
**eMethods 1.** Image preprocessing and segmentation

**eMethods 2.** Training procedure

**eMethods 3.** t-Distributed stochastic neighbor embedding visualization

**eReferences.**

This supplementary material has been provided by the authors to give readers additional information about their work.
eMethods 1. Image preprocessing and segmentation

Images were resized to a common resolution of 640 x 480 pixels, and the retinal vessels segmented using a U-Net architecture, developed as an open-source project on GitHub (https://github.com/orobix/retina-unet). The network was trained on 500,000 image patches (48 x 48 pixels) extracted from 200 retinal images, with manually labelled vessel images serving as the ground truth. Methods for manual segmentation were previously published. The trained U-Net was used to segment the images in full by extracting overlapping patches and averaging the model predictions to produce a “vesselnness” probability map. Example segmentations are shown in eFigure 1.

![Example segmentations](image_url)

eFigure 1. Examples of automated vessel segmentation using a U-Net versus manual segmentation. Examples of original (left), manually segmented (middle) and automatically segmented (right) images. The automated segmentation is a probability map, where the brightness of each pixel corresponds to the probability assigned by the network of belonging to a vessel.
The vesselness images were automatically masked by converting the original RGB images to grayscale and applying a binary threshold. This crude mask was used to estimate the centroid and radius of the retina, from which a circular mask of the appropriate size was generated. Images were resized to 302 x 226 (retaining the original aspect ratio) and centrally cropped to 224 x 224 pixels. To account for the significant class imbalance between normal (85%), pre-plus (12%) and plus (3%) disease images, images in the training sets were augmented eight-fold by means of 90 degree rotations and horizontal and vertical flips. The augmented datasets were then randomly sampled to achieve equal numbers of images from each class.

eFigure 2. Binary masking procedure. An initial mask of the retina region is produced by simple thresholding of the raw RGB image (top left), from which an estimate of its radius is used to generate a circular mask (bottom left). The raw output of the U-net often includes incorrectly labelled regions at the periphery (top right) which are removed using the circular mask (bottom right).
eMethods 2. Training procedure

The Inception V1 (GoogLeNet) neural network architecture was used to classify individual retinal images as normal, pre-plus, or plus.³ Adopting a strategy known as transfer learning,⁴ the network was pre-trained on the ILSVRC ImageNet dataset comprising 1.2 million images of objects belonging to 1,000 classes.⁵ Preliminary experiments showed that despite having grayscale images as input, pre-training on color images from the ImageNet database afforded better performance than training from scratch. The softmax output layer was modified to perform three class prediction, and all network layers were subsequently fine-tuned. The cross-entropy loss function was minimized by stochastic gradient descent (SGD) for 100 epochs, with a constant learning rate of 0.0001. A dropout rate of 0.4 was also used to mitigate overfitting.
eMethods 3. t-Distributed stochastic neighbor embedding visualization

The Inception-v1 architecture comprises a series of convolutional layers in “blocks” before culminating in an average pooling layer and softmax classifier. For t-SNE analysis, the flattened output of the average pooling layer is a high-dimensional vector $x \in \mathbb{R}^{1024}$. Feature vectors were extracted for every image in the training set, and reduced to 50 dimensions using principal components analysis (PCA). t-SNE was then used to reduce the data to two dimensions, with a perplexity value of 30.
eReferences.


