Results Appendix. Validation of individual detection of referable DR in Messidor-2.

As previous step to generate this dataset, diabetic patients from Brest University Hospital were recruited and new macula-centered paired color fundus (CF) images were obtained without pharmacological dilation, using a Topcon TRC NW6 non-mydriatic fundus camera with a 45-degree field of view. This set of images is known as Messidor-Extension (LaTIM 2011) and contains 690 images from 345 examinations, which were combined with the images from the original Messidor set (ADCIS, Messidor dataset) that came in pairs (one image per eye), that is, 1058 images from 529 examinations. The final set is known as Messidor-2 (LaTIM 2011) and includes in total 1748 images from 874 examinations.

We used as reference standard (RS) the gradings made publicly available by Abràmoff et al. (University of Iowa Health Care; Abràmoff et al. 2013). Three US-board retinal specialists graded each image assigning a level from the International Clinical Diabetic Retinopathy (ICDR) severity scale (Wilkinson et al. 2003), obtaining then a consensus between the specialists. The gradings were provided at participant level and regarding referability or non-referability, i.e., a person was classified as a non-referable case if both eyes were classified as level 0 or 1, whereas a participant was considered to have referable DR if one or both eyes belonged to higher stages. In total, 684 subjects (78%) were graded as referable DR and 190 (22%) were graded as non-referable DR.

For the validation of detection of referable DR in Messidor-2 using the mentioned RS, the participant-based score provided by RetCAD v.1.3.0 was computed as the maximum score of both eyes' image-based scores. The DL system achieved an area under the receiver operating characteristic (ROC) curve of 98.0% (95% CI, 96.8%-99.0%), sensitivity was 92.6% (95% CI, 88.4%-97.4%) and specificity was 93.4% (95% CI, 89.9%-97.2%). The corresponding ROC analysis and diagnostic performance of the DL framework can be found in **Figure S4** and **Table S4,** respectively.

References

Abràmoff MD, Folk JC, Han DP et al. (2013): Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol **131**: 351-357.

ADCIS. Messidor dataset. http://www.adcis.net/en/DownloadThirdParty/Messidor.html.

LaTIM (2011): Messidor-2 dataset. http://latim.univ-brest.fr/indexfce0.html.

University of Iowa Health Care, Department of Ophthalmology and Visual Sciences, Michael D. Abràmoff, MD, PhD. https://medicine.uiowa.edu/eye/abramoff.

Wilkinson CP, Ferris FL, Klein RE et al. (2003): Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology **110**: 1677-1682.