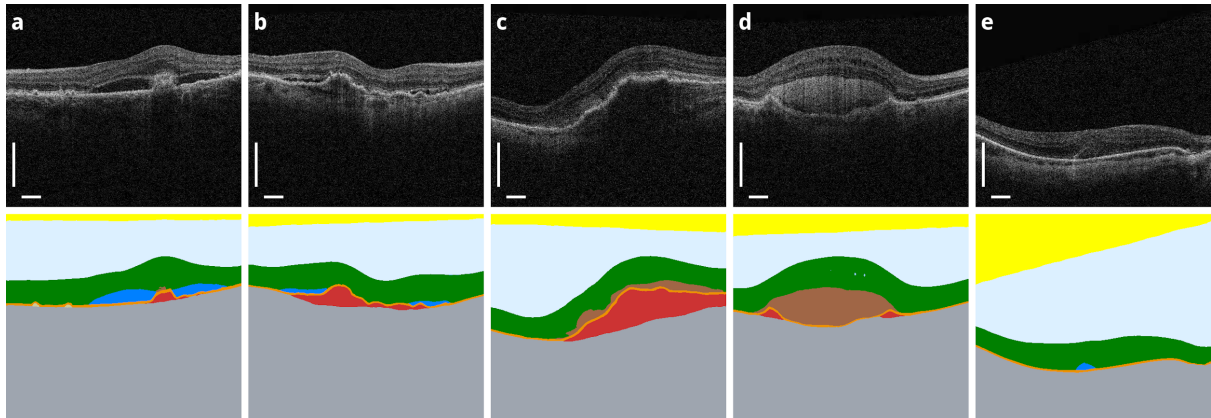
b Not Excluded from the test set



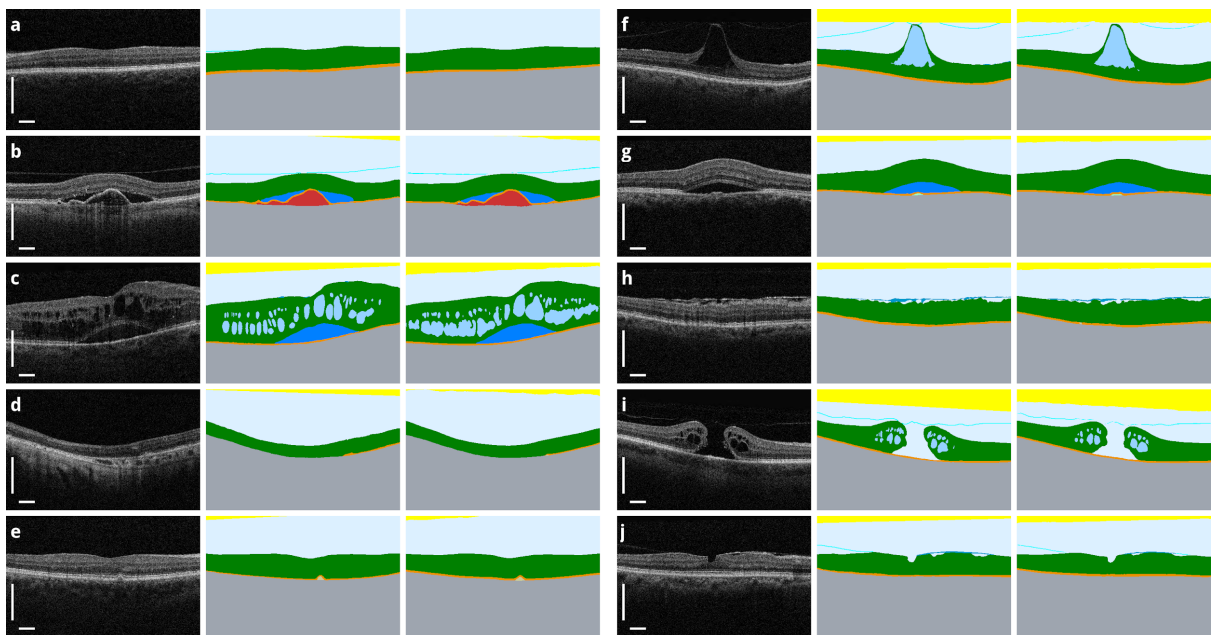
Supplementary Figure 10 | Excluded and included cases in the test set. (a) The three cases excluded from the test set due to insufficient signal in the OCT. Every 10th slice of the OCT scan is displayed. Note that in all cases the retina is either absent (empty scan) or barely visible, preventing the interpretation of the scan. (b) Four examples of cases that were included in the test set despite being of poor quality, or representing complex pathology. Scale bars: 1mm



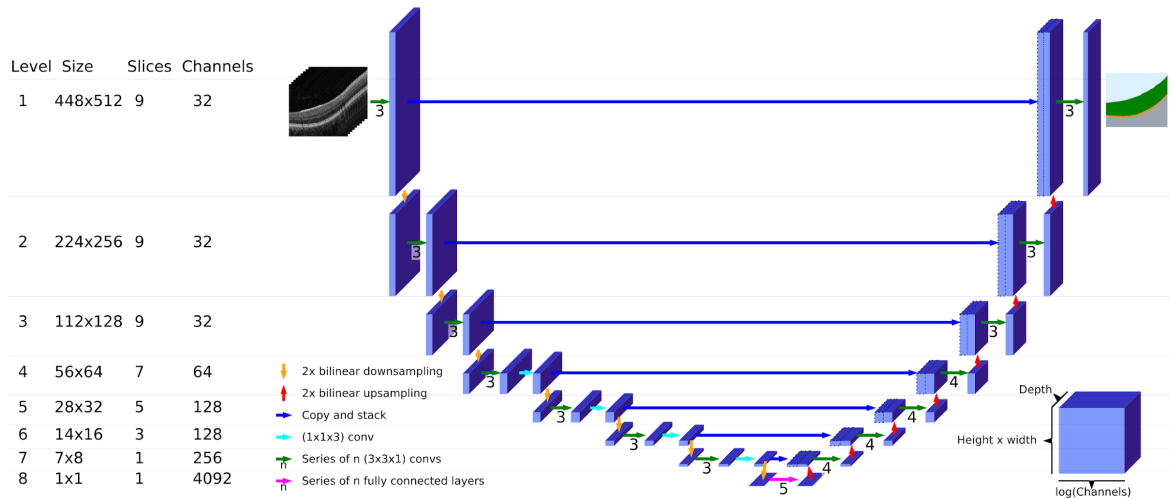
Supplementary Figure 11 | Sample selection at Moorfields Eye Hospital (MEH). Manual opt outs are not included as none of the patients who manually opted out had digital OCT within the study dates.



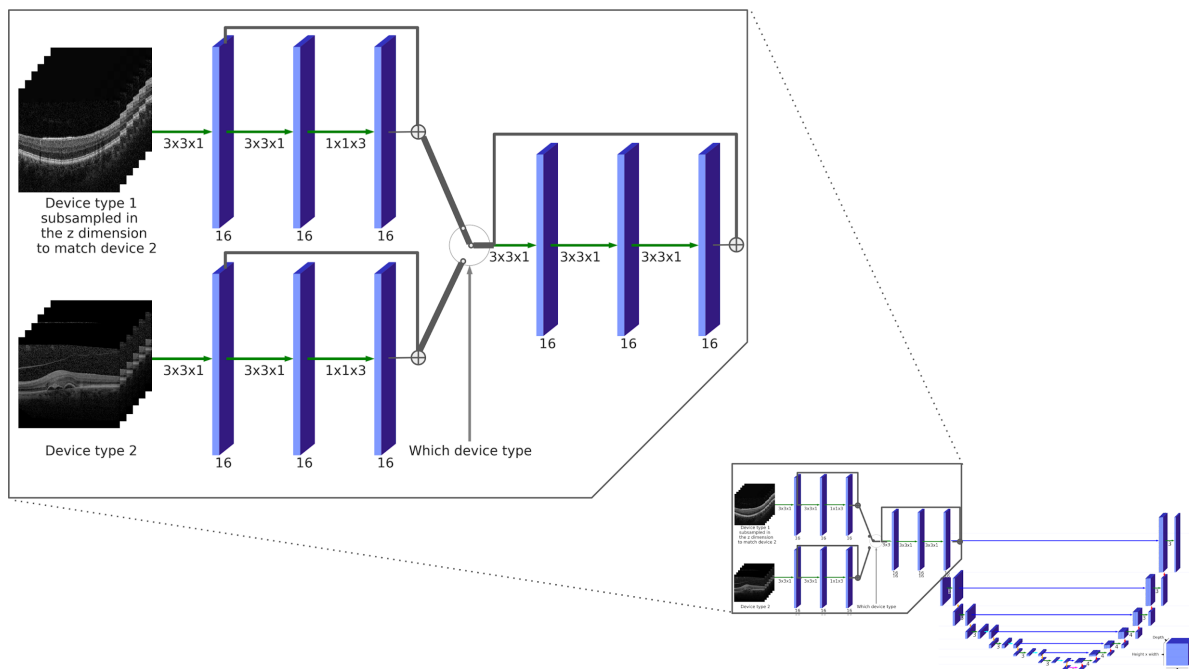
Supplementary Figure 12 | Five examples of patients in the test set with choroidal neovascularization (CNV). All images from the test set (n=997) show CNV requiring urgent referral with corresponding segmentations from our segmentation network (color legend in **Supplementary Table 2**). (a) A patient with choroidal neovascularization in the context of CSR. (b) A patient with choroidal neovascularization resulting from age related macular degeneration (AMD). (c) A patient with extensive fibrovascular pigment epithelium and subretinal hyperreflective material. (d) A patient with large amounts of subretinal hyperreflective material in the context of CNV. (e) A highly ambiguous case with a possible retinal angiomatous proliferation (RAP) lesion in a myopic patient. Scale bars: 0.5mm



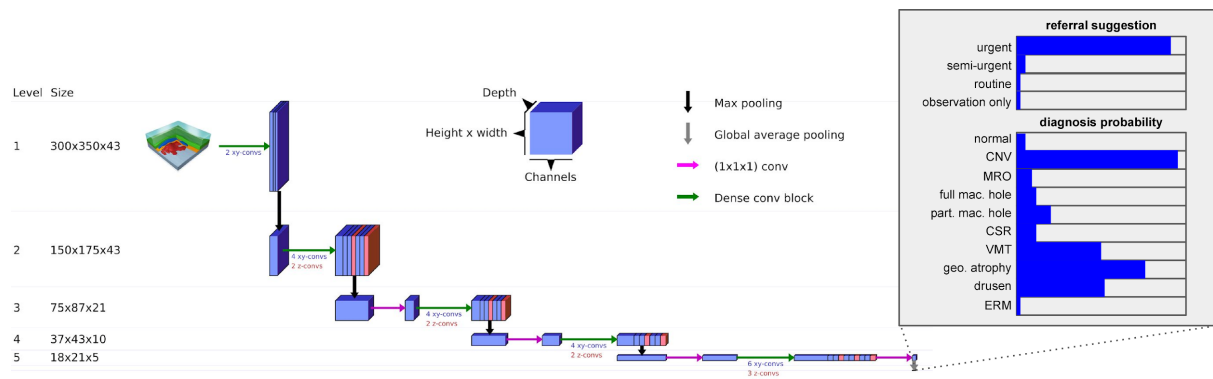
Supplementary Figure 13 | Examples of the segmentation model output for ten different retinal pathologies. 2D slices of OCT scans in the segmentation test set (n=224) with corresponding manual and predicted segmentation maps for the ten pathology classes included in our study (color legend in **Supplementary Table 2**). Classes are not mutually exclusive and multiple pathologies may be present in a single scan. (a) A normal retina as it appears in an OCT scan. (b) A patient with choroidal neovascularization due to age related macular degeneration. The segmentation map shows the area of fibrovascular pigment epithelium detachment associated with neovascularization. (c) Diabetic maculopathy and referable macular edema. (d) Geographic atrophy in late age related macular degeneration. (e) Drusen in early age related macular degeneration. (f) Vitreomacular traction. (g) Central serous retinopathy. Note that the segmentation model correctly identifies the pigment epithelium detachment as serous material. (h) A patient with epiretinal membrane. (i) Full thickness macular hole. (j) Partial thickness macular hole. Scale bars: 0.5mm



Supplementary Figure 14 | 3D U-Net model used in the first stage of our approach. At training time, the model receives 9 contiguous OCT slices. Blue boxes illustrate the 4D activation maps. Colored arrows stand for the different operations.



Supplementary Figure 15 | 2-branch U-Net for device type 2. The architecture of our segmentation network with "device adaptation branches" to segment scans of device type 2. In the top left we show an enlarged version of the differences compared to the original architecture for device type 1 (shown in **Supplementary Fig. 14**). Blue boxes illustrate the 4D activation maps with the number of channels shown below. Green arrows denote convolutional operations. We train on scans from both device type 1 and device type 2 but subsample those from device type 1 in the z-dimension to match the lower z-resolution of device type 2. Depending on which device the scan is from, the scan first goes through either the top branch, for device type 1, or the bottom branch, for device type 2. The output of the chosen branch is then used as input to a modified version of the first level of the analysis path of the original architecture. The rest of the architecture is identical.


















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.

Supplementary Tables

Supplementary Table 1 | Taxonomy of referral classes

Referral Category	Definition
Urgent	All causes of choroidal neovascularization, including age related macular degeneration, high myopia, central serous retinopathy, inherited retinal dystrophies (e.g., angioid streaks), posterior uveitis (e.g., multiple choroiditis), and post traumatic choroidal rupture.
Semi-urgent	Referable edema classed as semi-urgent included diabetic maculopathy, retinal vein occlusion, postoperative (Irvine-Gass syndrome), uveitis, Coat's disease, radiation and miscellaneous other cases.
Routine	All other non-urgent cases with a large variety, from uncomplicated central serous retinopathy to more rare conditions such as Macular Telangiectasia (MacTel) type 2.
Observation only	The absence of pathology classes described above.

Supplementary Table 2 | Taxonomy of segmentation regions

Color	Feature	Definition	Training		Test	
			Total number of scans with label segmented	Percent of total voxels	Total number of scans with label segmented	Percent of total voxels
	Vitreous and subhyaloid	Area above the internal limiting membrane not covered by other segmentation classes	856	20.63	220	20.30
	Posterior hyaloid	Hyperreflective membrane visible above the retina in cases of posterior vitreous detachment	356	0.12	95	0.13
	Epiretinal membrane	Hyperreflective band seen on the inner surface of the retina, often associated with distortion of the underlying neurosensory retina	326	0.06	95	0.04
	Neurosensory retina	All layers and contents of the retina, excluding the pathological features described below	856	10.76	220	11.12
	Intraretinal fluid	Areas of round or oval hyporefectivity located within the neurosensory retina	356	0.59	111	0.76
	Subretinal fluid	Hyporefective areas in the subretinal space - the space below the neurosensory retina but above the retinal pigment epithelium	255	0.48	68	0.33
	Subretinal hyper reflective material	Areas of hyperreflectivity between the retinal and RPE	92	0.12	22	0.08
	Retinal pigment epithelium (RPE)	Hyperreflective band underlying the neurosensory retina	853	0.92	220	0.95
	Drusenoid pigment epithelium detachment (PED)	Elevation of the RPE, often dome-shaped, with a hypo- or medium-reflective material separating the RPE from the underlying Bruch's membrane, and without the presence of fibrovascular material	268	0.07	61	0.06
	Serous PED	Dome-shaped elevation of the retinal pigment epithelium relative to Bruch's membrane, typically seen overlying a homogeneously hyporefective space devoid of fibrovascular material	58	0.02	11	0.003
	Fibrovascular PED	Irregular elevations of the retinal pigment epithelium relative to Bruch's membrane containing fibrovascular tissue of variable reflectivity	183	0.37	45	0.54
	Choroid and outer layers	Area below the RPE not covered by other segmentation classes	854	51.26	220	51.56
	Mirror artefact	Artefact caused by patient anatomy out of the OCT frame being reflected back onto the OCT	9	0.01	3	0.02
	Clipping artefact	Padding voxels introduced at the edges of OCT slices during image processing	877	7.21	224	7.36
	Blink artefact	Absent information due to patient blink	22	7.38	5	6.75

Supplementary Table 3 | Overview of datasets used for training, validation and testing of the different networks

	Dataset	Device type	Number of scans	Input	Labels	Label source
#1	Training set for segmentation	1	877	OCT scans	Sparse segm. maps (3-5 slices per scan)	Manually segmented by trained ophthalmologists, reviewed and edited by a senior ophthalmologist.
#2	Validation set for segmentation	1	224	OCT scans	Sparse segm. maps (3-5 slices per scan)	Manually segmented by trained ophthalmologists, reviewed and edited by a senior ophthalmologist.
#3	Training set for classification	1	14884	dense segm. maps, created automatically from OCT scans (5 segm. maps per scan)	Diagnoses and referral decision	Automated notes search + trained ophthalmologist and optometrist review of the OCT scans.
#4	Validation set for classification	1	993	dense segm. maps, created automatically from OCT scans	Diagnoses and referral decision	Graded by three junior graders. Disagreement in clinical labels arbitrated by a senior grader.
#5	Test set: Referral gold standard	1	997	OCT scans	Referral decision	Full patient clinical records to determine the final diagnosis and optimal referral pathway in the light of that (subsequently obtained) information.
#6	Test set: Diagnoses silver standard		(same OCT scans as #5)		Diagnoses	Majority vote from 8 experts (4 retinal specialists and 4 optometrists) grading using OCT scan, fundus image and clinical notes
#7	Human results: Experts on OCT only		(same OCT scans as #5)		Diagnoses and referral decision from 8 experts	8 experts (4 retinal specialists and 4 optometrists) grading on OCT scan only
#8	Human results: Experts on OCT + fundus + notes		(same OCT scans as #5)		Diagnoses and referral decision from 8 experts	8 experts (4 retinal specialists and 4 optometrists) grading on OCT scan, fundus image and clinical notes
#9	Training set 2 for segmentation	2	152	OCT scans	Sparse segm. maps (3-5 slices per scan)	Manually segmented by trained ophthalmologists, reviewed and edited by a senior ophthalmologist.
#10	Validation set 2 for classification	2	112	OCT scans	Referral decision	Full patient clinical records to determine the final diagnosis and optimal referral pathway in the light of that (subsequently obtained) information.
#11	Test set 2: Referral gold standard	2	116	OCT scans	Referral decision	Full patient clinical records to determine the final diagnosis and optimal referral pathway in the light of that (subsequently obtained) information.
#12	Human results: Experts on OCT + fundus + notes		(same OCT scans as #11)		Referral decision from 5 experts	5 retinal specialists grading on OCT scan, fundus image and clinical notes
#13	Training set for end-to-end model	(same cases as #3)		OCT scans		(same labels as #3)

Supplementary Table 4 | Overview of the OCT scan sizes used in this study. All sizes are given in A-scan, B-scan, C-scan direction

Dataset	image size [voxels]	real world voxel size [μm]	real world image size [mm]	comments
device type 1 raw OCT scans	885 · 512 · 128	2.6 · 11.7 · 47.2	2.3 · 6.0 · 6.0	
segmentation network input / output	448 · 512 · 128	5.2 · 11.7 · 47.2		device type 1 scans resampled in A-scan direction to 5.2 μm voxel size, and zero-padded to the next multiple of 64 (added 6 pixels)
classification network input	300 · 350 · 43	7.8 · 17.6 · 141.7		segmentation map resampled to 7.8 μm · 17.6 μm · 141.7 μm voxel size such that the full classification network fits into GPU memory
device type 2 raw OCT scans	496 · 512 · 49	3.9 · 11.3 · 120	1.93 · 5.79 · 5.88	
2-branch segmentation network input / output	448 · 512 · 49	5.2 · 11.7 · 120		device type 2 scans resampled in A,B-scan direction to 5.2 μm · 11.7 μm voxel size, and padded accordingly; device type 1 scans resampled in C-scan direction to 120 μm voxel size.

Supplementary Table 5 | Taxonomy of diagnostic labels

Condition	Definition
Normal	Absence of pathology.
Macular retinal edema (MRE)	Referable retinal edema, seen in the OCTs as intraretinal and subretinal fluid.
Choroidal neovascularization (CNV)	New vessel growth from the choroidal layer of the eye; associated with a variety of retinal conditions including neovascular age related macular degeneration, severe myopia and central serous retinopathy.
Drusen	Acellular polymorphous deposits in Bruch's membrane; the most common early sign of dry age-related macular degeneration.
Geographic atrophy	Loss of the retinal pigment epithelium with variable loss of the overlying photoreceptors and underlying choriocapillaris; a sign of late stage dry age-related macular degeneration.
Central serous retinopathy (CSR)	A disease where increased choroidal permeability leads to a build up of subretinal fluid, causing a detachment of the neurosensory retina.
Full thickness macular hole	A round, full-thickness defect of retinal tissue in the foveal retina, leading to loss of central vision.
Partial thickness macular hole	A partial thickness defect of retinal tissue in the foveal retina.
Vitreomacular traction (VMT)	A disorder of the vitreoretinal interface where an incomplete posterior vitreous detachment exerts tractional pull on the macula and results in morphologic alterations and consequent metamorphopsia or central visual loss.
Epiretinal membrane (ERM)	Fibrocellular tissue found on the inner surface of the retina which may be idiopathic or secondary to various retinal conditions. Small epiretinal membranes may not be clinically significant, and may be considered a normal aging feature.

Supplementary Table 6 | Same as table before but experts have access to OCT + fundus image + full summary notes.

Diagnosis	Area under ROC curve [percent]	N positive samples	Experts with significantly higher performance	Experts with indistinguishable performance	Experts with significantly lower performance
CNV	99.25	252	-	●●●▲▲	●▲▲
MRE	99.03	240	-	●●●▲▲	●▲▲
normal	99.51	242	●●▲▲▲	●●	▲
full mac. hole	100.0	46	-	●●●●▲▲▲▲	-
part.mac. hole	99.92	31	-	●●●●▲▲▲▲	-
CSR	99.49	44	●●●▲▲	●▲▲	-
VMT	97.95	46	●●●▲▲▲	●▲	-
geographic atrophy	99.02	51	●●●▲▲▲	●▲	-
Drusen	97.42 (from segm.)	170	●●▲▲	●●▲▲	-
ERM	96.63 (from segm.)	118	●●▲▲▲	●●▲	-

Supplementary Table 7 | Performance on additional diagnoses of experts using OCT only to our framework. The ground truth for CNV is derived from the full follow-up patient files. The ground truth for the other diagnoses is computed as majority vote from the 8 experts. Drusen and ERM predictions were derived directly from the segmentation map. Circles represent retinal specialists, triangles represent optometrists.

Diagnosis	Area under ROC curve [percent]	N positive samples	Experts with significantly higher performance	Experts with indistinguishable performance	Experts with significantly lower performance
CNV	99.25	252	-	●●	●●▲▲▲▲▲
MRE	99.03	240	-	●●▲▲▲▲	●●▲
normal	99.51	242	●▲▲	●●●▲	▲
full mac. hole	100.0	46	-	●●●●▲▲▲▲▲	-
part. mac. hole	99.92	31	-	●●●▲▲▲▲▲	●
CSR	99.49	44	-	●●●●▲▲▲▲▲	-
VMT	97.95	46	●●●▲	●▲▲▲	-
geographic atrophy	99.02	51	▲	●●●●▲▲▲▲	-
Drusen	97.42 (from segm.)	170	-	●●●▲▲▲▲▲	●
ERM	96.63 (from segm.)	118	-	●●●●▲▲▲▲▲	-

Supplementary Table 8 | Total OCT examinations (unique patients in brackets) in the dataset by triage category.

Triage category	Training	Validation	Test (Device Type 1)	Test (Device Type 2)
Urgent	4832 (3039)	251 (237)	252 (252)	34 (34)
Semi-urgent	3438 (1854)	268 (259)	230 (230)	28 (28)
Routine	5223 (1927)	247 (236)	266 (266)	35 (35)
Observation only	1391 (801)	227 (195)	249 (249)	19 (19)
Total	14884 (7621)	993 (927)	997 (997)	116 (116)

Supplementary Table 9 | Number of cases of referral classes in the training and validation set for the segmentation network for OCT device type 1.

Referral Category	Number in Training Set	Number in Validation Set
Urgent	227	58
Semi-Urgent	182	57
Routine	89	20
Observation	379	89
Total	877	224

Supplementary Table 10 | The experience and position of the nine experts against which the algorithm was compared.

Expert	Position	Years of Experience
1	Consultant Ophthalmologist in Medical Retina	21
2	Consultant Ophthalmologist in Medical Retina	21
3	Consultant Ophthalmologist in Medical Retina	12.5
4	Consultant Ophthalmologist in Medical Retina	11.5
5	Specialist Optometrist, Medical Retina	15
6	Specialist Optometrist, Medical Retina	9
7	Specialist Optometrist, Medical Retina	6
8	Specialist Optometrist, Medical Retina	2.5
9	Consultant Ophthalmologist in Medical Retina	10

Supplementary Videos

Supplementary Video 1 - OCT viewer | This video demonstrates the interaction with the OCT viewer. The OCT scan belongs to a 72 year old female presented with increasing visual distortion over a 4 month period; the OCT shows loss of RPE consistent with geographic atrophy. The view first goes through the whole volume (128 slices) for a fixed tissue map hypothesis, followed by showing the different tissue map hypotheses for a given slice. Finally, we let the collage cycle through the different hypotheses continually while scrolling through the volume, pausing on several slices briefly to show the variations. The color legend for all segmentation maps is available in **Supplementary Table 2**.

Supplementary Video 2 - wet AMD | Choroidal neovascularization (CNV) is the pathognomonic feature of the neovascular ("wet") form of age-related macular degeneration (AMD) and requires urgent treatment to prevent irreversible visual loss. A 72-year old man presented with a history of reduced vision in his left eye. Best corrected visual acuity in the affected eye was 38 Early Treatment Diabetic Retinopathy Study (ETDRS) letters. The model correctly selects the Most Urgent Diagnosis as "CNV", suggesting referral to an ophthalmologist on an urgent basis. The model segmentation highlights growth of the neovascular tissue in the sub-retinal pigment epithelium (RPE) space – a so-called fibrovascular pigment epithelium detachment (PED). Subretinal fluid can be seen surrounding the inferior margins of the fibrovascular PED indicating the presence of ongoing CNV leakage.

Supplementary Video 3 - Normal | Scans are quick and safe to perform and are thus commonly used in the screening of patients without visual symptoms or other ophthalmic findings. A 46-year old man who was referred for retinal specialist review. Best corrected visual acuity was 6/6. The model correctly selects the referral decision as "Observation Only", suggesting that the OCT findings in isolation do not require referral to an ophthalmologist. The model accurately delineates the neurosensory retina without the presence of any pathologic compartments. It also highlights partial separation of the posterior hyaloid of the vitreous – this is a normal finding as the vitreous gel increasingly liquefies with age.

Supplementary Video 4 - Diabetic Macular Edema | Accumulation of this fluid in the macula – diabetic macular edema (DME) – is the commonest cause of visual impairment in diabetes. A 54-year old man with diabetes was referred to Moorfields for ophthalmologist review with best corrected visual acuity in the affected eye of 45 ETDRS letters. The model correctly detects the presence of macular retinal edema (MRE) and suggests semi-urgent ophthalmology referral. The model highlights intraretinal fluid accumulation, with cystoid spaces in both the inner nuclear and outer plexiform layers, and a mixed petaloid/honeycomb appearance on the en face images. There is also an accompanying significant increase in total retinal thickness.

Supplementary Video 5 - Ambiguous Case (chronic CSR) | In chronic CSR, diagnosis of secondary CNV formation is often challenging due to the frequent presence of shallow irregular pigment epithelium detachments (PEDs). A 60-64 year old woman presented with a history of CSR in her left eye. The model correctly detects the presence of CSR but is far less certain about the presence of CNV. It highlights a gravitational tract of subretinal fluid with a discrete area of fibrovascular PED superior to the fovea.

Supplementary Video 6 - Ambiguous Case (advanced geographic atrophy) | In advanced forms of AMD, geographic atrophy (GA) may sometimes coexist with CNV formation. In such cases, the CNV component may be clinically silent, and the fundus appearance may be limited to that of GA, making the diagnosis difficult. A 84-year old man was referred to Moorfields. Best corrected visual acuity in the affected eye was 1/60. The ground truth diagnosis was GA and routine referral was recommended. While the model correctly diagnoses the presence of GA and drusen, it suggests urgent referral due to the possible presence of CNV. The presence of subretinal hyperreflective on model segmentation is suggestive of previous CNV formation.

Supplementary Video 7 - Difficult Case of CNV | A 30 year old male patient, with a known history of CSR, presented with acute visual loss in his left eye and was diagnosed with secondary CNV formation. At this visit, the OCT scans lack many of the prototypical features of CSR, such as subretinal fluid accumulation. The model correctly diagnoses the presence of CNV and suggests the presence of CSR, but with far less certainty.

Supplementary Video 8 - Failure case (partial thickness macular hole) | Ocular media opacities may sometimes cause artefactual reductions in OCT signal strength and this can make accurate image segmentation challenging. Due to localized reduction in OCT signal strength in this case, some of the models erroneously detect the presence of a partial thickness macular hole. As a result, the models are uncertain as to whether the eye is normal or whether routine referral is required.

Supplementary Video 9 - Integration with other clinical information | Retinal angiomatous proliferation (RAP) is a variant of choroidal neovascularization (CNV) due to age-related macular degeneration (AMD). A 75-79 year old woman presented with reduced vision in her left eye. The model segmentation highlights the presence of a fibrovascular pigment epithelium detachment (PED) with subretinal hyperreflective material, overlying intraretinal fluid, and surrounding drusen. These findings are highly suggestive of RAP - in its early stages, this can be misdiagnosed as macular retinal edema (MRE), particularly in elderly patients with diabetes. The interpretable representation reduces the risk of misdiagnosis and allows the clinician to easily correlate these findings with other clinical information, e.g., fundus fluorescein angiography.