SUPPLEMENTARY DATA

Data Supplement 2. Data Extraction Sheet

Study Identification	Journal published Publication type Country of study Setting of study
Study Design	Study methodology (consecutive or random; retrospective or prospective) Start date End date DR severity scale used Definition of RWDR, VTDR and/or STDR
Participants	Total number of participants included in study Total number of participants excluded from study Number of withdrawals Sample size (patients) Sample size (eyes) Inclusion criteria Exclusion criteria Mean age and age range Gender Type of diabetes Mean duration of diabetes Spectrum of presenting symptoms and comorbidities Current treatment
Index Test	Imaging technique Attachment used (Y/N) Type of attachment used Weight of attachment Type of smartphone used Cost excluding smartphone Fixation target Light source Mydriasis (Y/N) Number of fields Field of view (in degrees) Fields/Regions of retina imaged Stereoscopic (Y/N) Colour (Y/N) Image resolution Healthcare professional acquiring retinal images Healthcare professional grading images Time taken to acquire images Time taken to stitch images (if applicable) Stitching software used (if applicable) Other software used Was AI used for grading retinal images? (Y/N) AI software used for grading (if applicable)

Reference Standard	Type of reference standard Healthcare professional performing reference standard Time between index test and reference standard Blinding (Y/N)
2×2 Table	TP TN FP FN
Sensitivity (95% CI)	No DR Mild NPDR Moderate NPDR Severe NPDR Very severe NPDR Early PDR High-risk PDR Severe PDR Macular oedema Others
Specificity (95% CI)	No DR Mild NPDR Moderate NPDR Severe NPDR Very severe NPDR Early PDR High-risk PDR Severe PDR Macular oedema Others
PPV and NPV	Positive Predictive Value (DR) Negative Predictive Value (DR) Positive Predictive Value (DME) Negative Predictive Value (DME)
Image quality	Categories (e.g. Excellent, moderate…) Number of images in each category Number of ungradable images
Graders	Number of graders Number of eyes/patients assessed by each grader Agreement between graders

Quality assessment of included studies using QUADAS-2

Patient Selection	Description
	Was a consecutive or random sample of patients enrolled? Was a case-control design avoided? Did the study avoid inapprorpiate exclusions?
	Risk of bias (high/ low/ unclear)
	Applicability Concerns
Index Test	Description
	Were the index test results interpreted without knowledge of the results of the reference standard? If a threshold was used, was it prespecified?
	Risk of bias (high/ low/ unclear)
	Applicability Concerns
Reference Standard	Description
	Is the reference standard likely to correctly classify the target condition? Were the reference standard results interpreted without knowledge of the results of the index test?
	Risk of bias (high/ low/ unclear)
	Applicability Concerns
Flow and Timing	Description
	Was there an appropriate interval between index tests and reference standards? Did all patients receive a reference standard? Were all patients included in the analysis?
	Risk of bias (high/ low/ unclear)

Source: Whiting et al, 2011. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies