



**Voorlopig STANDPUNT** van de WERKGROEP MEDISCHE RETINA van het NEDERLANDS OOGHEELKUNDIG GEZELSCHAP

Onderwerp: **Geautomatiseerde screening op DR met klasse IIa EC-geregistreerde medical devices**

Auteur: Mw. Dr. Y. de Jong-Hesse, namens het bestuur van de Werkgroep MR

Belangenverstrengeling:

Dr. Y. de Jong-Hesse: geen

Prof. R.O. Schlingemann: geen (adviseur IDx tot 2019)

Prof C.J.F. Boon: geen

Prof C.B. Hoyng: geen

Dr. J. van Lith-Verhoeven: geen

Dr. S. Yzer: geen

Prof C.C.W. Klaver: geen

### **Inleiding**

In de laatste jaren is een aantal algoritmen voor geautomatiseerde screening op diabetische retinopathie op de markt gebracht na het verkrijgen van een Klasse I of Klasse IIa certificering door de EMA\*. Door de EMA als klasse IIa EC-geregistreerde medical devices moeten voldoen aan strenge eisen, waaronder een kwaliteit managementsysteem (1). Een klasse IIa device kan direct en autonoom een diagnose stellen en een advies voor verwijzing geven zonder controle door de zorgaanbieder. Klasse I CE medical devices kunnen niet autonoom een diagnose stellen maar wel als ondersteunend middel voor graders of oogartsen worden ingezet.

Deze screening algoritmen zijn gevalideerd met retrospectieve datasets of met prospectieve klinische cohorten (2,3), wat onderlinge vergelijking moeilijk maakt. De performance wordt weergegeven in 'diagnosability', dat is het percentage patiënten dat door het algoritme beoordeeld kan worden, en de sensitiviteit en specificiteit in het opsporen van 'referable' diabetische retinopathie (rfDR) ten opzichte van een per studie sterk wisselende 'gouden standaard'. Deze referentie standaard is soms een

classificatie door humane graders of door een 'expert' panel met een wisselende sensitiviteit, en in andere gevallen een uitgebreide set metingen met een hoge sensitiviteit. Dit maakt de gerapporteerde performance van verschillende algoritmen met de beschikbare literatuur moeilijk vergelijkbaar.

In het algemeen is rfDR gedefinieerd als meer dan milde niet proliferatieve DR (mtmNPDR) en/of kenmerken van diabetisch macula oedeem (DME), gebruik makend van de ICDR classificatie of de Scanlon classificatie (4,5). De algoritmen bereiken een sensitiviteit tussen 80-100% en een specificiteit tussen 60-100% voor rfDR (2,3), met in acht neming van het bovengenoemde voorbehoud dat de kwaliteit van de onderliggende data sterk verschilt.

\* Voor klasse IIa Medical Devices is een periodiek Quality Management System verplicht, zie ook: <https://support.ce-check.eu/hc/en-us/articles/360008712879-Classification-Of-Medical-Devices-And-Their-Routes-To-CE-Marking>

### **Klasse IIa EC-geregistreerde autonome medical devices**

Door de EMA als klasse IIa EC-geregistreerde medical devices moeten voldoen aan strenge eisen, waaronder een kwaliteit managementsysteem(1). Een klasse IIa device ('Autonomous medical device') kan direct en autonoom een diagnose stellen en een advies voor verwijzing geven zonder controle door de zorgaanbieder, waarbij de verantwoordelijkheid bij de producent van het device ligt, en niet bij de dokter.

Recent zijn meerdere klasse IIa EC-geregistreerde medical devices\* als 'Autonomous Medical Devices' door de EMA gecertificeerd, namelijk IDx-DR® (Digital diagnostics), EyeART® (EyeNuk), Retmarker® en RetCAD® (Thirona), waarvan er twee ook een FDA registratie hebben (IDx-DR® en EyeArt®).

IDx-DR en EyeArt worden hieronder besproken. RetCAD® bereikte in een retrospectieve studie met routine foto's een sensitiviteit/specificiteit van 90,1%/90,6% bij het opsporen van referable DR (6). Voor Retmarker® was dat 85%/53% in een retrospectieve vergelijkende real-life studie (7,8). De bruikbaarheid van de output van deze twee algoritmen in de klinische setting blijft onduidelijk. Zo geeft b.v. RetCAD® alleen aan of er tekenen van DR gevonden zijn met een waarschijnlijkheid score, zonder een specifieke DR-classificatie. De uitslag van Retmarker® lijkt een arbitraire score van gevonden afwijkingen te zijn.

### **'Autonomous Medical Devices' met FDA goedkeuring**

Twee systemen zijn inmiddels door de FDA goedgekeurd als 'autonomous device': IDx-DR in 2018 en Eyeart® in 2020. In de 'real life' registratie studies van deze twee devices was de referentie standard strenger dan in de andere genoemde studies.

De registratiestudie voor IDx-DR was een prospectieve observationele studie met als referentie standard een combinatie van beoordeling van 4-velden stereo fundusfoto's (equivalent aan 7-velden stereo volgens ETDR protocol) en OCT door professionele graders van het Wisconsin Reading center (9). De sensitiviteit/specificiteit voor detectie van more-than-mild DR was 87,2% en 90,7% en de sensitiviteit voor visus bedreigende DR (ETDRS level >53 en/of diabetisch maculaedeem) was 97,4%. De 'diagnosability' was 96.1%, wat betekent dat bij 96,1% van de foto's een gradering door de device mogelijk was. 23% van de patiënten had daarvoor mydriasis nodig. In eerdere retro- en prospectieve studies met IDx-DR lag de sensitiviteit voor rfDR tussen 68-100% en de specificiteit tussen 81-86% (10-12), maar in deze studies werd geen gebruik gemaakt van de ingebouwde beeld kwaliteit feedback die

IDx-DR ook levert. In de in de VS beschikbare versie maakt IDx-DR onderscheid tussen geen referable DR, more-than-mild DR, diabetic macular edema en visus bedreigende DR. De versie voor de Europese markt kan daarnaast ook onderscheid maken tussen geen DR en milde DR in een of beide ogen.

EyeArt® is in 2020 goedgekeurd op basis van een prospectieve klinische multicenter studie, die nog niet gepubliceerd is (<https://clinicaltrials.gov/ct2/show/NCT03112005>; [https://www.accessdata.fda.gov/cdrh\\_docs/pdf20/K200667.pdf](https://www.accessdata.fda.gov/cdrh_docs/pdf20/K200667.pdf)). De referentiestandaard was 4-velden stereo fundusfoto's beoordeeld door expert graders van het Fundus Photography Reading Center. De 'diagnosability' was 96,5%. De sensitiviteit/specifiteit voor meer dan milde DR was 96%/88% en voor visus-bedreigende DR (ETDRS level >53 en/of diabetisch maculaoedeem) 92%/84%. In een aantal andere retro- en prospectieve klinische studies met EyeArt was de gerapporteerde sensitiviteit voor detectie van referable DR tussen 90-95,7% en voor visus bedreigende DR 98-99% (7,8,13-16). De specificiteit lag tussen 20 en 91% (7,8,13-16). In sommige studies werd gebruik gemaakt van de ICDR-classificatie voor DR, in andere van de Scanlon-classificatie (2,3) (zie tabel).

EyeArt kan ook per oog de graad van diabetische retinopathie bepalen en de aan- of afwezigheid van diabetisch maculaoedeem, maar deze performance per oog is niet gerapporteerd in de literatuur. Het kan zo geconfigureerd worden dat het gebruikt maakt van de ICDR classificatie of van de NHS DESP Scanlon classificatie.

### **Algoritmen versus humane graders**

Hoe verhouden de prestaties van algoritmes zich tot de standaard 'manual grading'? Dit is in een aantal retro- en prospectieve validatiestudies onderzocht. Hierbij werd de performance van een aantal algoritmes, die overigens niet tot de eerder genoemde in Europa gecertificeerde systemen behoren, vergeleken met humane graders en met een 'expert' panel als 'gouden standard' (17-19). De betreffende algoritmen bereikten een duidelijk hogere sensitiviteit (89-97%) en specificiteit (91-96%) voor detectie van referable DR dan human graders (sensitiviteit 73-84% en specificiteit 84-98%) (17-19) (zie tabel 1 voor meer details). De enige prospectieve studie in dit verband werd uitgevoerd in 2 centra. In centrum 1 bereikte de getrainde grader een sensitiviteit en specificiteit van 75,5% en 94,2% voor het opsporen van moderate DR. Het algoritme in hetzelfde centrum bereikte een sensitiviteit/specifiteit van 88,9%/ 92,2% (18). In centrum 2 waren sensitiviteit/specifiteit van de grader 84,2%/98,6% en van het algoritme 92,1%/95,2% (18).

### **Kosteneffectiviteit**

Op dit moment is de kosteneffectiviteit van geautomatiseerde screening ten opzichte van manuele screening door graders weinig onderzocht.

In de studie van Tufail (7) uit de U.K werden twee strategieën op kosteneffectiviteit onderzocht: 1. De geautomatiseerde screening als vervanging van manuele screening door level 1 graders. 2. Geautomatiseerde screening als filter voorafgaand aan screening door level 1 graders. EyeART and Retmarker bespaarden kosten in beide scenario's, waarbij de 2<sup>de</sup> strategie minder kosten-effectief was (7). In een andere studie uit de V.S. werden middels een model-gebaseerde kosteneffectiviteits-analyse de kosten voor AI gebaseerde screening met de EyeArt als prescreening vergeleken met standaard screening middels jaarlijkse funduscopie. De geautomatiseerde screening bleek even effectief maar reduceerde de kosten met 23% (20). IDx-DR is in V.S. op kosteneffectiviteit onderzocht bij screening van kinderen (<21j) met DM type 1 of 2 ten opzichte van jaarlijkse screening middels funduscopie (21). Als >23% van deze populatie gebruikt maakt van de screening dan is screening met

een autonoom systeem kosten effectiever dan screening door oogartsen (21). Aangezien screening programma's tussen landen zeer verschillen, zal de kosteneffectiviteit van geautomatiseerde screening per land verschillen. De kosteneffectiviteit van geautomatiseerde screening in Nederland moet nog onderzocht worden.

### **DR classificaties**

Naast het probleem van de verschillen in per studie gehanteerde 'gouden standaard' zijn bij vergelijking van algoritmen ook de gehanteerde grading/classification scales bij de geleverde output van belang. IDx-DR gebruikt de ICDR classificatie, waarbij 'more than mild DR' bijvoorbeeld overeenkomt met ETDRS level >35 (9-12). Bij IDx-DR ligt door gebruik van een "strengere" grading scale (ICDR) de drempel voor rfDR iets lager dan in de Nederlandse richtlijn die gebruikt maakt van de Scanlon-classificatie, waarin milde DR (R1) overeenkomt met ETDRS level 20-43.

Hoewel conform de Richtlijn DM op dit moment bij milde DR in **één** oog de screening herhaald dient worden na een jaar, maar bij milde DR in **twee** ogen verwijzing naar de oogarts is geïndiceerd, lijkt bij gebruik van de IDx-DR en EyeArt van wege de hoge sensitiviteit en 'diagnosability' de kans op het ontwikkelen van visusbedreigende DR verwaarloosbaar klein, ook in de gevallen waarbij milde DR in 2 ogen wordt geconstateerd op basis van de ICDR classificatie.

Volgens de Nederlandse richtlijn DR kan bij RO (geen DR) het screening-interval verlengd worden naar 2 of 3 jaar. IDx-DR en de EyeART kunnen patiënten zonder DR identificeren, wat verlenging van het screeningsinterval met deze algoritmen ook mogelijk maakt.

### **De werkgroep Medische Retina van het Nederland Oogheelkundig Gezelschap concludeert op basis van bovenstaande dat:**

- 1. geautomatiseerde screening op DR met klasse IIa EC-geregistreerde medical devices\* even goed of beter is dan screening door human graders, en een acceptabel alternatief daarvoor is in de huisartsen- of oogartsenpraktijk.**
- 2. er op dit moment een voorkeur uitgaat naar het gebruik van IDx-DR of EyeArt omdat alleen deze twee klasse IIa medical devices additioneel ook een FDA goedkeuring hebben gebaseerd op een objectieve referentiestandaard van hoge kwaliteit in 'state of the art' en 'real life' registratiestudies, en omdat de geleverde output goed aansluit bij de klinische praktijk.**
- 3. bij gebruik van geautomatiseerde screening met klasse IIa devices die zijn gebaseerd op de ICDR classificatie, kan het onderscheid in vervolg- en verwijsbeleid bij milde DR in een of twee ogen vervallen, en kan ook bij milde DR in twee ogen de patiënt een jaar later opnieuw met een klasse IIa device gescreend worden.**

### **Disclaimer:**

- Dit standpunt is opgesteld op basis van de bij de auteur beschikbare informatie en kan worden aangepast n.a.v. nieuwe inzichten.
- Dit standpunt kan worden gebruikt om beleid op te baseren, maar is geen richtlijn.
- Er is de grootst mogelijke zorg besteed aan de inhoud van dit standpunt. Desondanks accepteert het NOG geen aansprakelijkheid voor eventuele onjuistheden in dit document,

voor enigerlei schade of voor andersoortige gevolgen die voortvloeien uit of samenhangen met het gebruik van dit standpunt.

## Referenties

- 1 [https://ec.europa.eu/health/sites/health/files/md\\_topics-interest/docs/md\\_meddev-guidance-216\\_en.pdf](https://ec.europa.eu/health/sites/health/files/md_topics-interest/docs/md_meddev-guidance-216_en.pdf)
- 2 Nielsen, K.B. et al. (2019) Deep Learning-Based Algorithms in Screening of Diabetic Retinopathy: A Systematic Review of Diagnostic Performance. *Ophthalmology Retina*; 3:294-304
- 3 Raman, R. et al. (2019) Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. *Eye* 33:97-109
- 4 Wu L, Fernandez-Loaiza P, Sauma J, et al. Classification of diabetic retinopathy and diabetic macular edema. *World J Diabetes* 2013;4:290–4
- 5 Scanlon, PH. 2019. Update on Screening for Sight-Threatening Diabetic Retinopathy. *Ophthalmic Res.* 2019;62(4):218-224.
- 6 Gonzalez-Gonzalo C et al. (2020) Evaluation of a deep learning system for the joint automated detection of diabetic retinopathy and age-related macular degeneration. *Acta Ophthalmol*:98:368-377
- 7 Tufail, A. (2017) Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared With Human Graders. *Ophthalmology* 124:343-351
- 8 Tufail, A et al. (2016) An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technol Assess.* 20(92):1-72.doi:103310/hta20920
- 9 Abramoff, M.D. et al. (2018) Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine* volume 1, article number 39
- 10 Shah, A et al. (2020) Validation of automated screening for referable diabetic retinopathy with an autonomous diagnostic artificial intelligence system in a Spanish population. *J Diabetes Science and Technology* DOI:10.1177/19322968206212
- 11 Verbraak, F.D. et al. (2019) Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. *Diabetes Care* 42:651-656
- 12 van de Heijden, A.A. et al. (2018) Validation of automated screening for referable diabetic retinopathy with the IDx-DR Device in the Hoorn diabetes care system. *Acta Ophthalmol.* 96(1):63-68
- 13 Bhaskaranand, M. et al. (2016) Automated Diabetic Retinopathy Screening and Monitoring Using Retinal Fundus Image Analysis. *J Diabetes Sci Technol.* 16: 254-61
- 14 Heydon, P et al. (2020) Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30000 patients. *Br J Ophthalmol* Epub ahead of print. doi:10.1136/bjophthalmol-2020-316594

- 15 Bhaskaranand, M. et al. (2019) The value of automated diabetic retinopathy screening with the EyeArt system: A study of more the 100000 consecutive encounters from people with diabetes. *Diabetes Technology & Therapeutics*, (21), DOI:10.1089/dia.2019.0164
- 16 Olvera-Barrios, A et al. (2020) Diagnostic accuracy of diabetic retinopathy grading by an artificial intelligence-enabled algorithm compared with human standard for wide-field true-colour confocal scanning and standard digital retinal images. *Br J Ophthalmol* (Epub ahead of print), doi:10.1136/bjophthalmol-2019-315394
- 17 Krause J. et al. (2018) Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*. 125:1264-1272
- 18 Gulshan V. et al. (2019) Performance of a Deep-Learning Algorithm vs Manual Grading for Detecting Diabetic Retinopathy in India. *JAMA Ophthalmol*. 137:987-993
- 19 Raumviboonsuk P. et al. (2019) Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *NPJ Digit. Med*. 2:25. doi: 10.1038/s41746-019-0099-8
- 20 Fuller S.D. et al.(2020) Five-Year Cost-Effectiveness Modeling of Primary Care-Based, Nonmydriatic Automated Retinal Image Analysis Screening Among Low-Income Patients with Diabetes.” *Journal of Diabetes Science and Technology*,, 193229682096701
- 21 Wolf R.M. et al. (2020) Cost-effectiveness of Autonomous Point-of-Care Diabetic Retinopathy Screening for Pediatric Patients With Diabetes. *JAMA Ophthalmol*.;138(10):1063-1069. Doi

Tabel 1

study	opzet	DRP scale	Reference standard	Evidence label	results	algorithm
6 Gonzalez-Gonzalo C 2020 <i>Acta Ophthalmol</i>	600 routine images Messidor dataset (1200) AREDS (AMD)  Comparison with observers	ICDR Non-rfDRP: stage 0 en 1 rfDR: stage 2,3,4	One experienced ophthalmologist	3	Sensitivity: rfDR: 90,1% specificity: 90,6% human perf: sens: 61,5% spec: 97,8%	RetCAD v.1.3.0. (Thirona) Joint detection DR and AMD
7 Tufail, A. (2017) <i>Ophthalmology</i> (= study 11)	Observational retrospective Real-life Consecutive patients routine screening N=20258	Scanlon rfDRP (R2, R3, M1)	Standard national screening protocol, reading center for disagreements	3	EyeArt: Sensitivity rfDR: 93,8% PDR 99,6% Spec R0:20% Retmarker: Sensitivity: rfDR:85% PDR: 97,9% Spec R0:53% iGradingM (rfDR): sens: 100% spec 0%	3 systems iGradingM Retmarker v 0.8.2 EyeArt

8 Tufail, A (2016) Health Technol Assess.	Observational retrospective N=20258	Scanlon rfDRP (R2, R3, M1)	Final manual grade	3	EyeArt: Sensitivity rfDR: 93,8% PDR 99,6% Spec RO:20% Retmarker: Sensitivity: rfDR:85% PDR: 97,9% Spec RO:53%	Retmarker version 0.8.2  EyeArt
9 Abramoff, MD (2018) npj Digital Medicine FDA registration study	Prospective observational N=900 Primary care centers	ICDR mtmDR: ≥35 ETDRS +/- ME vtDR: ≥53 ETDRS +/- ME	Wisconsin reading center: center and 7- fields+ OCT	1	mtmDR: sensi:87,2% Spec: 90,7% vtDR: sens: 97,4%	IDx-DR
10 Shah, A. et al. J Diabetes Science and Technology (2020)	Prospective real-life Multicenter (DR-screening program in Valencia, Spain)	ICDR Referable DR:mDR of vtDR vtDR: ≥53 mDR:ETDRS 35-47	Grading by 3 ophthalmol ogists	2-3	Sensi/Spec rfDR:100%/ 81,82%  vtDR:100%/ 94,64%	IDx-DR v2
11 Verbraak, F.D. (2019) Diabetes Care	Retrospective Primary care screening center N=1616	ICDR No/mild DR mtmDR vtDR	Rotterdam Study Reading center (2 graders)	3	Sens/spec vtDR: 100%/97,8% Sens/spec mtmDR: 79,4%/93,8%	IDx-DR- EU-2.1
12 van de Heijden, A.A. (2018) Acta Ophthalmol.	Retrospective observational, Hoorn Diabetes primary care system N=1415	ICDR and EURODIAB modDR = ICDR level2 vtDR≥ ICDR level 3	3 retinale specialists	3	Sens/spec rfDR: 68%/86% Sens/spec vtDR: 62%/95%	IDx-DR 2.0
13 Bhaskaranand, M. (2016) J Diabetes Sci Technol	EyePACS telescreening system (360 primary care clinics) Screening data set N=5084	ICDR No refer: 0,1, en no MO rfDRP: ≥2 or MO	Human expert	3	rfDR: sens: 90% spec 63,2% sens vtDR: 96,8%	EyeArt v1.2
14 Heydon, P 2020 Br J Ophthalmol	Prospective, multicenter N=30405 screening episodes (NHS DESP)	Scanlon R1:ETDRS 20-35 R2: ETDRS 43-53	Human grading following standard national (U.K)	2	Sensitivity: rfDRP:95,7% R1M1: 98,3% R2: 100% R3:100% R1M0: 89,1%	EyeArt v.2.1.0

		R3: ETDRS 61+	protocol (DESP)		ROM0 (spec) 68%	
15 Bhaskaranand, M 2019 Diabet Technog&Ther	Retrospective real-world Multicenter (primary care visits) N=107001	ICDR Non-ref (no or mild DR) rfDR: $\geq$ moderate DR +/- MO	Eye-PACS DR telescreening Certified graders	3	Sensitivity Ref DR: 91,3% Specificity: nonrfDR: 91,1% Sens Treat DR (mod of M1): 98,5%	EyeArt v2.0
16 Olvera-Barrios,A 2020 Br J Ophthalmol	Cross-sectional, multicenter N=1257 patients form screening program	Scanlon: R0, R1, R2, R3, M0, M1 vtDR: R2,R3,M1	Human grade of standard NDESP images	2-3	Sensitivity anyDRP 92,27% vtDR:99% PDR: 100% Specificity: ROM0:74%	EyeArt V.2.1.0
17 Krause, J (2018) Ophthalmology	Retrospective analyses 3 datasets: EyePACS, Messidor-2 Validation n=999	ICDR	Adjudicated consensus of retinal specialists	3	modDR or worse (rDR): Sens 97,0% Spec 91,7% Ophthalmologists: Sens 83,8% Spec 98,1%	google
18 Gulshan, V (2019) JAMA Ophthalmol	Prospective observational clinical validation N=3049 from eye care centers in India Algorithm vs manual grading (2 centers)	ICDR	Consensus of 3 senior retinal specialists	2	$\geq$ mod DR (rfDR): Algorithm: Sensitivieit center1/center2 88,9/92,1% Spec center1/2 92,2/95,2% Trained Grader: Sens center 1/2 75,7% /84,2% Spec center 1/2 94,2%/98,6%	google
19 Raumviboonsuk P. (2019) NJP digital Medicine	Retrospectieve clinical validation, n=25326 DR screening progam in Thailand; algorithm vs graders of national screening progam	ICDR	Consensus of pannel of international retinal specialists	3	$\geq$ mod DR (rfDR): Algorithm: Sens 97% Spec 96% Graders: Sens 74% Spec 98%	google