

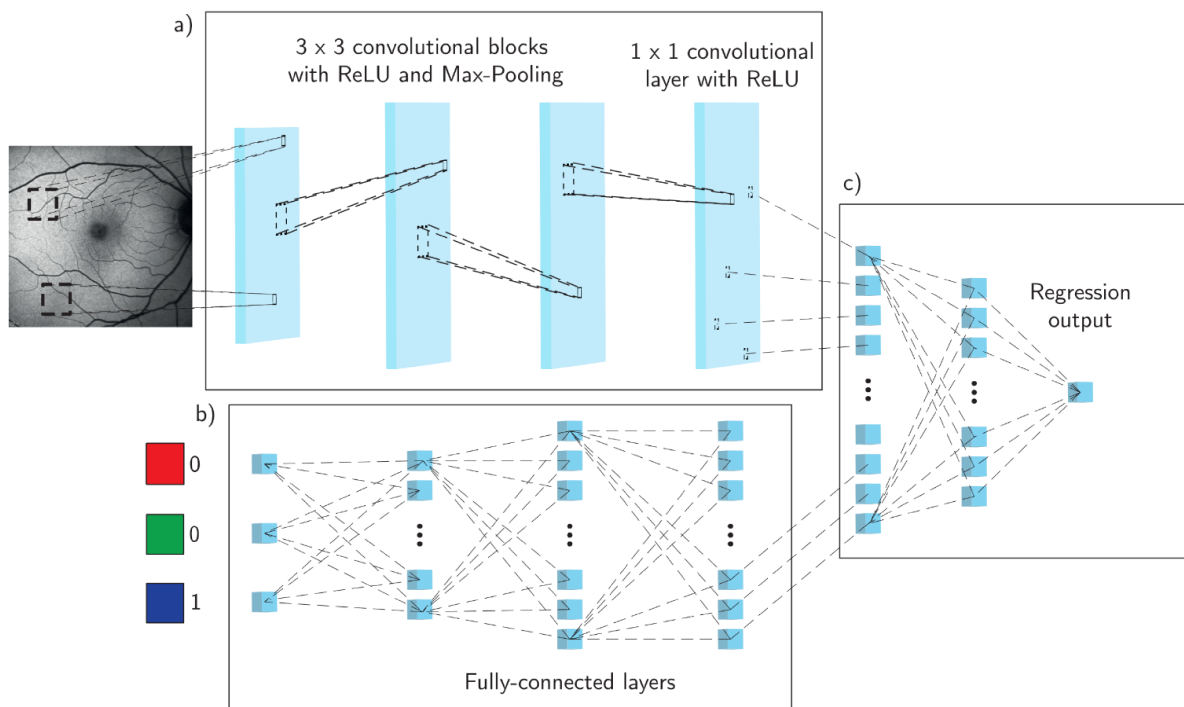
1 **Supplementary Figure 1: Neural network architecture.**

2 The convolutional neural network architecture used to obtain grading predictions is shown.
3 The processing of a single-channel greyscale AF image is shown in **A**. The retinal images
4 are centrally cropped to dimensions of 680 x 760, and transformed by applying three blocks
5 of 3 x 3 convolutions followed by ReLU activation and Max-Pooling (with 32, 64, and 64
6 filters, respectively). This is followed by a layer of 1x1 convolutions with 64 filters, and by
7 flattening to obtain a latent space representation of the input image. **B** represents the
8 processing of the one-hot-encoded imaging modality (0 = not this image modality/ 1 =this
9 image modality; in the below depicted example a BAF image is analysed) information. The
10 red, green and blue square represents the BAF, GAF or IR image modality. It undergoes
11 transformation by three fully-connected layers with Swish activation and 16, 32 and 32
12 hidden units (these numbers represent the number of neurons in each of the three fully-
13 connected (dense) layers; shown as cubes) each. The resulting representations are
14 concatenated and passed into a fully-connected regression head, shown in **C**, with the first
15 layer having 64 neurons, the second 32 neurons, and the final layer having just 1 neuron,
16 and using Swish activation. For training, the mean-squared error (MSE) loss function was
17 chosen, and minimised using the Adam optimiser with an initial learning rate of 10^{-4} , in
18 batched optimisation with a batch size of 512. In order to improve convergence, we used a
19 learning rate decay scheduler, reducing the learning rate by a factor of $\sqrt{2}$ whenever the
20 training loss did not improve by at least 10^{-3} over a patience period of 10 epochs.

21

22

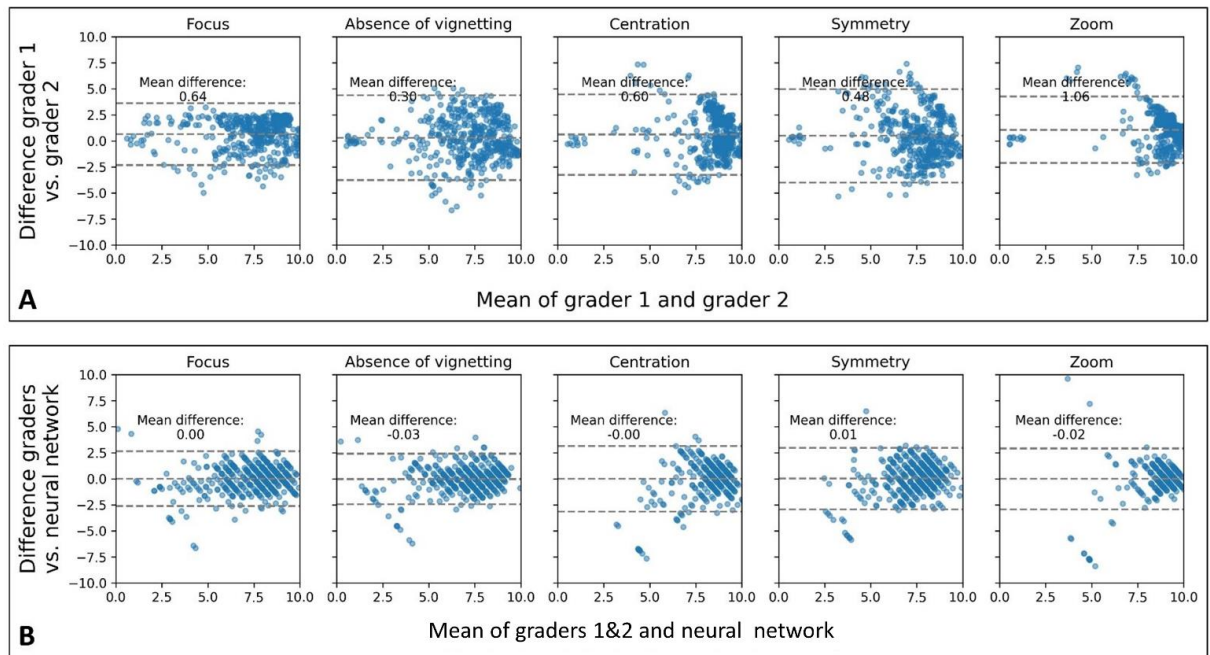
23



24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 **Supplementary Figure 2: Bland-Altman analysis.**

44 To visualize the agreement between the two human graders (A), as well as between the
45 human and neural network grading (B), we performed Bland-Altman analysis.³⁷ Random jitter
46 was added for better readability due to a discrete response scale.



47

48

49