# Supplementary Online Content

Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*.
doi:10.1001/jama.2016.17216



This supplementary material has been provided by the authors to give readers additional information about their work.

**eAppendix.** Supplemental Information

**Development dataset**

For developing our algorithm, a data set of macula-centered fundus images were acquired from 3 eye hospitals in India (Aravind Eye Hospital, Sankara Nethralaya and Narayana Nethralaya) and EyePACS in the USA. All images were de-identified according to HIPAA Safe Harbor prior to transfer to study investigators. The datasets from India were obtained from both eye hospital clinics and screening camps. The EyePACS data consists of patients that were screened using the EyePACS tele-ophthalmology platform from January 2013 to April 2015. EyePACS clinics serve higher percentages of the latino population in the U.S., therefore, the EyePACS dataset was enriched for Hispanic patients (~55%), with Caucasian, Black, and Asian patients each comprising approximately 5-10% of the population. Cameras were used to acquire the images include Centervue DRS, Optovue iCam, Canon CR1/DGi/CR2, Topcon NW8 using 45-degree fields of view. In total, the development dataset consists of 128,175 macula-centered images of which 33,894 were from India and the rest from EyePACS sites.

For the development set, between 3-7 grades were obtained for each image. The graders for the development set were U.S. licensed ophthalmologists or ophthalmology trainees in their last year of residency (PGY-4). All graders were asked to grade a 19-image test set prior to starting grading to ensure that they were proficient in reading diabetic retinopathy fundus images and were monitored for inter-grader and intra-grader consistency (details below *Grading Quality Control* section).

**Grading instructions**

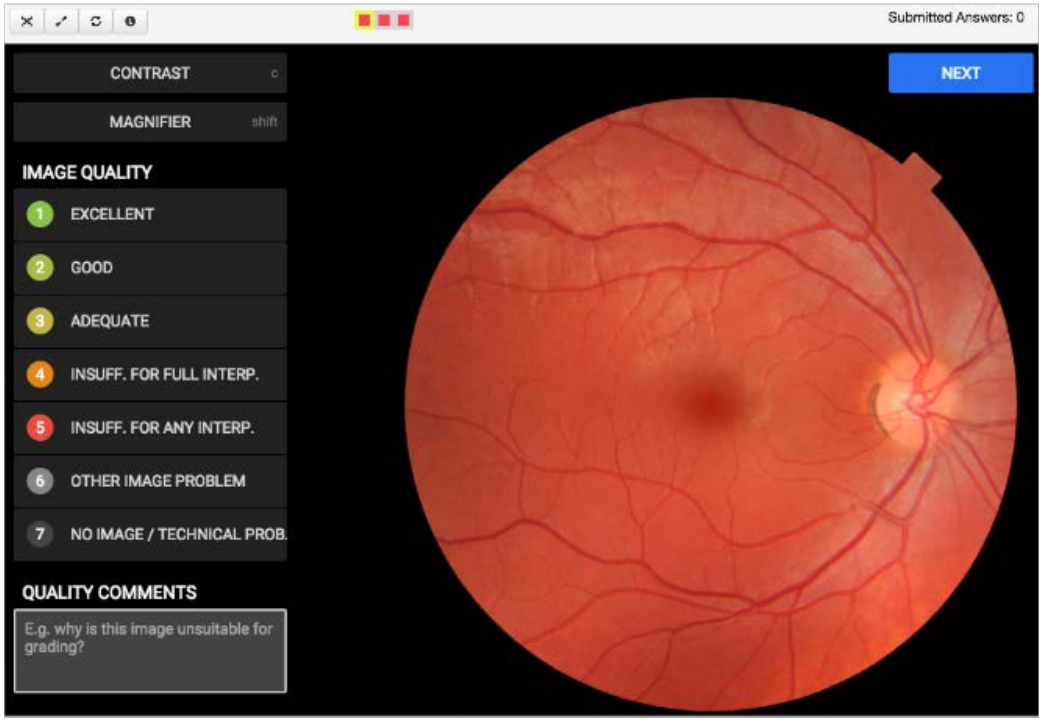
Graders were asked to grade all images according to the rubric below:

I. **Image Quality**. Select the image quality for each image before grading.
   A. Examine each image for the following image quality factors:
      1. Focus: Is the focus good enough for grading of smaller retinal lesions (e.g. MA, IRMA)?
      2. Illumination: Is the image too dark, or too light? Are there dark areas or washed-out areas that interfere with detailed grading?
      3. Image field definition: Does the primary field include the entire optic nerve head and macula?
      4. Artifacts: Is the image sufficiently free of artifacts (e.g. dust spots, arc defects, and eyelash images) to allow adequate grading?
   B. Select the correct classification for image quality:
      1. Excellent: No problems with any image quality factors. All retinopathy lesions gradable.
      2. Good: Problem with 1-2 image quality factors. All retinopathy lesions gradable.
      3. Adequate: Problems with 3-4 of the image quality factors. All retinopathy lesions gradable.
      4. Insufficient for Full Interpretation: One or more retinopathy lesions cannot be graded, but part of the image was gradable (e.g. neovascularization noted so likely PDR, but obscured view of the macula, so DME could not be graded)
      5. Insufficient for Any Interpretation.
      6. Other: Some other quality factor(s) interfered with grading. Please specify in Quality Comments section (e.g. not an retinal image).
      7. No image / technical problem

8. Quality Comments: if the image is not gradable, please specify the reasons (e.g. blurry, too dark)

C. Selecting 1-4 will bring you to an additional screen that will ask you to grade the image. If you select 4, you may skip the sections that are ungradable. Selecting 5-7 will bring up the next image for evaluation.

**eFigure 1.** Screenshot of First Screen of Grading Tool Which Asks Graders to Assess Image Quality




II. **DR/DME Severity**. If image is gradable, you will be asked to grade of diabetic retinopathy (DR)

A. Select DR grade based upon the attached international clinical diabetic retinopathy disease severity scale

B. Select referable diabetic macular edema (DME) if there are any hard exudates within one disc diameter of the center of the macula.

C. If any of the image is only partially gradable, only select a DR / DME grade only if that section is gradable. If you mistakenly select an option, you can reset the image by pressing the reset button on the top toolbar (see b in screenshot). This will reset the grading process for this image, so you will have to select image quality again.

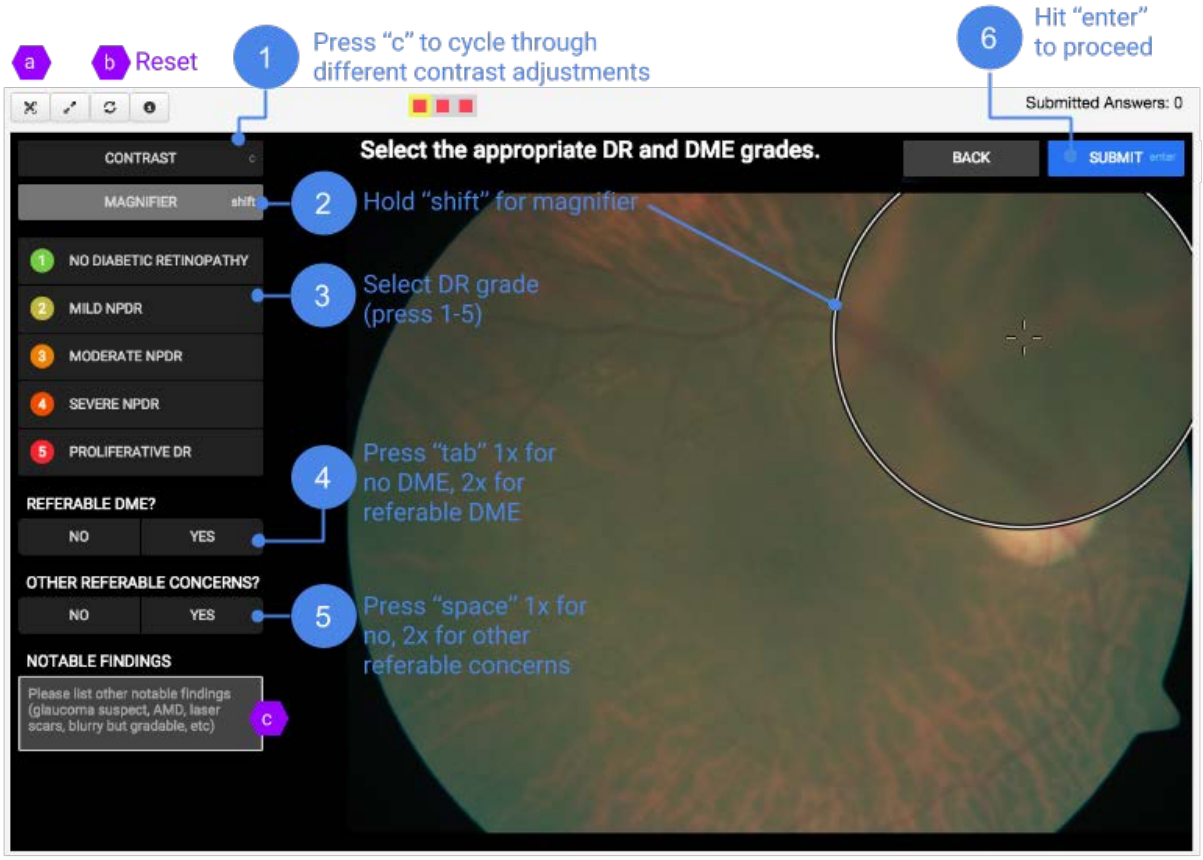
D. Note any remarkable findings in the image, even if it is not referable.

1. Examples include choroidal retinal scars, glaucoma suspect, AMD, laser scars, occlusions
2. PRP scars should be graded as PDR with “prp” noted in the comment box.
3. Focal laser scars should be noted as “focal laser” in the comment box

E. Optional tools (see purple hexagons in screenshot):

a) Fullscreen mode
b) Hit reset to reload this image. This will reset all of the grading.
c) Comment box for other pathologies you see

**eFigure 2.** Screenshot of the Second Screen of the Grading Tool, Which Asks Graders to Assess the Image for DR, DME and Other Notable Conditions or Findings




NPDR stands for non-proliferative diabetic retinopathy.

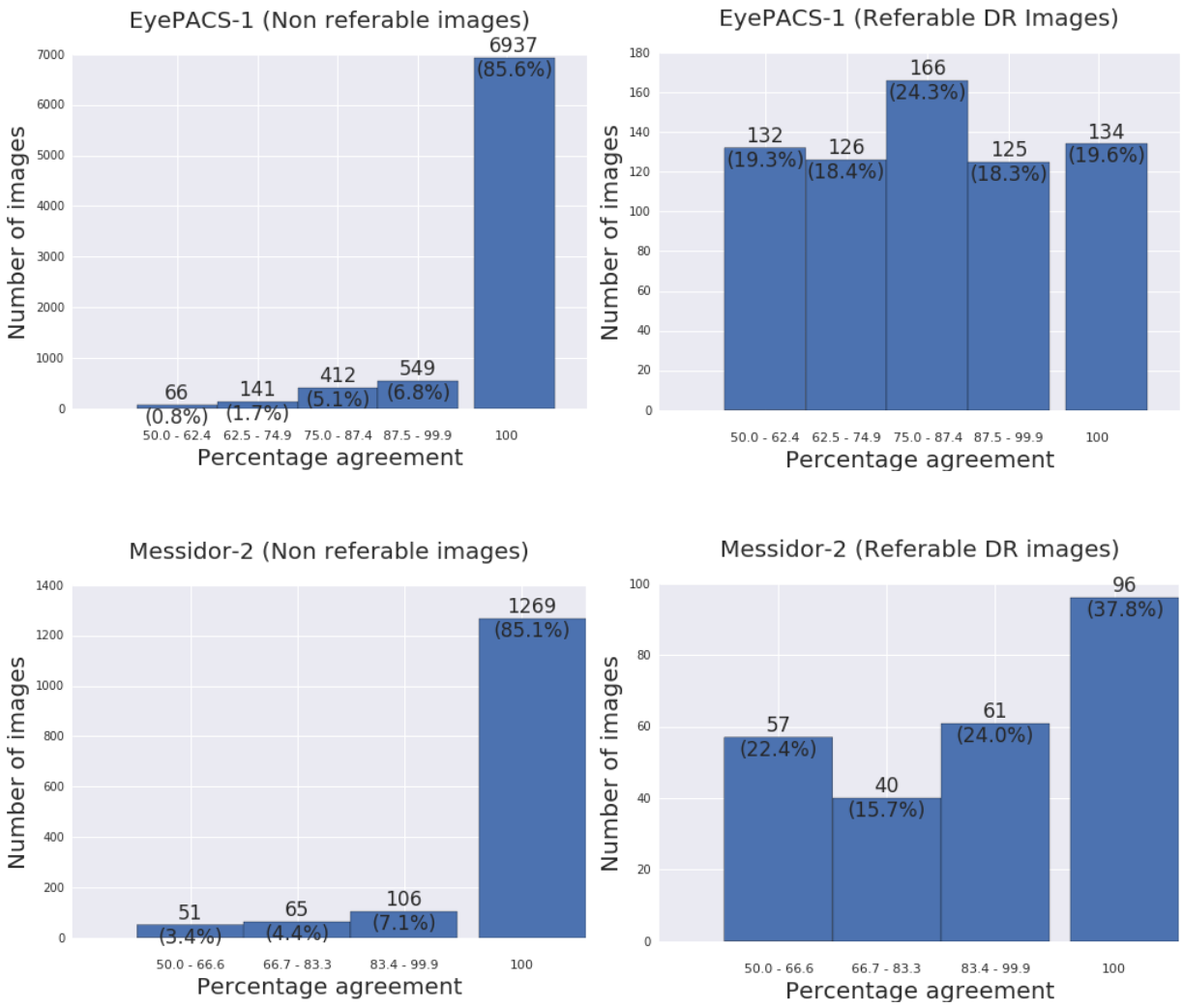
### Grading quality control

Inter-grader reliability of the panel was calculated for each physician using pairwise comparisons by taking the number of times a grader was in agreement with another grader over the total number of pairwise comparisons. Approximately 10% of the development set were overread to determine intra-grader reliability.

### Concordance amongst ophthalmologists

The plots below show the agreement amongst ophthalmologists for referable diabetic retinopathy.

**eFigure 3.** Distribution of Agreement Amongst Ophthalmologists on EyePACS-1 (8 Ophthalmologists) and Messidor-2 (7 Ophthalmologists**)**




The histograms are broken down by the majority decision for referable diabetic retinopathy, and restricted to images considered fully gradable by the majority. As an illustrative example, in the bottom-right histogram representing the distribution of agreements on Messidor-2 (referable DR images), the agreement range 50.0-66.6 (marked on the x-axis) has 57 images which corresponds to 22.4% of all the referable images (the numbers on the top of the bar). Not all images are assigned a grade by every ophthalmologist because of the option to mark an image ungradable. Hence the denominator for measuring the percentage agreement varies across images. A larger bin size is used for Messidor-2 (as compared to EyePACS-1), because the average number of grades per image is less in Messidor-2, providing for a coarser quantization in the agreement.

**Data pre-processing**

For algorithm training, input images were scale normalized by detecting the circular mask of the fundus image and resizing the diameter of the fundus to be 299 pixels wide. Images for which the circular mask could not be detected

were excluded from the development and the clinical validation sets. This corresponded to 117 out of 128,175 on the development set, 17 out of 9,963 in EyePACS-1, and none in Messidor-2.

**Performance on individual diabetic retinopathy subtypes**

The performance of the algorithm for individual subtypes of diabetic retinopathy (moderate or worse diabetic retinopathy only, severe or worse diabetic retinopathy only, and referable diabetic macular edema only) are described in the eTable 1.

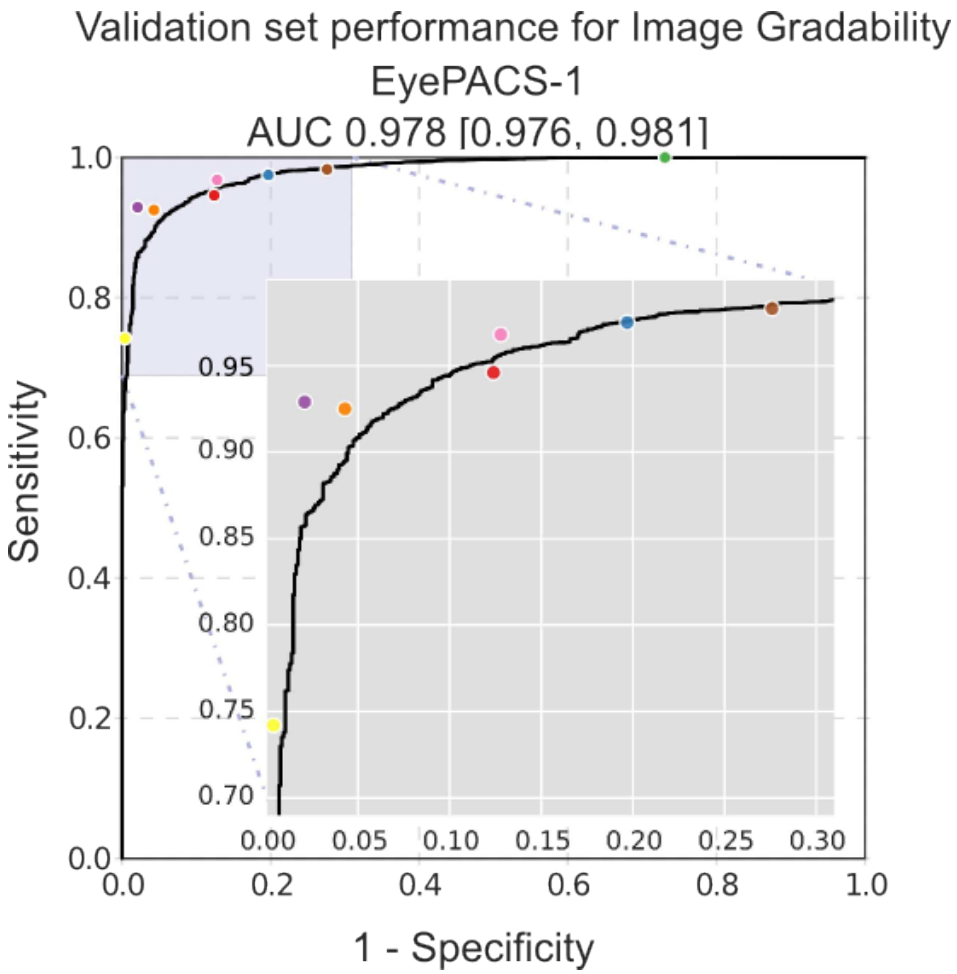
**eTable 1.** Performance of the Algorithm at the Ophthalmologist Operating Point in the EyePACS-1 (8788 Fully Gradable Images) and the Messidor-2 Datasets (1745 Fully Gradable Images). [95% Confidence Intervals]

| | EyePACS-1 | | Messidor-2 | |
|---|---|---|---|---|
| | Sensitivity | Specificity | Sensitivity | Specificity |
| Moderate or worse diabetic retinopathy only | 90.1% [87.2, 92.6] | 98.2% [97.8, 98.5] | 86.6% [80.5, 90.7] | 98.4% [97.5, 99.0] |
| Severe or worse diabetic retinopathy only | 84.0% [75.3, 90.6] | 98.8% [98.5, 99.0] | 87.8% [73.4, 96.0] | 98.2% [97.4, 98.9] |
| Diabetic macular edema only | 90.8% [86.1, 94.3] | 98.7% [98.4, 99.0] | 90.4% [81.9, 94.8] | 98.8% [98.1, 99.3] |

**Performance on image gradability**

Because patients with ungradable images are often referred to an ophthalmologist for further evaluation, the algorithm's performance was also evaluated for predicting gradable and ungradable images (Figure 2). The analysis was reported for the EyePACS-1 dataset only as there was only three image that were not fully in the Messidor-2 dataset. The algorithm's sensitivity was 93.9% and specificity was 90.9% for detecting fully gradable images.

**eFigure 4.** Performance of the Algorithm (Black Curve) and Ophthalmologists (Colored Dots) for Predicting Gradable Images on the EyePACS-1 Dataset (9946 Images)




| **Image Gradability** | EyePACS-1 | |
|---|---|---|
| | Sensitivity | Specificity |
| Algorithm | 93.9% [93.3, 94.5] | 90.9% [88.9, 92.7] |

The black dot denotes the performance of the algorithm at the first operating point [95% confidence intervals]. Data is shown only for EyePACS-1 because there were only 3 ungradable images in the Messidor-2 dataset. There were 8 ophthalmologists (shown as colored dots) who graded EyePACS-1.

**Performance on mydriatic and non-mydriatic eyes**

The performance of the algorithm was measured on known mydriatic and non-mydriatic subsets of the full EYEPACS-1 image set. Model performance on each subset and the combined set is shown in eTable 2.

**eTable 2.** Performance of the Algorithm for Detecting Referable Diabetic Retinopathy at the Ophthalmologist Operating Point on All EyePACS-1 Images for Which We Have dilation Data, Mydriatic Images Only, and Non-mydriatic Images Only [95% Confidence Intervals]

| **Image** | **Image Count** | EyePACS-1 | |
|---|---|---|---|
| | | Sensitivity | Specificity |
| Mydriatic | 4236 | 89.6% [85.6, 92.8] | 97.9% [97.3, 98.4] |
| Non-Mydriatic | 4534 | 90.9% [86.1, 94.4] | 98.5% [98.0, 98.8] |
| Both | 8770 | 90.1% [87.2, 92.6] | 98.2% [97.8, 98.5] |