Survey of Ophthalmology 69 (2024) 24–33

Contents lists available at ScienceDirect

Survey of Ophthalmology

journal homepage: www.elsevier.com/locate/survophthal

Review article

Early diabetic eye damage: Comparing detection methods using diagnostic power

Bhim B. Rai, PhD [a,b,*,1], Joshua P. van Kleef, PhD [a,b,2], Faran Sabeti, PhD [a,c,3], Robin Vlieger, PhD [d,4], Hanna Suominen, PhD [b,d,e,5], Ted Maddess, PhD [a,b,6]

[a] John Curtin School of Medical Research, Australian National University, Canberra, ACT, Australia
[b] ANU Eccles Institute of Neuroscience, Australian National University, Canberra, ACT, Australia
[c] School of Optometry, Faculty of Health, 2 University of Canberra, Canberra, ACT, Australia
[d] ANU School of Computing, Australian National University, Canberra, ACT, Australia
[e] University of Turku, Turku, Finland

ARTICLE INFO

Keywords:
Young people with diabetes
Early diabetic retinopathy
Diagnostic tests
Effect size
Visual function
Retinal structure
Combined measures
Machine Learning
Deep Learning
Endpoints

ABSTRACT

It is now clear that retinal neuropathy precedes classical microvascular retinopathy in diabetes. Therefore, tests that underpin useful new endpoints must provide high diagnostic power well before the onset of moderate diabetic retinopathy. Hence, we compare detection methods of early diabetic eye damage. We reviewed data from a range of functional and structural studies of early diabetic eye disease and computed standardized effect size as a measure of diagnostic power, allowing the studies to be compared quantitatively. We then derived minimum performance criteria for tests to provide useful clinical endpoints. This included the criteria that tests should be rapid and easy so that children with type 1 diabetes can be followed into adulthood with the same tests. We also defined attributes that lend test data to further improve performance using Machine/Deep Learning. Data from a new form of objective perimetry suggested that the criteria are achievable.

1. Introduction – comparing diagnostic power in early diabetic retinopathy

Diabetes mellitus (DM) continues as an epidemic,[85] with an estimated global prevalence at 463 million.[84] The complications of DM, both systemic and ocular, are serious.[58,97] Diabetic retinopathy (DR) is a major ocular complication and is a leading cause of vision loss.[101] The global prevalence of DR among people with diabetes (PwD) is 19–25%.[96] The later-stage presentations are proliferative DR (PDR),[17,58] with 6% of PwD developing vision-threatening DR (VTDR) and 4% developing diabetic macular edema (DME), which affects central vision directly.[96] By 2045, it is projected that there will be 700 million PwD, 160.5 million with DR, 44.8 million with VTDR, and 28.6 million with DME.[84,96]

Although treatments such as retinal laser or antivascular endothelial growth factor (anti-VEGF) injections are available, they are not without complications.[25,88,100] Drugs like candesartan may prevent earlier-stage DR,[15] and fenofibrate may prevent or reverse progression of early-stage DR in type 2 diabetes.[35,44] New treatments may arise from those findings.

Importantly, people with early-onset diabetes, type 1 diabetes (T1D), have higher lifetime risk of developing DR, suffering blindness and experiencing greater socioeconomic impact.[89] It is estimated that 95–97% of people with T1D will be negatively affected in their

* Corresponding author: Bhim Bahadur Rai, PhD, John Curtin School of Medical Research, Australian National University, 131 Garran Road, Acton, Canberra, ACT 2607, Australia. Phone: +61 2 6125 9253.
E-mail address: bhim.rai@anu.edu.au (B.B. Rai).
[1] ORCID ID: 0000 0003 0748 4581
[2] ORCID ID: 0000 0003 2167 0348
[3] ORCID ID: 0000 0001 9187 7569
[4] ORCID ID: 0000 0001 8492 6203
[5] ORCID ID: 0000 0002 4195 1641
[6] ORCID ID: 0000 0003 4591 3658

https://doi.org/10.1016/j.survophthal.2023.09.002
Received 17 March 2023; Received in revised form 22 August 2023; Accepted 6 September 2023
Available online 4 October 2023

lifetime,[8,34] and 3–8% of them will suffer blindness.[33,46] Therefore, we would like to be able to identify early eye damage and track it over the lifetime of these people employing the same tests. Given the need to restrict DR to moderate nonproliferative DR (NPDR) over the lifetimes of PwD, we require tests that can detect different stages of functional damage, from eyes with no apparent DR to mild DR.

We now know that functional damage precedes classical vasculopathy of mild NPDR[57,63,67] due in part to diabetic retinal neurodegeneration preceding other structural changes.[64] The classical visible markers of DR present only after considerable diabetic retinal neurodegeneration has occurred.[87] Taken together this information indicates that to provide useful endpoints for future clinical trials, we need to discover and develop tests that discriminate normal controls (NCs) from PwD with no classical retinopathy with high sensitivity, specificity, and precision. Obviously, diagnostic performance should be higher still in moderate DR.

We compare detection methods of early diabetic eye damage by reviewing data from a range of functional and structural studies of early diabetic eye disease and computing measures of diagnostic power of their methods, allowing the diagnostic tests to be compared quantitatively. A comprehensive review of functional tests addressing the full range of DR was published in 2021.[16] Most of the reviewed studies did not compare PwD who did not yet show DR (No-DR) with NC subjects. Further, that review did not compare the diagnostic power of the methods used, providing the impetus for this review.

Here we use standardized effect sizes to compare the diagnostic power of functional and structural tests for discriminating NC subjects and PwD with No-DR. Diagnostic power can often be estimated from published data using standardized effect sizes, for example, Cohen's *d*: the standardized mean difference of pairs of distributions.[19] A slightly more conservative version, Hedge's *g*,[38] provides a conservative penalty for smaller study group sizes. Standardized effect sizes are closely related to area under receiver operating characteristic (AUROC) plots.[75] Effect sizes approaching 2, corresponding to AUROCs of >90%,[61] are the target level of diagnostic power for a clinically useful test.[90] We also discuss an evolving form of objective perimetry that in several studies seems to meet those criteria.[6,54,79,82] Finally we discuss the statistics of how independent measures may usefully be combined to increase diagnostic power and the requirements for validating more complicated diagnostic models. Related factors such as correlation with complications screening variables, ease of use and test duration are also discussed. These are important factors given that we would like to employ the same tests for PwD over their entire lives, especially those who are diagnosed in childhood.

2. Methods

We searched online between September, 2022, and July, 2023, for published studies reporting on various methods for discriminating NC subjects from PwD with No-DR: *the NC* vs *No-DR comparison*. A few studies comparing NCs and prediabetes, or No-DR vs mild nonproliferative retinopathy (NPDR) were included. Studies of more severe DR were excluded. For comparison, we examined both functional and structural tests. Some studies examined several tests or provided independent assessments for several variables per test. Functional tests included behavioral tests, generally involving some visual threshold, electrophysiology, and pupillographic testing. Structural tests were largely based upon some sort of imaging of the retinal or cornea and could involve forms of retinal angiography.

The main search tool was the Web of Science (WoS, Clarivate, London, UK). Key starting points were known papers and those highlighted by a recent review of methods for more serious DR[16] and WoS was used to examine all papers citing those to discover more recent studies on the same or related topics. Searches for other papers were done containing disease terms including diabetes + eye, diabetes + retinopathy (and substituting diabetic for diabetes), and with vasculopathy and neuropathy were combined with those for functional and structural test methods. Functional terms included retinogram, electroretinogram, ERG, evoked + potential, VEP, threshold, perimeter, perimetry, pupillometry, and pupil. Structural terms included retinal + photography, fundus + photography, fundus + camera, optical + coherence + tomography, OCT, microscopy, macular + pigment, angiography, and cornea were used. When a suitable reference was found WoS was again used to find papers that cited it.

To assess relative diagnostic power we used Hedge's g[38] because many studies provide the necessary inputs: the means, standard deviations, and numbers of subjects from each of their study groups. A few studies provided AUROC data, so in these cases were converted them to Hedge's g.[75] In this way Hedge's g and AUROC values were provided for each reported variable. Studies not providing the inputs to calculate Hedge's g, or AUROC values were therefore excluded. To aid interpretation we report AUROCs as percentages, 50% being chance performance and 100% perfect discrimination. We compared the diagnostic performance of different groups of methods using linear models (fitlm function of Matlab 2020b, The MathWorks, Natick, MA). Two separate models were formed for the Hedge's g and AUROC data.

We briefly provide an illustration of the difference between Cohen's d and Hedge's g. Using data from Joltkov et al[42] for the pattern standard deviation for the 10-2 test of the matrix perimeter gives mean, and standard deviation (SD), and sample size (N) for NCs and No-DR diabetics of: 2.43 ± 0.28 and 2.86 ± 0.67 respectively with $N = 18$ and 23. Hedge's g is 0.819, while Cohen's d (which assumes an N of ∞) is 0.837. Doubling the N values gives a value for g of 0.828. Thus, the number of subjects, N, does not change the effect size greatly but using g makes the values obtained from different studies comparable. Hedge's g is also the basis for many power calculations,[26,27] and effect sizes of 0.2, 0.5, 0.8, 1.2 and 2.0 are the accepted cut-offs for small, medium, large, very large, and huge effect sizes, respectively. As mentioned, clinically useful tests will minimally have effect sizes around 2 and therefore AUROCs >90%.

Fig. 1 illustrates why we cannot use P-values to demonstrate clinical utility. P-values indicate our confidence in the difference in the means of the empirical distributions for NC and No-DR people.[90] Put simply, P-values are heavily determined by N, the number of subjects in the study groups. For diagnosis $N = 1$, the person being examined. In that case, the only thing that matters is the relative separation the distributions for NC and PwD. Standardized effect sizes give us that.

3. Functional studies

Eight functional methods studies met the study criteria, and the data for 17 of the reported measures are given in Table 1. Hedge's g ranged from 0.05 to 1.20, yielding a median of 0.47. The corresponding AUROCs ranged from 51 to 80%, providing a median of 63%. In a few cases, the same measure was reported by different researchers. The first 13 measures were from perimeters, the first two from microperimetry targeting the central 20° (6 mm), and the remaining tests examined the central 48°. The AST contrast sensitivity was measured within the central 5–6°,[42] and the high temporal frequency responses of the two ERG methods were assessed with a ganzfeld stimulus.[55,103] The median effect size of 0.47 was small to medium, indicating low power. The highest value of g was 1.2 reported for the SWAP mean defect (MD), but another group reported a g of 0.39 for the same measure. The absence of DR in the Afrashi and coworkers study[1] was determined by slit lamp ophthalmoscopy (SL).

4. Studies using pupillometry

Pupillometry is a methodology in which pupil diameters are monitored in response to a single large light stimulus. Generally, one pupil/eye at a time is tested. There have been several studies comparing NC and No-DR subjects and so we summarize them separately to the other functional methods. The stimulus duration employed is often quite long,

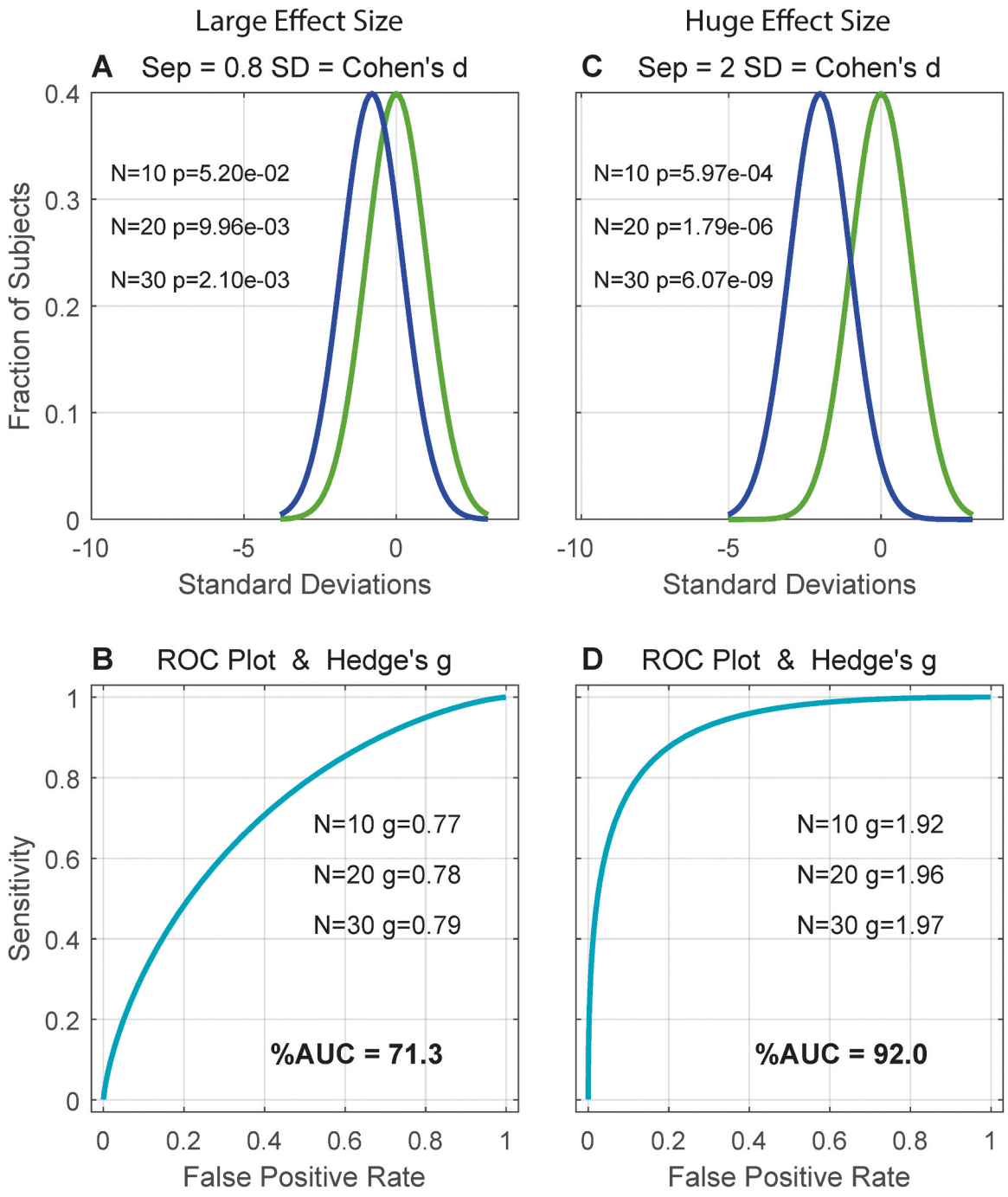

Fig. 1. In this example the pairs of distributions in A and C have standard deviations (SD) of 1. In that case, the separation of the distributions in SD is Cohens' *d*. Notionally the blue distributions are for those with diabetes and the green the control group. *A,B:* Are for the case of a large effect size of $d = 0.8$; (C,D) are for a huge effect size, $d = 2$. Thus, in C the distributions are more widely separated than in A. In each of A and C there are three lines of text beginning "N = " and then a P-value. The Ns are the number of persons in each group in the study, and the P-value is the outcome for a t-test given the distributions and Ns. In C the P-value for N = 30 is 6×10^{-9} B and D give the ROC plots for the distributions in A and C. The larger bold text in B and D shows the AUROC. In each of B and D there are 3 lines of text beginning with "N = " and then the value of Hedge's *g*. As in the example in the text from Joltkov et al[42] the effect of small N upon *g* is slight. Thus, in D when $d = 2$, for N = 10 per group $g = 1.92$. Therefore, the estimated diagnostic power depends little on N, only the separation of the distributions. AUROC = area under receiver operating characteristic.

5–20 seconds,[5,9,28,40,95] and recovery from those flashes is frequently monitored. Others have used shorter pulses of between 1 second and 40 ms.[36,62,102] The baseline pupil diameter to particular adapting light levels also has diagnostic value.[62] Table 3 summarizes the 1 or 2 methods with the highest diagnostic power from each study. The median performance of Hedge's $g = 0.84$ (large effect size) and AUROC = 72% is somewhat better than the measures of Table 1, but the maximum performance of the pupillometry methods was similar at $g = 1.25$ (very large effects size) and AUROC = 81%. Two studies that examined the NC vs No-DR comparison were left out of the table as they found no significant differences between those groups and so would have registered a *g* close to 0 and an AUROC close to 50%.[5,45] Adding those values to the table would have changed the median values to $g = 0.41$ and AUROC = 62%.

5. Studies with the ObjectiveField Analyzer functional test

Four studies with ObjectiveField Analyzer (OFA) as their functional testing method met the study criteria and the data for nine of the reported measures are given in Table 3. The OFA is a novel functional test that has been evolving over time (see Discussion). It provides objective perimetry by combining multifocal stimulation, as in a multifocal ERG or evoked potential, and noncontact measurement of relative changes in pupil diameter. The method has therefore been termed multifocal pupillographic objective perimetry (mfPOP), and thus provides spatially resolved sensitivity and delay data for up to 44 test regions/eye. Both eyes are tested concurrently and generate localized afferent field defects.[6] The OFA is a version of mfPOP is cleared by the United States Food and Drug Administration (FDA) and manufactured by Konan Medical USA (Irvine, CA)—which provides five variants of mfPOP. Fourth-generation stimuli test both eyes in 7 minutes and produce 30-2 style reports or a half-scale macular variant. Unlike most multifocal systems, both eyes are tested concurrently. OFA uses very transiently presented pseudorandom stimuli, and the resulting responses appear to be driven by the extrastriate cortex,[80] matching the neuro-anatomical expectation.[29]

The effect sizes in the included studies with the OFA ranged from very large to huge (Table 3). One entry in Table 3 gave *g* and AUROC values for comparing No-DR and mild to moderate DR (Mild-Mod). The most recent 2023 results were for fifth-generation stimuli that test both eyes in under 90 seconds. Overall, the Hedge's *g* values ranged from 1.53 (very large) to >2.5 (huge) with the median of 1.74 (very large) and the corresponding AUROC values from 86 to 100% (89% median). The between-eye asymmetries reported were calculated from anatomically equivalent regions of the two eyes that are tested concurrently. By contrast, if one eye is measured at a time, as in standard perimetry, test time is confounded with tested eye.

6. Structural studies

We found 4 studies where NC eyes were compared with No-DR eyes, and one where eyes of people with prediabetes or early diabetes were examined (Table 4). The measures reported were from polarization-sensitive optical coherence tomography, retinal nerve fiber layer birefringence, several OCT-angiography variables, corneal nerve fiber density, and vessel diameter change in response to a flickering 30° diameter light. The Hedge's *g* values ranged from 0.28 to 0.92 (median 0.52) and the corresponding AUROC values from 58 to 87% (65% median).

7. Comparisons

We next compared the outcomes for Tables 1–4 using separate linear models of the Hedge's *g* and AUROC values. Both models were highly significant (see bottom of Table 5). The fits were to an additive model using a categorical variable that identified tests as being from Tables 1, 2, 3 or 4. In these models the reference or constant term was the mean of the functional tests of Table 1. The models contained additive amounts fitted to the structural, pupillographic and OFA data to quantify the mean difference between each and the mean of the functional tests. The t-statistic and P-values thus quantified the significance of those

Table 2
Pupillometry tests comparing No-DR and control subjects.

First author	Year	Measure	Discrimination	Hedge's *g*	AUROC
Cankurtaran[9]	2020	2 or 4 s post flash size	Controls vs No-DR	0.50	64%
Feigl[28]	2012	Re-dilation kinetics	Controls vs No-DR	0.89	73%
Halperin[36]	2016	Initial latency	Controls vs No-DR	0.78	71%
Jain[40]	2018	Dilation velocity	Controls vs No-DR	0.29	58%
Park[62]	2017	Scotopic baseline diameter	Controls vs No-DR	0.94	75%
Park	2017	Photopic baseline diameter	Controls vs No-DR	0.68	69%
Tan[95]	2022	Maximum blue constriction	Controls vs No-DR	1.25	81%
Tan	2022	Maximum red constriction	Controls vs No-DR	0.99	76%
Medians				0.84	72%
Maximums				1.25	81%

AUROC = area under receiver operating characteristic.

Table 1
Functional tests comparing No-DR and control subjects.

First author	Year	Measure	Discrimination	Hedge's *g*	AUROC
Gella[31,32]	2016	Microperimetry (MP1) 8°	Controls vs No-DR	1.09	78%
Montesano[57]	2021	Microperimetry (Maia) MS	Controls vs No-DR	0.29	58%
Montesano	2021	Matrix/FDP 24-2 perimetry MD	Controls vs No-DR	0.19	55%
Nitta[59]	2006	SWAP 24-2 MD	Controls vs No-DR	0.05	51%
Afrashi[1]	2003	SWAP 24-2 MD	Controls vs No-DR (SL)	1.20	80%
Nitta	2006	SWAP 24-2 PSD	Controls vs No-DR	0.35	60%
Afrashi	2003	SWAP 24-2 PSD	Controls vs No-DR (SL)	0.16	55%
McAnany[56]	2023	Octopus 900, photopic MD	Controls vs No-DR	0.69	69%
McAnany	2023	Octopus 900, scotopic PSD	Controls vs No-DR	0.71	70%
Joltikov[42]	2017	SWAP 24-2 MD	Controls vs No-DR	0.39	60%
Joltikov	2017	SAP 24-2 perimetry PSD	Controls vs No-DR	0.41	62%
Joltikov	2017	Matrix/FDP 24-2 perimetry FT	Controls vs No-DR	0.47	63%
Joltikov	2017	Matrix/FDP 10-2 perimetry PSD	Controls vs No-DR	0.82	72%
Joltikov	2017	Rarebit perimetry fovea MHR	Controls vs No-DR	0.37	61%
Joltikov	2017	AST contrast sensitivity	Controls vs No-DR	0.99	77%
McAnany[55]	2018	ERG - temporal frequency	Controls vs No-DR	1.09	*78%
Zeng[103]	2019	RETeval ERG + Pupil + Age	Controls vs No-DR	0.52	67%
Medians				0.47	63%
Maximums				1.20	80%

AUROC = area under receiver operating characteristic; FDP = frequency doubling perimetry; SWAP = short-wavelength automated perimetry; SAP = standard white-on-white automated perimetry; ERG = electroretinogram; MD = mean defect; PSD = pattern standard deviation; MS = mean sensitivity; FT = foveal threshold; MHR = mean hit rate.
(SL) in the discrimination column means the lack of DR was assessed by slit-lamp bio-microscopy.[36]
* The AUROC was reported and we converted it to Hedge's *g*.

Table 3
ObjectiveFIELD Analyzer (OFA) studies comparing No-DR and control subjects.

First author	Year	Measure	Discrimination	Hedge's *g*	AUROC
Bell[6]	2010	OFA sensitivities	Controls vs No-DR	1.60	87%
Bell	2010	OFA sensitivity asymmetries	Controls vs No-DR	2.33	95%
Sabeti[79]	2015	OFA sensitivities	Controls vs No-DR	1.65	88%
Sabeti	2015	OFA delay asymmetries	Controls vs No-DR	>2.5	100%
Sabeti[82]	2022	OFA sensitivities and delays	Controls vs No-DR	1.80	90%
Sabeti	2022	OFA sensitivities and delays	No-DR vs Mild-Mod	1.53	86%
Maddess[54]	2023	OFA M18 sensitivities and delays	Controls vs No-DR	1.80	90%
Maddess	2023	OFA W20 sensitivities and delays	Controls vs No-DR	1.74	89%
Maddess	2023	OFA M18 and W20 sensitivities	Controls vs No-DR	1.60	87%
Medians				1.74	89%
Maximums				2.50	100%

AUROC = area under receiver operating characteristic.

Table 4
Structural tests comparing No-DR and control subjects.

First author	Year	Measure	Discrimination	Hedge's *g*	AUROC
Pollreisz[65]	2021	PS-OCT RNFL birefringence	Controls vs No-DR	0.57	66%
Sung[91]	2022	OCT-angiography VD - center 1 mm	Controls vs No-DR	0.28	58%
Sung	2022	OCT-angiography VD - ring	Controls vs No-DR	0.92	74%
Sung	2022	OCT-angiography VD - full 3 mm	Controls vs No-DR	0.47	63%
Chao[14]	2020	Corneal nerve fiber density	Controls vs prediabetes	0.32	59%
Chao	2020	Corneal nerve fiber density	Controls vs early-diabetes	0.92	74%
Lim[48]	2014	Retinal vessel diam - flicker - artery	No-DR vs mild NPDR	0.39	61%
Lim	2014	Retinal vessel diam - flicker - vein	No-DR vs mild NPDR	0.70	72%
Zeng[104]	2023	OCT-angiography VD - SVC	Controls vs No-DR	1.57	87%
Zeng	2023	OCT-angiography VD - DVC	Controls vs No-DR	1.54	86%
Medians				0.64	69%
Maximums				1.57	87%

AUROC = area under receiver operating characteristic; PS-OCT = polarization-sensitive optical coherence tomography; RNFL = retinal nerve fiber layer; NPDR = nonproliferative diabetic retinopathy; VD = vessel density; SVC = superficial vascular complex; DVC = deep vascular complex.
In early T2 diabetes group 2/14 had moderate NPDR.[42]

Table 5
Comparing methods by Hedge's *g* and AUROC.

Hedge's *g*				
	Estimate	SE	t-statistic	P-value
Functional	0.51	0.10	5.17	0.0001<
Pupillography	0.28	0.16	1.68	0.102
Structural	0.25	0.15	1.65	0.107
OFA	1.33	0.16	8,34	0.0001<
AUROC				
	Estimate%	SE %	t-statistic	P-value
Functional	63.9	2.2	29.0	0.0001<
Pupillography	7.0	3.7	1.92	0.063
Structural	6.1	3.4	1.80	0.080
OFA	26.4	3.5	7.49	0.0001<

AUROC = area under receiver operating characteristic; OFA = ObjectiveFIELD Analyzer.
Hedge's *g* model: Adjusted R^2 = 0.641; F-stat 24.8, P = 5.8e-09.
AUROC model: Adjusted R^2 = 0.580; F-stat 19.4, P = 6.4e-08.

differences. In the case of the functional tests, the t-statistic and P-value values indicate the significance of the difference from 0.

Thus, in the case of Hedge's *g*, the mean value for the functional tests was 0.51 ± 0.10. The mean difference for the pupillographic tests was marginally significantly larger at 0.28 ± 0.16 (P = 0.102). At 0.25 ± 0.15, the structural tests were also marginally significantly different (P = 0.107). The mean *g* for the OFA tests was significantly larger by 1.33 ± 0.16 (P < 0.0001). Notice that the predicted OFA total *g* of 0.51 + 1.33 = 1.84 is very similar to the median value of Table 3 (cf. 1.74). Similar outcomes occurred for the AUROCs where the mean difference of the pupillography tests relative to the functional tests was a possible increase of 7.0 ± 3.7% (P = 0.080), and for the OFA tests 26.4 ± 3.5% (P < 0.0001). The predicted total *g* for OFA tests was thus 63.9 ± 26.4 = 90.3%, close to the median value of Table 3 (cf. 89%).

8. Discussion

As mentioned at the outset, by 2045 it is projected that there will be 700 million PwD, 160.5 million with DR, 44.8 million with VTDR, and 28.6 million with DME.[84,96] The magnitude of the problem is illustrated by the fact that diabetes is no longer a disease that impacts only the affluent societies. Even in one small developing country, Bhutan, DR/DME was the third most common retinal disorder,[68] the most common indication for retinal laser therapy,[69] and the third most common indication for retinal surgery.[70] Similarly in Nepal, DR was the second most common retinal disorder among patients[66] with significant systemic associations.[98] Thus, more than ever it is important to compare diagnostic power of potential tests and to understand other aspects that can augment that information.

We therefore posit 6 criteria that clinically useful tests should have:

1. Standardized effect sizes of >1.80, that is, AUROC >90%, to discriminate NC subjects from PwD but who show ***no classical DR***, and higher power for mild to moderate DR,
2. High reproducibility to maximize ability to track change over time,
3. Good ease of use and short test duration to permit the same test to be used in all age groups, from children to all adults, including the elderly and infirm,
4. Ability to easily compare structure and function measures at the same retinal regions,
5. Good correlation of single structural and functional measures with independent factors within standard complications screening variables,
6. Multiple measures that are as uncorrelated with each other as possible, but which are correlated with diabetic eye and/or tissue damage.

Tables 1–5 illustrated that OFA has clinically useful diagnostic power

in early-stage disease. Higher power is desirable—or even required—for useful clinical endpoints. This means future studies with new technologies need to report diagnostic power. Analogously to calls for the use of effect sizes,[90] journal editors and referees need to demand this, for example, though reporting guidelines.[86]

With respect to our second criterion above, the measures in Tables 1 and 3 have not had their reproducibility assessed. It is worth noting that the very small Goldmann size III stimulus of standard automated perimeters covers <0.5% of the 6° square cells of the test grid, contributing to[49,51,60] their high test-retest variability.[4] Two fourth-generation OFA tests, which provide 30-2 style visual field reports, have lower test-retest variability of standard automated perimeters in 40 people living with glaucoma and 95 matched NCs.[52] For example, for initial visit sensitivities of 4–16 dB the mean (±SD) of the interquartile range of SITA Fast perimetry is 14.6 ± 4.2 dB,[4] where for OFA it is 7.8 ± 0.7 dB.[52] Supplementary Fig. 1 presents test-retest variability data for persons with type 2 diabetes using the same two fourth-generation OFA tests as the glaucoma group.[52] As expected for lower test-retest variability, OFA has recently shown significant ability to track progression of mild DME over 1.5 years when matrix perimetry cannot.[83] That paper and another[73] have supplementary data sets showing OCT, OFA sensitivity and delay, and matrix data in an easy-to-compare (over time) format for up to 128 data sets. Those fourth-generation tests have been reported to have no safety issues in people with migraine[2] or epilepsy.[3]

With respect to our third criterion, OFA tests can be easily performed on persons of all ages and all disease stages, from an eye without any damage to severe DR, in as little as 90 seconds for both eyes. Older third-generation OFA tests have already been shown to be well correlated with standard diabetic tissue damage and metabolic variables in older people (41.8 ± 12.1 years) with T1D.[82] Being rapid, noncontact, and requiring no subject input, the latest fifth-generation OFA tests can be easily performed on persons of all ages and are also ideal for elderly and infirm people; the 2023 OFA study of Table 2 used two different fifth-generation tests that each assess both eyes in <90 seconds.[54] Those tests were performed on 52 young people with T1D (15.2 ± 3.77 years, 26 females), the youngest subjects were 8 years. Being noncontact, and requiring no subject input, those OFA tests can be easily performed on people of all ages and are ideal for the young, elderly, and infirm. Thus, the same tests can be used to track children over their lifetime.

The two tests of the fifth-generation OFA, called W20 and M18, have 20 and 18 stimuli/eye, respectively, and of them, the M18 test satisfies our fourth criterion. Specifically, the wide-field W20 test stimuli comprise an array of 20 stimuli/eye that span the central 60°. W20 has demonstrated high diagnostic power in adult people with multiple sclerosis.[53] M18 assesses the central 20° (6 mm, i.e., the macula) with an array of 18 stimuli. The M18 stimuli are the size and shape of the 9 Early Treatment Diabetic Retinopathy Study (ETDRS) retinal thickness grid regions cut in two (Fig. 2). Thus, it is easy to compare OCT retinal thickness data from the ETDRS grid and with the per-region sensitivity and delay data of M18. This satisfies point 4 above.

Returning to the age considerations of our third criterion above, M18 has also been shown to have high diagnostic power in people with early- to late-stage age-related macular degeneration (AMD).[74] That study indicated M18 had superior diagnostic power to either best corrected visual a(BCVA) or ETDRS retinal thickness and volume measures in all but late-stage disease, where M18 performed as well as the OCT measures.

With respect to our first, third, and sixth criterion above, in the T1D study of young people,[54] whose data are shown in Table 2, the best diagnostic power for a single measure was for combined sensitivities and delays, providing AUROCs around 89% in the persons under age 16. Without combining delays, the M18 and W20 sensitivities on their own yielded AUROCs <71%. The Kullback–Leibler divergence of the M18 and W20 per-region sensitivity data indicated that they were highly independent. As might be expected for independent measures that are both correlated with disease but not with each other (our sixth criterion above) a linear combination of the M18 and W20 per-region sensitivities produced AUROCs of 92%. It seems reasonable that M18 and W20 would produce quite independent data given that W20 covers 9 times greater retinal area than M18.

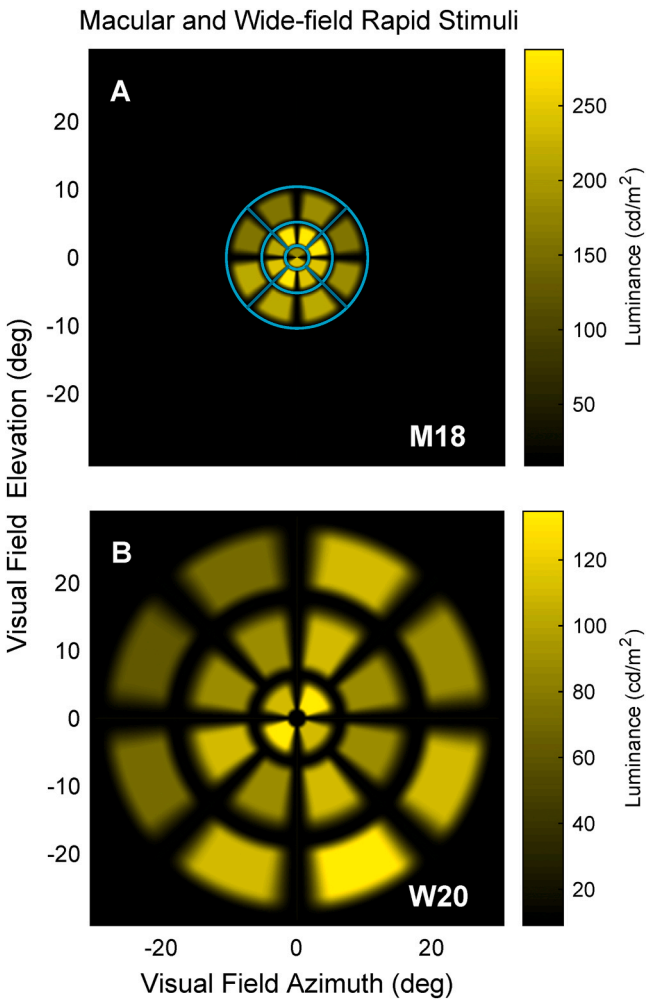

Fig. 2. The spatial layout of the fifth-generation M18 and W20 multifocal stimulus arrays. In practice stimuli of these sizes and shapes are presented pseudo-randomly in time, each for a period of 33 ms. The stimuli shown are for the left eye. Right eye stimuli are the left-right mirror image. The different brightness levels of the stimuli elicit approximately equal-sized responses in a normal person. Each region of the two eyes is tested 44 times in 86 seconds. *A:* The M18 macular stimuli. The cyan lines give the borders of the Early Treatment Diabetic Retinopathy Study retinal thickness grid used by most optical coherence tomography manufacturers. *B:* The much larger W20 stimulus array, which tests a 9-times larger area in the same amount of time.

Returning to the sixth criterion, why are uncorrelated variables desirable and powerful? The lack of correlation between different measures, but good correlation with disease, means the different tests are measuring independent aspects of the disease. This is the statistical basis of evidence-based medicine[37]: different tests characterizing different aspects of the disease are needed to confirm diagnoses and prognoses. If the results of several different tests are correlated there is little benefit to doing more than one of the tests. In statistics this is related to the *likelihood ratio* (LR) for a test: the sensitivity/false-positive-rate. For example, if we chose a false-positive-rate of 5% and the test gives a sensitivity of 95% the LR is 0.95/0.05 = 19, that is, the surety you have of a true result is 1 part in 19. If you have 2 tests that are uncorrelated with each other *the combined LR is the product of the LRs for each*, providing a surety of say 1 part in 19 × 20 = 380.[24] If the 2 tests are perfectly correlated, then the

combined LR is the same as that for either test, that is, 1 part in 19.

In the case above a linear combination of M18 and W20 per-region sensitivities greatly increased diagnostic power. Nonlinear combinations of uncorrelated variables can be even more powerful. Imagine you have a vector of measures from 100 patients using method *a*, and a similar one for method *b*. It turns out the new vector of the element-wise product (Hadamard/Schur product), $d = a \circ b$, can be independent of either *a* or *b* (i.e., somewhat uncorrelated). In other words, the derived vector *d* can provide new independent diagnostic information that could be used to enhance combined diagnostic power. Fig. 3 shows an optimistic example of this. Higher order *kernels* can also be considered like $a \circ b \circ c$, $a \circ b^2 \circ c$, $a \circ c^2 \circ b^2$, and so on.

These sorts of kernel functions and other complex interactions are the basis for much of machine/deep learning.[18,43,47] The downside is that the resulting models are complex to explain[20,39] and so require strict procedures (e.g., resubstitution, hold-out, cross-validation, and bootstrapping) to assure independence of model development and evaluation in training, its possible parameter tuning, and testing.[92] In the field of medical applications of machine learning, recent publications have put a spotlight on the necessity of adherence to these practices to avoid overly optimistic results.[22,30] Typically, these procedures would require datasets with at least 100 subjects—a prerequisite that hinders applications of machine/deep learning in diagnostics.[21] A lack of data can in some cases be alleviated by techniques such as oversampling, but special attention needs to be paid here to avoid data leakage between test and training sets.[94]

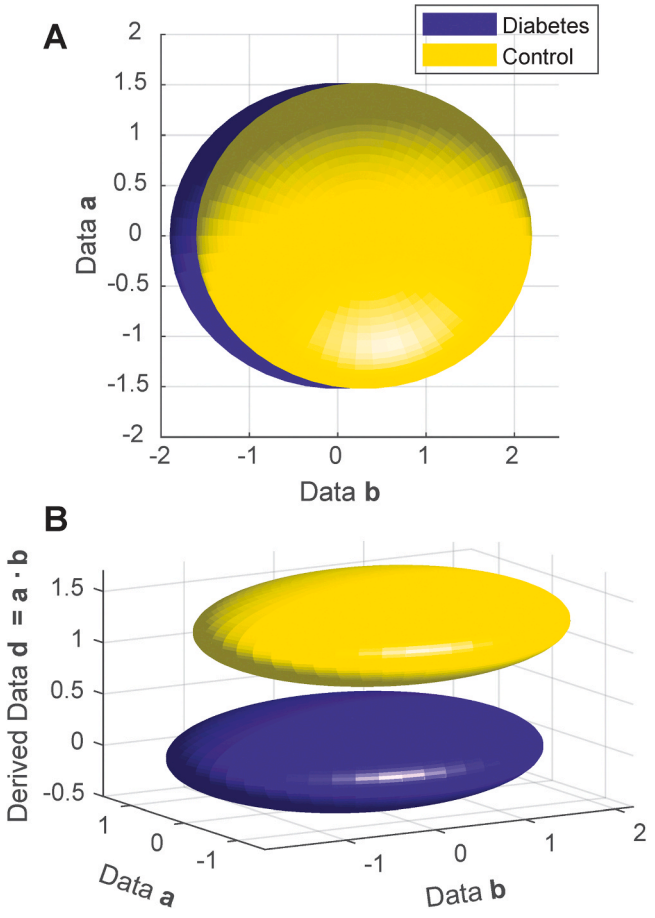

Fig. 3. An example of the possible outcomes of a derived data set $d = a \circ b$. *A:* Contains two elliptical figures the boundaries of which represent something like 1 SD of two clouds of data plots in *a,b* space. Each of *a* and *b* contain data from people with diabetes (blue) and controls (yellow). The distributions are highly over-lapping and so linear combinations of a and b will not be able to discriminate the groups well. *B:* Shows the case of introducing a derived variable $d = a \circ b$. By increasing the dimensionality of the so-called *feature space* the data of the two groups are now well separated.

BCVA and its variants are tried and true clinical tools for monitoring DMO and DR from moderate/severe nonproliferative disease to proliferative disease, but these later disease stages are not where new tests and treatments need to operate. As we know early-stage retinopathy is pointillistic, with microaneurysms and other earlier features appearing at isolated locations across the retina.[23] Therefore the standard ETDRS composite photos use multiple 30°–45° photos to canvas much of the retina for any sign. Visual field testing typically examines a smaller area of ±21° to ±30° radius, that is, 1400°–2800°.[2] Visual acuity is determined by the foveola, which is 0.35 mm across, or 1.1°.[2] Thus, acuity fails to sample the minimally required part of the retina by several thousand times.

Sight-threatening DME in its early stages can be off-axis and therefore fails to alter visual acuity, even though highly significant changes to nearby macular physiology are measurable. For example in early off-axis DME peripheral macular thickness was correlated with central and peripheral OFA sensitivities, although functional change could occur when retinal structure was stable.[73] When tested head-to-head over 457 ± 198 days in DME, OFA sensitivities and delays appear to be more correlated with retinal thickness changes than matrix sensitivities.[83] Similar differential changes between fovea and surrounding macular regions has been reported in neovascular age related macular degeneration (AMD) along with interactions with anti-VEGF treatment using OFA. Thus, changes in peripheral retinal sensitivity and delays can significantly precede the decision to treat on pro re nata (PRN) management by 1–2 months.[72] Similar fovea vs peripheral macular interactions have also been reported for advanced AMD comparing matrix perimeter data and retinal thickness. There BCVA and central OCT thickness (within 4°) were reported to be strongly correlated with peripheral macular sensitivities (4°–10°).[81] Similarly, Quality of Life measures in people with advanced AMD were correlated with global (e.g., pattern standard deviation) but not central matrix sensitivity data (e.g., monocular or binocular central 4°). Thus, it seems that for retinal diseases, more attention needs to be paid to contrast between peripheral and central macular structural and functional changes for diagnosis and prognosis.

At the request of the reviewers, we supply some further information about the OFA. We have mentioned fourth- and fifth-generation OFA methods. We have tested several hundred OFA variants. Many have been published including data on 34 early variants.[10,41] The fifth-generation stimuli of Fig. 2 are the 139th and 140th methods tried on >50 persons, and details of them have been provided.[53,74] The defining features of the five OFA generations are a series of patents on aspects of OFA stimuli. The fourth- and fifth-generation methods include a new patent application, which provides a novel form of multifocal response estimation,[99] and which is part of the OFA analysis software.

An important part of the pupillary system is its dynamic gain control system.[12] It makes responses to single stimuli quite large, but also very noisy. Increasing the number of stimuli presented within a short time period decreases the response size obtained from each tested region. For this reason, the mean interval between OFA stimuli at any location is about 4 seconds. The fourth-generation stimuli test 44 field regions/eye. That means the dichoptic stimuli are delivered at 22/s, allowing each region to be tested 90 times in about 4 min/eye. Environmental contrasts rarely exceed 40% and the visual system is designed to detect rapid changes in those small contrasts.[7] We have found stimuli of about 33 ms duration to be diagnostically the most effective.[41] Nevertheless overly bright brief stimuli can lead to saturating neural responses, which mean damaged and undamaged parts of the field can respond similarly, masking field defects. This caused one of our early studies to have poor diagnostic performance.[50] The response of the pupils to stimuli of the same brightness across the field is exaggerated in the temporal fields.[10] The gain control mechanism means that providing dimmer stimuli to those regions enhances the responses of other regions, so called

luminance balancing.[11] Presenting about equal numbers of stimuli to subsections of the field, so called *clustered volleys*, can also enhance signal-to-noise ratios[13] and diagnostic power.[78]

A limitation of the present study is that the method for calculating AUROC values from standardized effect sizes assumes Gaussian distributions.[76,93] We have found cases where authors reported both the data to calculate *g* and a conventionally derived AUROC, and the agreement has been good. For example, the OCTA study by Zeng and coworkers in Table 4 cited an AUROC of "0.8608" and the result based upon *g* was 0.870 (87%).[104] It is also clear that results from the NC vs No-DR comparisons have not been shown to predict future progression. Later stages of DR are important and need to be assessed with new technologies in cross-sectional and longitudinal studies. Such studies would hopefully inform the application of exiting treatment options like fenofibrate[29,37] or candesartan,[10] as well as new therapeutics. It is worth noting that OFA has demonstrated ability to track change in response to anti-VEGF therapy.[71,77]

9. Conclusions

A growing body of evidence shows that the functional changes precede the structural damage of classical DR. Therefore, we should consider new functional tests with high diagnostic power for discriminating eyes of PwD, but who show no to mild DR. Those tests should be reproducible to aid progression analysis. Ideally multiple measures, for example, sensitivity, delay, asymmetry between eyes, or test variants (e.g., M18, W20), should be available and relatively independent of each other while being well correlated with eye and tissue damage. Such independence can aid things like Machine/Deep Learning to provide more accurate measures with no increase in test duration. Structural measures can be usefully independent but should be assessed at the same retinal locations as the functional measures to aid comparisons and analysis. Finally, tests should be quick and easy to use to allow children to be followed over their lifetimes. The OFA seems to provide these features. We need studies of new tests to report on diagnostic power and other relevant measures.

10. Disclosures

Prof. Ted Maddess and Prof. Hanna Suominen have received grant support from Konan Medical. Prof. Ted Maddess, and Dr. Joshua van Kleef could be paid royalties for the sale of the OFA.

Method of literature search

We searched online (between September, 2022, and July, 2023) for published studies reporting on various methods for discriminating NCs from PwD with No-DR. A few studies comparing controls and prediabetes, or No-DR vs NPDR were included. Studies of more serious DR were excluded. For comparison, we examined both functional and structural tests. Some studies examined several tests or provided independent assessments for several variables per test. Functional tests included behavioral tests, generally involving some visual threshold, electrophysiology, and pupillographic. Structural tests were largely based upon some sort of imaging of the retinal or cornea and could involve forms of retinal angiography.

The main search tool was the Web of Science (WoS, Clarivate, London, UK). Key starting points were known papers and those highlighted by a recent review of methods for more serious DR[16] and WoS was used to examine all papers citing those to discover more recent studies on the same or related topics. Searches for other papers were done containing disease terms including diabetes+eye, diabetes+retinopathy (and substituting diabetic for diabetes), and with vasculopathy and neuropathy were combined with those for functional and structural test methods. Functional terms included retinogram, ERG, evoked+potential, VEP, threshold, perimeter, and perimetry. Structural terms included retinal+photography, fundus+photography, fundus+camera, optical+coherence+tomography, OCT, microscopy, macular+pigment, angiography, and cornea were used. When a suitable reference was found WoS was again used to find papers that cited it.

Declaration of Competing Interest

None.

Acknowledgment

This research was funded in part by and has been delivered in partnership with Our Health in Our Hands (OHIOH), a strategic initiative of the Australian National University, which aims to transform health care by developing new personalized health technologies and solutions in collaboration with patients, clinicians, and health care providers. We also express our gratitude to Konan Medical USA and the Australian Government MRFF Biotechnology Bridge (BTB) Program (research grant number: BTBR100196) for partnering and funding in part this study. We gratefully acknowledge the funding from the ANU School of Computing for Robin Vlieger's PhD studies and JCSMR funding of Dr. van Kleef.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.survophthal.2023.09.002.

References

1 Afrashi F, Erakgun T, Kose S, et al. Blue-on-yellow perimetry versus achromatic perimetry in type 1 diabetes patients without retinopathy. *Diabetes Res Clin Pract.* 2003;61:7–11.
2 Ali EN, Carle CF, Lueck CJ, et al. Assessing migraine patients with multifocal pupillographic objective perimetry. *BMC Neurol.* 2021;21:1–12.
3 Ali EN, Lueck CJ, Martin K, et al. Response characteristics of multifocal objective pupillographic perimetry in persons living with epilepsy. *J Neurol Sci.* 2022;436(1):5.
4 Artes PH, Hutchison DM, Nicolela MT, et al. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Investig Ophthalmol Vis Sci.* 2005;46:2451–2457.
5 Ba-Ali S, Brondsted AE, Andersen HU, et al. Pupillary light responses in type 1 and type 2 diabetics with and without retinopathy. *Acta Ophthalmol.* 2020;98:477–484.
6 Bell A, James AC, Kolic M, et al. Dichoptic multifocal pupillography reveals afferent visual field defects in early type 2 diabetes. *Investig Ophthalmol Vis Sci.* 2010;51:602–608.
7 Bernardete EA, Kaplan E, Knight BW. Contrast gain control in the primate retina: P cells are not X-like, some M cells are. *Vis Neurosci.* 1992;8:483–486.
8 Broe R, Rasmussen ML, Frydkjaer-Olsen U, et al. The 16-year incidence, progression and regression of diabetic retinopathy in a young population-based Danish cohort with type 1 diabetes mellitus: the Danish cohort of pediatric diabetes 1987 (DCPD1987). *Acta Diabetol.* 2014;51:413–420.
9 Cankurtaran V, Ilhan C, Tekin K, et al. Use of automated quantitative pupillometric evaluation for monitoring the severity of diabetic retinopathy. *Arquiv Brasil Oftalmol.* 2020;84:37–44.
10 Carle CF, Maddess T, Kolic M, et al. Contraction anisocoria: segregation, summation and saturation in the pupil light reflex pathway. *Investig Ophthalmol Vis Sci.* 2011;52:2365–2371.
11 Carle CF, James AC, Kolic M, et al. Luminance and colour variant pupil perimetry in glaucoma. *Clin Exp Ophthalmol.* 2014;42:815–824.
12 Carle CF, James AC, Rosli Y, Maddess T. Localization of neuronal gain control in the pupillary response. *Front Neurol.* 2019;10:1–9.
13 Carle CF, James AC, Sabeti F, et al. Clustered volleys stimulus presentation for multifocal objective perimetry. *Transl Vis Sci Technol.* 2022;11(2):1–10.
14 Chao C, Wang R, Jones M, et al. The relationship between corneal nerve density and hemoglobin A1c in patients with prediabetes and type 2 diabetes. *Investig Ophthalmol Vis Sci.* 2020;61:26.
15 Chaturvedi N, Porta M, Klein R, et al. Effect of candesartan on prevention (DIRECT-Prevent 1) and progression (DIRECT-Protect 1) of retinopathy in type 1 diabetes: randomised, placebo-controlled trials. *Lancet.* 2008;372:1394–1402.
16 Chen XD, Gardner TW. A critical review: psychophysical assessments of diabetic retinopathy. *Surv Ophthalmol.* 2021;66:213–230.
17 Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet.* 2010;376:124–136.
18 Cho Y., Saul L. Kernel Methods for Deep Learning: Advances in Neural Information Processing Systems 22 (NIPS 2009). Vancouver, Canada; 2009.
19 Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Routledge; 1988.

20 Cutillo CM, Sharma KR, Foschini L, et al. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit Med*. 2020;3:47.
21 Daskalaki E, Parkinson A, Brew-Sam N, et al. The potential of current noninvasive wearable technology for the monitoring of physiological signals in the management of type 1 diabetes: literature survey. *J Med Internet Res*. 2022;24, e28901.
22 Demircioglu A. Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics. *Insights Imaging*. 2021;12:172.
23 Early treatment diabetic retinopathy study research group: Early Treatment Diabetic Retinopathy Study design and baseline patient characteristics. ETDRS report number 7. *Ophthalmology*. 1991;98:741–756.
24 Egan J. *Signal Detection Theory and ROC Analysis*. New York: Academic Press; 1975.
25 Falavarjani KG, Nguyen QD. Adverse events and complications associated with intravitreal injection of anti-VEGF agents: a review of literature. *Eye*. 2013;27:787–794.
26 Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39:175–191.
27 Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 2009;41:1149–1160.
28 Feigl B, Zele AJ, Fader SM, et al. The post-illumination pupil response of melanopsin-expressing intrinsically photosensitive retinal ganglion cells in diabetes. *Acta Ophthalmol*. 2012;90:e230–234.
29 Gamlin PD. The pretectum: connections and oculomotor-related roles. *Prog Brain Res*. 2006;151:379–405.
30 Ge W, Lueck C, Suominen H, Apthorp D. Has machine learning over-promised in healthcare? A critical analysis and a proposal for improved evaluation, with evidence from Parkinson's disease. *Artif Intell Med*. 2023;139:1–16. https://doi.org/10.1016/j.artmed.2023.102524.
31 Gella L, Nittala MG, Raman R. Retinal sensitivity in healthy Indians using microperimeter. *Indian J Ophthalmol*. 2013;62:284–286.
32 Gella L, Raman R, Kulothungan V, et al. Retinal sensitivity in subjects with type 2 diabetes mellitus: Sankara Nethralaya Diabetic Retinopathy Epidemiology and Molecular Genetics Study (SN-DREAMS II, Report No. 4). *Br J Ophthalmol*. 2016;100:808–813.
33 Grauslund J, Green A, Sjolie AK. Blindness in a 25-year follow-up of a population-based cohort of Danish type 1 diabetic patients. *Ophthalmology*. 2009;116:2170–2174.
34 Grauslund J, Green A, Sjolie AK. Prevalence and 25 year incidence of proliferative retinopathy among Danish type 1 diabetic patients. *Diabetologia*. 2009;52:1829–1835.
35 Group AS, Group AES, Chew EY, et al. Effects of medical therapies on retinopathy progression in type 2 diabetes. *N Engl J Med*. 2010;363:233–244.
36 Halperin A, Pajuelo M, Tornheim JA, et al. Pupillary light reflexes are associated with autonomic dysfunction in bolivian diabetics but not chagas disease patients. *Am J Trop Med Hyg*. 2016;94:1290–1298.
37 Hawkins RC. The evidence based medicine approach to diagnostic testing: practicalities and limitations. *Clin Biochem Rev*. 2005;26:7–18.
38 Hedges LV. Distribution theory for glass's estimator of effect size and related estimators. *J Educ Stat*. 1981;6:107–128.
39 Huang W, Suominen H, Liu T, et al. Explainable discovery of disease biomarkers: the case of ovarian cancer to illustrate the best practice in machine learning and Shapley analysis. *J Biomed Inform*. 2023;141:1–24.
40 Jain M, Devan S, Jaisankar D, et al. Pupillary abnormalities with varying severity of diabetic retinopathy. *Sci Rep*. 2018;8:5636.
41 James AC, Kolic M, Bedford SM, Maddess T. Stimulus parameters for multifocal pupillographic objective perimetry. *J Glaucoma*. 2012;21:571–578.
42 Joltikov KA, de Castro KA, Davila JR, et al. Multidimensional functional and structural evaluation reveals neuroretinal impairment in early diabetic retinopathy. *Investig Ophthalmol Vis Sci*. 2017;58. BIO277-BIO290.
43 Jordan M, Mitchell T. Machine learning: trends, perspectives, and prospects. *Science*. 2015;349:255–260.
44 Keech AC, Mitchell P, Summanen PA, et al. Effect of fenofibrate on the need for laser treatment for diabetic retinopathy (FIELD study): a randomised controlled trial. *Lancet*. 2007;370:1687–1697.
45 Kiziltoprak H, Tekin K, Sekeroglu MA, et al. Static and dynamic pupillary responses in patients with different stages of diabetic retinopathy. *Neuroophthalmology*. 2020;44:226–235.
46 Klein R, Lee KE, Gangnon RE, Klein BE. The 25-year incidence of visual impairment in type 1 diabetes mellitus the Wisconsin epidemiologic study of diabetic retinopathy. *Ophthalmology*. 2010;117:63–70.
47 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
48 Lim LS, Ling LH, Ong PG, et al. Dynamic responses in retinal vessel caliber with flicker light stimulation in eyes with diabetic retinopathy. *Investig Ophthalmol Vis Sci*. 2014;55:5207–5213.
49 Maddess T. Relative Effects of Sampling Errors and Eye Movements Upon SAP Test-Retest Variablity: Imaging and Perimetry Society. Melbourne; 2012.
50 Maddess T, Kolic M, Essex RW, et al. High- versus low-density multifocal pupillographic objective perimetry in glaucoma. *Clin Exp Ophthalmol*. 2013;41:140–147.
51 Maddess T. Modelling the relative influence of fixation and sampling errors on test-retest-variability in perimetry. *Graefes Arch Ophthalmol*. 2014;252:1611–1619. https://doi.org/10.1007/s00417-014-2751-y.
52 Maddess T, van Kleef JP, Kolic M, et al. Comparing Macular and Wide-field Objective Perimetry: World Glaucoma Congress, Vol. ABSUB-824; 2021.
53 Maddess T, van Kleef JP, Rohan EMF, et al. Rapid, non-contact multifocal visual assessment in multiple sclerosis. *Neurol Sci*. 2022;43:1–7.
54 Maddess T, Rohan EMF, Rai BB, et al. Diagnostic power of rapid objective perimetry in young people with type 1 diabetes. *IOVS*. 2023;64.
55 McAnany JJ, Park JC. Temporal frequency abnormalities in early-stage diabetic retinopathy assessed by electroretinography. *Investig Ophthalmol Vis Sci*. 2018;59:4871–4879.
56 McAnany JJ, Park JC, Lim JI. Visual field abnormalities in early-stage diabetic retinopathy assessed by chromatic perimetry. *Investig Ophthalmol Vis Sci*. 2023;64:8.
57 Montesano G, Ometto G, Higgins BE, et al. Evidence for structural and functional damage of the inner retina in diabetes with no diabetic retinopathy. *Investig Ophthalmol Vis Sci*. 2021;62:35.
58 Nathan DM, Group DER. The diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: overview. *Diabetes Care*. 2014;37:9–16.
59 Nitta K, Saito Y, Kobayashi A, Sugiyama K. Influence of clinical factors on blue-on-yellow perimetry for diabetic patients without retinopathy: comparison with white-on-white perimetry. *Retina*. 2006;26:797–802.
60 Numata T, Maddess T, Matsumoto C, et al. Exploring test-retest variability using high-resolution perimetry. *Transl Vis Sci Technol*. 2017;6:1–9.
61 Pahikkala T, Airola A, Suominen H, et al. Efficient AUC Maximization with Regularized Least-squares. In: Holst A, Kreuger P, Funk P, eds. Stockholm, Sweden: IOS Press; 2008:12–19. Tenth Scandinavian Conference on Artificial Intelligence, SCAI 2008; Vol. 173.
62 Park JC, Chen YF, Blair NP, et al. Pupillary responses in non-proliferative diabetic retinopathy. *Sci Rep*. 2017;7:44987.
63 Parravano M, Oddone F, Boccassini B, et al. Functional retinal impairment in type 1 diabetic patients without any signs of retinopathy. *Ophthalmic Res*. 2013;50:108–112.
64 Picconi F, Parravano M, Ylli D, et al. Retinal neurodegeneration in patients with type 1 diabetes mellitus: the role of glycemic variability. *Acta Diabetol*. 2017;54:489–497.
65 Pollreisz A, Desissaire S, Sedova A, et al. Early identification of retinal neuropathy in subclinical diabetic eyes by reduced birefringence of the peripapillary retinal nerve fiber layer. *Investig Ophthalmol Vis Sci*. 2021;62:24.
66 Rai BB, Shresthra MK, Thapa R, et al. Pattern and presentation of vitreo-retinal diseases:an analysis of retrospective data at a tertiary eye care centre in Nepal. *Asia-Pac J Ophthalmol*. 2019;8:481–488.
67 Rai BB, Maddess T, Carle CF, et al. Comparing retinal thickness and Matrix 10-2 functional testing in diabetic macular edema. *Investig Ophthalmol Vis Sci*. 2020;61,4866–4866.
68 Rai BB, Morley MG, Bernstein PS, Maddess T. Pattern of vitreo-retinal diseases at the national referral hospital in Bhutan: a retrospective, hospital-based study. *BMC Ophthalmol*. 2020;20.
69 Rai BB, Morley MG, Zangmo P, et al. Retinal laser services in Bhutan: a 3-year national survey. *BMC Ophthalmol*. 2020;20:404.
70 Rai BB, Morley MG, Zangmo P, et al. Surgical management of vitreo-retinal diseases in Bhutan: a 3-year national study. *New Front Ophthalmol*. 2020;6:1–6.
71 Rai BB, Essex RW, Sabeti F, et al. An objective perimetry study of central versus peripheral sensitivities and delays in age-related macular degeneration. *Transl Vis Sci Technol*. 2021;10:1–14.
72 Rai BB, Essex RW, Sabeti F, et al. An objective perimetry study of central versus peripheral sensitivities and delays in age-related macular degeneration. *Transl Vis Sci Technol*. 2021;10, 24–24.
73 Rai BB, Maddess T, Carle CF, et al. Comparing objective perimetry, matrix perimetry, and regional retinal thickness in mild diabetic macular oedema. *Transl Vis Sci Technol*. 2021;10:1–12.
74 Rai BB, Sabeti F, Carle CF, et al. Rapid objective testing of visual function matched to the ETDRS-grid, and its diagnostic power in AMD. *Ophthalmol Sci*. 2022;2:1–9.
75 Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law Hum Behav*. 2005;29:615–620.
76 Rosenthal R. Defining Research Results: Meta-analytic Procedures for Social Research. Newbury Park, CA; 1991, pp. 13–35.
77 Sabeti F, Maddess T, Essex RW, James AC. Multifocal pupillography identifies ranibizumab induced changes in retinal function for exudative age-related macular degeneration. *Investig Ophthalmol Vis Sci*. 2012;53:253–260.
78 Sabeti F, Maddess T, Saikal A, et al. Multifocal pupillography in early age-related macular degeneration. *Opt Vis Sci*. 2014;91:904–915.
79 Sabeti F, Nolan C, Essex R, et al. Multifocal pupillography identifies changes in visual sensitivity according to severity of diabetic retinopathy in type 2 diabetes. *Investig Ophthalmol Vis Sci*. 2015;56:4504–4513.
80 Sabeti F, James AC, Carle CF, et al. Comparing multifocal pupillographic objective perimetry (mfPOP) and multifocal visual evoked potentials (mfVEP) in retinal diseases. *Sci Rep*. 2017;7:45847.
81 Sabeti F, Lane J, Rohan EMF, et al. Correlation of central versus peripheral macular structure-function with acuity in age-related macular degeneration. *TVST*. 2021;10:1–12.
82 Sabeti F, Carle CF, Nolan C, et al. Multifocal pupillographic objective perimetry for assessment of early diabetic retinopathy and generalised diabetes-related tissue injury in persons with type 1 diabetes. *BMC Ophthalmol*. 2022;22:1–13.
83 Sabeti F, Rai BB, van Kleef JP, et al. Objective perimetry identifies functional progression and recovery in mild diabetic macular oedema. *PLoS One*. 2023;18:1–20.
84 Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the

International Diabetes Federation Diabetes Atlas, 9(th) edition. *Diabetes Res Clin Pract*. 2019;157, 107843.
85 Schmidt AM. Highlighting diabetes mellitus: the epidemic continues. *Arterioscler Thromb Vasc Biol*. 2018;38:e1–e8.
86 Shelmerdine SC, Arthurs OJ, Denniston A, Sebire NJ. Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. *BMJ Health Care Inform*. 2021;28.
87 Simó R, Hernández C. Neurodegeneration in the diabetic eye: new insights and therapeutic perspectives. *Trends Endocrinol Metab*. 2014;25:23–33.
88 Sivaprasad S, Elagouz M, McHugh D, et al. Micropulsed diode laser therapy: evolution and clinical applications. *Surv Ophthalmol*. 2010;55:516–530.
89 Song SH. Significant retinopathy in young-onset type 2 vs. type 1 diabetes: a clinical observation. *Int J Clin Pract*. 2016;70:853–860.
90 Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. *J Grad Med Educ*. 2012;4:279–282.
91 Sung JY, Lee MW, Lim HB, et al. The ganglion cell-inner plexiform layer thickness/vessel density of superficial vascular plexus ratio according to the progression of diabetic retinopathy. *Investig Ophthalmol Vis Sci*. 2022;63:4.
92 Suominen H., Pahikkala T., Salakoski T. Critical Points in Assessing Learning Performance Via Cross-Validation. In: 2nd International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning. Espoo, Finland; 2008, pp. 9–22.
93 Swets JA. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychol Bull*. 1986;99:100–117.
94 Tampu IE, Eklund A, Haj-Hosseini N. Inflation of test accuracy due to data leakage in deep learning-based classification of OCT images. *Sci Data*. 2022;9:580.
95 Tan TE, Finkelstein MT, Tan GSW, et al. Retinal neural dysfunction in diabetes revealed with handheld chromatic pupillometry. *Clin Exp Ophthalmol*. 2022;50:745–756.
96 Teo ZL, Tham YC, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology*. 2021. Patent No.: P0040304AU.
97 Tesfaye S, Boulton AJ, Dyck PJ, et al. Diabetic neuropathies: update on definitions, diagnostic criteria, estimation of severity, and treatments. *Diabetes Care*. 2010;33:2285–2293.
98 Thapa R, Bajimaya S, Sharma S, et al. Systemic association of newly diagnosed proliferative diabetic retinopathy among type 2 diabetes patients presented at a tertiary eye hospital of Nepal. *Nepal J Ophthalmol*. 2015;7:26–32.
99 van Kleef JP, Carle CF, Maddess T. Response Estimation to Efficiently Capture Dynamic Response Gain Changes in Multifocal Responses. Australia; 2022, pp. 1–53.
100 van Wijngaarden P, Coster DJ, Williams KA. Inhibitors of ocular neovascularization: promises and potential problems. *JAMA*. 2005;293:1509–1513.
101 Wang W, Lo ACY. Diabetic retinopathy: pathophysiology and treatments. *Int J Mol Sci*. 2018;19.
102 Yuan D, Spaeth EB, Vernino S, Muppidi S. Disproportionate pupillary involvement in diabetic autonomic neuropathy. *Clin Auton Res*. 2014;24:305–309.
103 Zeng Y, Cao D, Yu H, et al. Early retinal neurovascular impairment in patients with diabetes without clinically detectable retinopathy. *Br J Ophthalmol*. 2019;103:1747–1752.
104 Zeng Y, Liu M, Li M, et al. Early changes to retinal structure in patients with diabetic retinopathy as determined by ultrawide swept-source optical coherence tomography-angiography. *Front Endocrinol*. 2023;14, 1143535.