



ARCHIVOS DE LA SOCIEDAD ESPAÑOLA DE OFTALMOLOGÍA

www.elsevier.es/oftalmologia



Artículo original

Uso en la práctica clínica, de un método de cribado automatizado de retinopatía diabética derivable mediante un sistema de inteligencia artificial de diagnóstico



Cristina Peris-Martínez^{a,b,*}, Abhay Shaha^{a,b,c,d,e,f,g,h,i,j}, Warren Clarida^c, Ryan Amelon^c, María C. Hernández-Ortega^d, Amparo Navea^{e,f}, Jesús Morales-Olivas^a, Rosa Dolz-Marco^g, Pablo Pérez-Jordá^a, Frank Verbraak^h y Amber A van der Heijden^{i,j}

^a FISABIO Oftalmología Médica (FOM), Valencia, España

^b Universidad de Valencia, Valencia, España

^c IDx Technologies Inc., Coralville, United States of America

^d European Innovative Biomedicine Institute (EIBI), Castro-urdiales, España

^e Instituto de la retina, Valencia, España

^f Universidad Cardenal Herrera CEU, Valencia, España

^g Oftalvist, Valencia, España

^h Departamento de Oftalmología, VUmc, Centros Médicos de la Universidad de Ámsterdam, Ámsterdam, Países Bajos

ⁱ Departamento de Medicina General y Geriátrica, Centro Médico de la Universidad VU, Ámsterdam, Países Bajos

^j Instituto de Investigación en Salud Pública de Ámsterdam, Centro Médico de la Universidad VU, Ámsterdam, Países Bajos

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 27 de abril de 2020

Aceptado el 3 de agosto de 2020

On-line el 3 de noviembre de 2020

Palabras clave:

Retinopatía diabética

Inteligencia artificial

Cribado automatizado

R E S U M E N

Antecedentes y objetivo: Comparar el rendimiento diagnóstico de un sistema de inteligencia artificial (IA) de diagnóstico autónomo para el diagnóstico de retinopatía diabética derivable (RDR) con la clasificación manual.

Materiales y métodos: Sujetos con diabetes tipo 1 y 2 participaron en un programa de cribado de retinopatía diabética (RD) entre 2011-2012. Se recogieron dos imágenes de cada ojo. Se obtuvieron imágenes retinianas no identificables, una centrada en el disco y otra en la fóvea. Los exámenes se clasificaron con el sistema de IA autónomo y manualmente por parte de oftalmólogos anónimos. Los resultados del sistema de IA y de la clasificación manual se compararon en cuanto a sensibilidad y especificidad para el diagnóstico tanto de RDR como de retinopatía diabética con disminución de la visión (VTDR).

Resultados: Se incluyeron 10.257 retinografías de 5.360 ojos de 2.680 sujetos. De acuerdo con la clasificación manual, la prevalencia de RDR fue de 4,14% y la de VTDR de 2,57%. El sistema de IA registró un 100% (95% IC: 97-100%) de sensibilidad y un 81,82% (95% IC: 80-83%) de especificidad para RDR, y un 100% (95% IC: 95-100%) de sensibilidad y un 94,64% (95% IC: 94-95%) de especificidad para VTDR.

* Autor para correspondencia.

Correo electrónico: cristinaperismartinez0@gmail.com (C. Peris-Martínez).

<https://doi.org/10.1016/j.oftal.2020.08.007>

0365-6691/© 2020 Sociedad Española de Oftalmología. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Conclusiones: En comparación con la clasificación manual, el sistema de IA de diagnóstico autónomo registró una sensibilidad (100%) y una especificidad (82%) elevadas en el diagnóstico de RDR y edema macular en personas con diabetes. Debido a su diagnóstico inmediato, el sistema de IA de diagnóstico autónomo puede aumentar la accesibilidad al cribado de RDR en contextos de atención primaria.

© 2020 Sociedad Española de Oftalmología. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Use in clinical practice of an automated screening method of diabetic retinopathy that can be derived using a diagnostic artificial intelligence system

A B S T R A C T

Keywords:

Diabetic retinopathy
Artificial intelligence
Automated screening

Background and objective: To compare the diagnostic performance of an autonomous diagnostic artificial intelligence (AI) system for the diagnosis of derivable diabetic retinopathy (RDR) with manual classification.

Materials and methods: Patients with type 1 and type 2 diabetes participated in a diabetic retinopathy (DR) screening program between 2011-2012. 2 images of each eye were collected. Unidentifiable retinal images were obtained, one centered on the disc and one on the fovea. The exams were classified with the autonomous AI system and manually by anonymous ophthalmologists. The results of the AI system and manual classification were compared in terms of sensitivity and specificity for the diagnosis of both (RDR) and diabetic retinopathy with decreased vision (VTDR).

Results: 10,257 retinal images of 5,630 eyes of 2,680 subjects were included. According to the manual classification, the prevalence of RDR was 4.14% and that of VTDR 2.57%. The AI system recorded 100% (95% CI: 97-100%) sensitivity and 81.82% (95% CI: 80-83%) specificity for RDR, and 100% (95% CI: 95-100%) of sensitivity and 94.64% (95% CI: 94-95%) of specificity for VTDR.

Conclusions: Compared to the manual classification, the autonomous diagnostic AI system registered a high sensitivity (100%) and specificity (82%) in the diagnosis of RDR and macular edema in people with diabetes. Due to its immediate diagnosis, the autonomous diagnostic AI system can increase the accessibility of RDR screening in primary care settings.

© 2020 Sociedad Española de Oftalmología. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

A pesar de los avances en la atención de los pacientes diabéticos, la retinopatía diabética (RD) continúa siendo una de las principales causas de ceguera en España y la principal causa de pérdida de visión en personas en edad laboral¹. La prevalencia de RD en personas diabéticas oscila entre el 5 y el 30%²⁻⁶. El cribado de la RD ha demostrado su efectividad en la prevención de la ceguera, y muchas organizaciones profesionales han adoptado recomendaciones para un cribado de la RD realizado de manera regulada⁷⁻⁹.

Los sistemas de inteligencia artificial (IA) para la clasificación y el cribado de la RD pueden aumentar la accesibilidad al cribado de la RD de personas diabéticas y la eficiencia, la productividad, la reproducibilidad y la precisión del diagnóstico, así como reducir el tiempo que se tarda en conseguir el acceso y mejorar los resultados de los pacientes. Se han desarrollado numerosos algoritmos académicos de IA para el cribado de RD¹⁰⁻¹⁵ que han registrado una amplia

variedad de rendimientos en conjuntos de validación de calidad y patrón de referencia diferentes. IDx-DR[®] v2 (IDx Technologies Inc, Coralville, Iowa, EE. UU.) es un sistema de IA de diagnóstico autónomo comercializado para identificar pacientes con retinopatía diabética derivable (RDR) analizando imágenes retinianas para detectar signos de RDR y edema macular diabético (EMD) sin supervisión humana o de un médico¹⁶. El sistema de IA cuenta con un marcado de conformidad europea (marcado CE) de clase IIa¹⁷ y es uno de los sistemas de IA de diagnóstico autónomo aprobado por la FDA (*Food and Drug Administration*) en EE. UU.¹⁸. La elevada precisión diagnóstica de este sistema de IA se ha demostrado previamente en conjuntos de datos disponibles al público¹⁶, en datos retrospectivos^{19,20} y en un ensayo clínico prospectivo en contraste con un patrón de resultados de pacientes²¹.

El objetivo de este estudio es comparar la precisión diagnóstica del IDx-DR v2 con la clasificación manual de los oftalmólogos españoles de acuerdo con la escala clínica

internacional de gravedad de la retinopatía diabética (ICDR)²² en una población real de pacientes en España.

Material y métodos

Población del estudio

Se obtuvieron imágenes retinianas de pacientes diabéticos consecutivos participantes en un programa regional de cribado de RD en atención primaria en 2011-2012 en Valencia (España) de acuerdo con un protocolo de obtención de imágenes estandarizado. Los pacientes se sometieron al cribado en 43 centros de salud de diversos departamentos de salud de la Comunidad Valenciana (Alcoy, Requena, Vinaroz y Xàtiva-Ontinyent)²³. De acuerdo con los criterios determinados, los pacientes se incluyeron cuando su médico de cabecera determinó que necesitaban una revisión ocular anual y podían colaborar con el protocolo de obtención de imágenes. Se obtuvo el consentimiento informado de todos los participantes con anterioridad al protocolo del estudio, que se realizó de acuerdo con la Declaración de Helsinki de 1983²⁴. La aprobación del estudio, desde el punto de vista ético se obtuvo del Comité Ético de Investigación Clínica de FISABIO Oftalmología Médica (FOM) para el uso de imágenes y datos anónimos de las clasificaciones.

Protocolo de obtención de imágenes

En la visita de cada paciente se realizó una fotografía de fondo de ojo con una cámara retiniana no midriática Topcon® TRC-NW200 (Topcon, Tokio, Japón) utilizando un protocolo estandarizado⁶. Enfermeros cualificados captaron imágenes retinianas de 45° de campo en color, de dos campos (un campo centrado en la fóvea y otro en el disco óptico) de cada ojo.

Las imágenes retinianas de cada visita del cribado se incluyeron en el sistema de gestión de pacientes del cribado (PMS) de FISABIO-Oftalmología Médica (FOM), Valencia, España. Para fines de este estudio, se obtuvieron datos retrospectivos de la base de datos del PMS en forma de retinografías (14.124) de 3.531 de pacientes únicos no identificables, seleccionados aleatoriamente. Se obtuvieron imágenes de ambos ojos de cada paciente (dos imágenes por ojo).

Clasificación del patrón de referencia

Tres oftalmólogos españoles clasificaron los exámenes de forma anónima tanto entre ellos como con respecto al sistema de IA de diagnóstico autónomo. La gravedad de la RD y la presencia de edema macular diabético (EMD) se clasificaron utilizando la escala ICDR²². Posteriormente, esa clasificación se sistematizó para las siguientes categorías:

- Retinopatía diabética con disminución de la visión (VTDR) definida según el *Early Treatment Diabetic Retinopathy Study* (ETDRS), nivel de gravedad 53 y superiores (o Retinopatía Diabética no proliferativa [RDNP] y Retinopatía Diabética Proliferativa [RDP] según la escala ICDR), o presencia de EMD significativo desde el punto de vista clínico (CSDME).
- Retinopatía diabética moderada (RDM) definida según el ETDRS²⁵, nivel de gravedad 35-47 (o RD moderada, según la escala ICDR), sin CSDME.

- Retinopatía diabética más que leve negativa definida según el ETDRS²⁵, gravedad 14-20 y 10 (o leve y sin RD según la escala ICDR) sin CSDME. Si es leve, se incluye en los resultados negativos de los sistemas de IA, en función del riesgo observado en el estudio ETDRS de que un paciente con enfermedad leve se quede ciego en un intervalo de cribado de un año²⁶.
- Cicatrices por láser. El clasificador registró la presencia de cicatrices de fotocoagulación pan-retiniana o focal. Cualquier paciente que hubiera recibido previamente tratamiento con láser se excluyó del análisis, de acuerdo con los criterios de exclusión para el cribado.
- Calidad de examen insuficiente. Evaluación pericial subjetiva de que las imágenes del examen presentaban una calidad insuficiente para una clasificación correcta. Esos exámenes se excluyeron del estudio.

La retinopatía diabética derivable (RDR) puede definirse como retinopatía diabética moderada (RDM) o retinopatía diabética con disminución de la visión (VTDR), mientras que la ausencia de RDM y VTDR no se considera RDR. La clasificación final del examen se adjudicó de la siguiente forma: se aceptó cuando existía un consenso en la clasificación de los tres clasificadores en cuanto a RDM, VTDR y RD no derivable. Cuando no existía un consenso en cuanto a los exámenes, un especialista en retina cualificado de un grupo independiente (F.V.) volvió a clasificar los exámenes de forma anónima, y su clasificación se aceptó si estaba de acuerdo con una de las tres clasificaciones originales de los oftalmólogos. En caso contrario, se consideró que la calidad del examen era insuficiente y se excluyó del estudio. El árbol de decisión de cada clasificador se muestra en la [figura 1](#).

Sistema de IA de diagnóstico automático

El sistema de IA ofrece como resultado un nivel específico de RD basado en la escala ICDR analizando las imágenes retinianas del examen y empleando algoritmos basados en inteligencia artificial y filtros aprendidos que forman un conjunto de detectores parcialmente dependientes¹⁶. Entre las principales características del sistema de IA se encuentran detectores de lesiones de alto rendimiento redundantes, validados de forma independiente y basados en aprendizaje profundo^{27,28}, un algoritmo de detección de la enfermedad final que evalúa la información proporcionada por esos detectores de lesiones, y un algoritmo independiente que evalúa si la calidad de las imágenes facilitadas es suficiente y ofrece la cobertura retiniana necesaria para permitir la exclusión de la presencia de la enfermedad de forma segura para el paciente.

Específicamente, el dispositivo determina en un minuto los siguientes niveles de enfermedad: VTDR, RDM y negativo. Asimismo, el sistema de IA de diagnóstico autónomo realiza inmediatamente un análisis de calidad de imagen. Cuando la calidad de imagen es insuficiente para descartar RDR, se determina que la calidad del examen es insuficiente y el dispositivo no ofrece ningún resultado. Si el paciente todavía se encuentra presente en ese momento, en la práctica clínica, el operador de la cámara puede volver a obtener inmediatamente las imágenes defectuosas y someter de nuevo el examen a análisis para obtener un resultado. De igual forma, el sistema de IA evalúa el protocolo de obtención de imágenes. El análisis del protocolo incluye la detección automática de las imágenes del

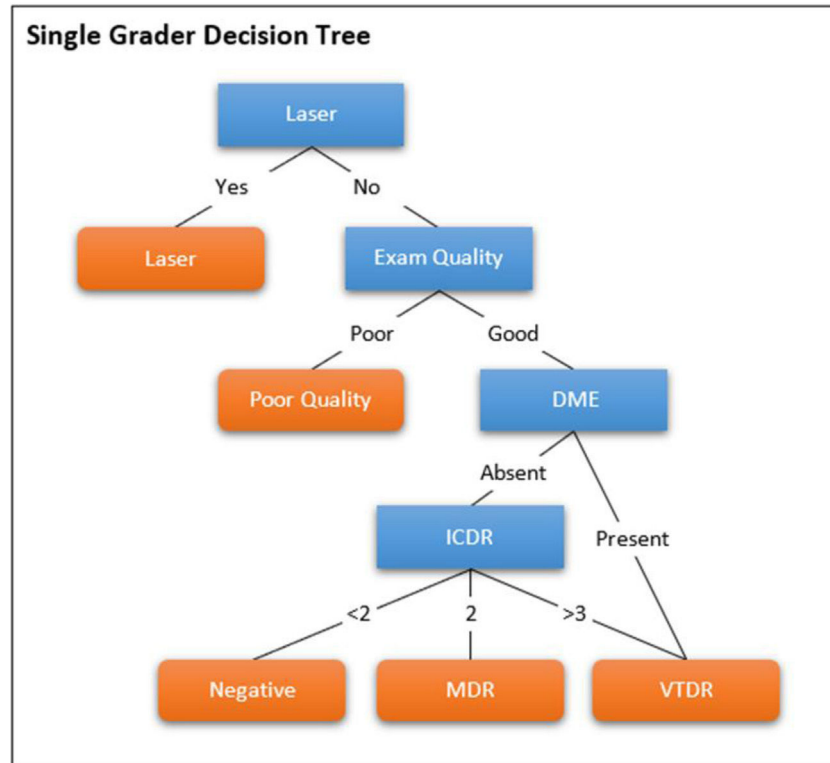


Figura 1 – Árbol de decisión de cada clasificador. El color azul indica las etapas de decisión y el naranja los resultados de cada una.

RDM: retinopatía diabética moderada; VTDR: retinopatía diabética con disminución de la visión.

ojo izquierdo y el ojo derecho, así como la localización de la fovea y la cabeza del nervio óptico para garantizar que se analicen imágenes centradas tanto de la fovea como de la cabeza del nervio óptico de cada ojo del paciente. En caso de que el sistema de IA detecte un fallo en el protocolo de obtención de imágenes, se considera que el protocolo del examen es incorrecto y el dispositivo no ofrece ningún resultado. En ese caso, se indica al operador de la cámara que debe volver a realizar el examen.

El resultado del sistema de IA de diagnóstico autónomo se genera utilizando métodos del valor umbral y combinando dos probabilidades posteriores diferentes, una de las cuales refleja la probabilidad de presencia de VTDR y la otra de RDM.

Análisis estadístico

Los propios pacientes informaron de la duración de la diabetes. Los resultados indexados continuos del sistema de IA se utilizaron para crear curvas características operativas del receptor (ROC) para VTDR y RDM, y se calculó el área bajo la curva (ABC) ROC. Se indicó la sensibilidad y la especificidad, el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN), el ABC del sistema de IA en contraste con la clasificación manual adjudicada y sus intervalos de confianza del 95%. Asimismo, se examinaron los efectos potenciales de sexo, edad y duración de la diabetes en la sensibilidad y la especificidad. Para cada subcategoría, se examinó la importancia de la diferencia de los parámetros observados utilizando la prueba de χ^2 de Pearson, donde un valor p de 0,05/3 (según la

corrección de Bonferroni) se consideró significativo. Para completar el análisis, se calculó la diferencia entre el ABC obtenida en este estudio y aquella obtenida a partir de la base de datos pública MESSIDOR-2 de libre acceso^{29,30} utilizando el mismo algoritmo, donde un valor p de 0,05 se consideró estadísticamente significativo^{27,28}. Por último, se calculó el aumento de eficiencia derivado de la reducción de la carga de trabajo de los oftalmólogos utilizando el sistema de IA de diagnóstico autónomo.

Resultados

De los 3.531 pacientes incluidos en este estudio, 250 se excluyeron debido a cicatrices por láser, 195 debido a una calidad de imagen insuficiente determinada únicamente por los clasificadores, 404 debido a una calidad de imagen insuficiente determinada por el sistema de IA, y 2 exámenes se excluyeron debido a una calidad de imagen insuficiente determinada tanto por los clasificadores como por el sistema de IA. Tras las eliminaciones pertinentes, se realizó el estudio 2.680 pacientes y 5.360 ojos. A cada ojo, se le realizaron dos retinografías: una centrada en la fovea y la otra en la papila. Por lo tanto, se analizaron 10.720 retinografías.

En los 3.531 sujetos, la edad media de los pacientes fue de 74 años, y la duración media de la diabetes de 6,9 años. El 55% de los pacientes eran mujeres y el 89% padecían diabetes tipo 2. En los 2680/3531 (76%) sujetos, de acuerdo con la clasificación

Tabla 1 – Clasificación del sistema de IA comparada con el patrón de referencia para retinopatía diabética derivable (RDR)

Patrón de referencia	Sistema de IA		
	Sin RDR	RDR	Total
Sin RDR	2.102	467	2.569
RDR	0	111	111
Total	2.102	578	2.680

Tabla 2 – Clasificación del sistema de IA comparada con el patrón de referencia para retinopatía diabética con disminución de la visión (VTDR)

Patrón de referencia	Sistema de IA		
	Sin RDR	RDR	Total
Sin RDR	2.471	140	2.611
RDR	0	69	69
Total	2.471	209	2.680

RDR: retinopatía diabética derivable.

Tabla 3 – Sensibilidad, especificidad, valor predictivo negativo (VPN) y valores predictivos positivos (VPP) del sistema de IA para la detección de RDR y VTDR

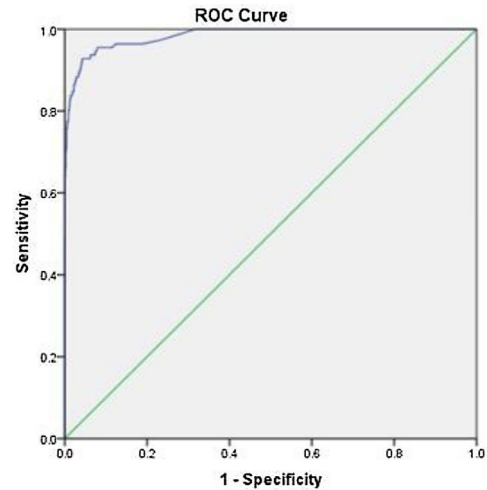
	RDR	VTDR
Sensibilidad	111= 100% (96,73-100%)	69= 100% (94,79-100%)
Especificidad	2.102= 81,82% (80,27-83,30%)	2.471= 96,64% (93,70-95,47%)
VPN	100%	100%
VPP	19,20% (17,96-20,51%)	33,01% (29,55-36,67%)

RDR: retinopatía diabética derivable; VTDR: retinopatía diabética con disminución de la visión.

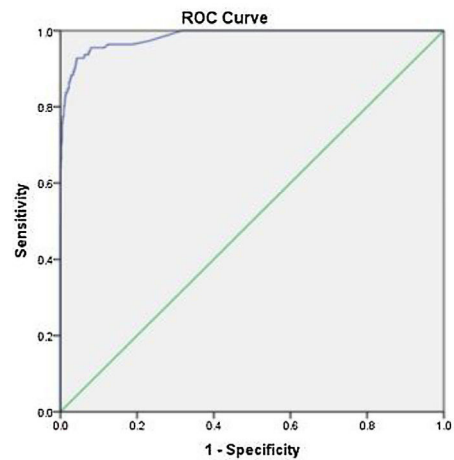
manual, la prevalencia de RDR fue de 111/2.680 (4,14%) y la de VTDR de 69/2.680 (2,57%).

Las tablas 1 y 2 relacionan los resultados del sistema de IA con respecto a la clasificación manual para RDR y VTDR. El rendimiento del sistema de IA se resume en la tabla 3. El sistema de IA registró cero falsos negativos tanto para RDR como para VTDR, lo que resultó en una sensibilidad de 111/111 = 100% (IC 95%: 97-100%) para RDR y una sensibilidad de 69/69 = 100% (IC 95%: 95-100%) para VTDR. Las especificidades correspondientes fueron de 2.102/2.569 = 82% (IC 95%: 80-83%) para RDR y 2.471/2.611 = 95% (IC 95%: 94-95%) para VTDR. El VPP correspondiente fue de 19,20% (IC 95%: 18-21%) para RDR y 33% (IC 95%: 30-37%) para VTDR, mientras que el VPN fue de 100% tanto para RDR como para VTDR.

El ABC fue de 0,984 (IC 95%: 0,97-0,99) al analizar la capacidad del sistema de IA para detectar RDR en comparación con la clasificación manual. La ROC se ilustra en la figura 2. Para la detección de VTDR por parte del sistema de IA, el ABC fue de 0,998 (IC 95%: 0,997-0,999). La ROC se muestra en la figura 1. Para RDR, el ABC registrada en este estudio (0,984) no presentó una diferencia significativa ($p > 0,70$) con aquella registrada con el mismo sistema de IA utilizando el conjunto de datos francés (0,980)¹⁶. Asimismo, el IC 95% del ABC registrada para



(a)



(b)

Figura 2 – Curva operativa del receptor: (a) RDR (b) VTDR.

RDR del sistema de IA tanto en este estudio (IC 95%: 0,97-0,99) como en el conjunto de datos francés (IC 95%: 0,968-0,992)¹⁶ se solapa con el ABC máxima teórica de 0,99 mensurable para un sistema de detección perfecto con la intervención de varios clasificadores²⁸.

El sistema de IA dio como resultado 111 verdaderos positivos (VP) y 467 falsos positivos (FP) para RDR de los 2.680 exámenes analizados (EA), lo que resultó en un aumento de la eficiencia (AE) del 78,43% ($AE = (1 (VP + FP)/EA) \times 100$).

Se estratificaron la sensibilidad y la especificidad para las siguientes subcategorías: sexo, edad superior a 65, duración de la diabetes > 10 años. No hubo efectos significativos de ninguna de esas categorías en la sensibilidad ($p > 0,05/3$). En cuanto a la especificidad, no hubo un efecto significativo del sexo ($p > 0,655$), aunque hubo un efecto significativo de la duración de la diabetes ($p < 0,05/3$) y la edad ($p < 0,05/3$): el sistema de IA mostró una especificidad más elevada del 86% en sujetos con una duración de la diabetes inferior a 10 años, mientras que la especificidad en sujetos con una duración de la diabetes superior a 10 años fue del 71% ($p < 0,0001$). De igual forma, el sistema de IA mostró una especificidad más elevada en sujetos más jóvenes, siendo del 89% en sujetos con menos de 65

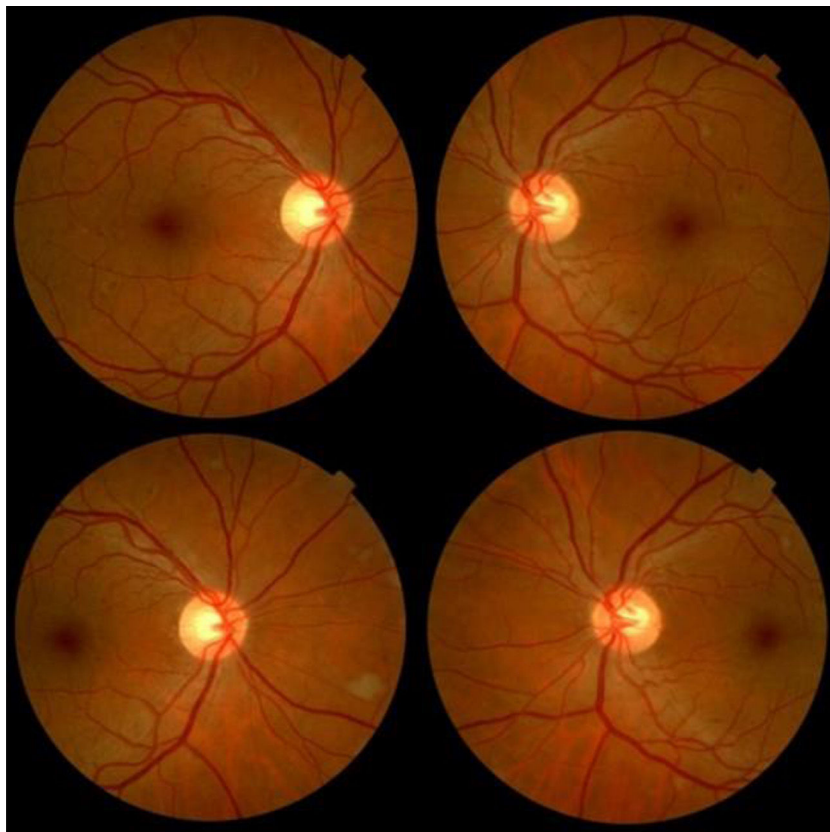


Figura 3 – Imagen de fondo de ojo de ejemplo de examen clasificado como negativo para RDR por clasificadores humanos y como positivo para RDM por el sistema de IA.

años, comparada con el 79% en sujetos mayores de 65 años ($p < 0,0001$).

Discusión

En este estudio retrospectivo, se compara el rendimiento de un sistema de IA de diagnóstico autónomo para la detección de RDR y VTDR, con un patrón de clasificación manual establecido por cuatro oftalmólogos en una población real de pacientes en España. El sistema de IA estudiado en este artículo, se ha estudiado anteriormente utilizando conjuntos de datos retrospectivos de la base de datos pública MESSIDOR-2 en la población francesa^{29,30}, así como de forma también retrospectiva en una muestra estadounidense clínica EyePACS registrada previamente^{31,32}. El conjunto de datos EyePACS-1 consistió en 9.963 imágenes de 4.997 pacientes (edad media, 54,4 años; 62,2% mujeres; prevalencia de RDR, 683/8.878 imágenes totalmente calificables (7,8%). El conjunto de datos MESSIDOR-2 tenía 1.748 imágenes de 874 pacientes (edad media, 57,6 años; 42,6% mujeres; prevalencia de RDR, 254/1.745 imágenes totalmente calificables (14,6%)²⁹.

Se han realizado estudios retrospectivos similares en diversas poblaciones utilizando otros algoritmos de IA y protocolos de estudio¹⁰⁻¹⁶. Cabe esperar una variación en el rendimiento entre esos estudios, teniendo en cuenta las diferencias existentes entre los clasificadores, la distribución de la

enfermedad y los patrones de referencia. Por tanto, es importante que este estudio se considere en el contexto de estudios adicionales que validen más ampliamente el rendimiento del sistema de IA utilizado.

Por ejemplo, cuando se diseña adecuadamente, el uso de una clasificación manual adjudicada por parte de varios oftalmólogos para determinar el rendimiento de IA puede ofrecer una excelente comparación con el patrón de atención en un contexto o ubicación determinados. Al mismo tiempo, esos estudios no ofrecen necesariamente un indicador para la relación de un sistema de IA con los resultados de los pacientes, dado el fenómeno documentado de la desviación diagnóstica, término acuñado para describir cómo, con el tiempo, los patrones que los profesionales sanitarios utilizan para diagnosticar una enfermedad se desvían de sus definiciones originales basadas en resultados^{33,34}. En un ensayo prospectivo clave publicado recientemente²¹ se utilizó un patrón de referencia con el nivel más elevado posible; un patrón de referencia basado en resultados clínicos de cada paciente, donde diagnósticos precisos de la enfermedad se correlacionan con la probabilidad real de pérdida de visión²⁵. Para la retinopatía diabética, esto se determina a través de la fotografía digital estéreo de cuatro campos en color de fondo de ojo dilatado, así como una tomografía de coherencia óptica macular (para el edema macular central) realizada por un fotógrafo oftalmológico titulado y experimentado y leída por un centro de lectura aprobado (en este caso, el Wisconsin Reading Center),

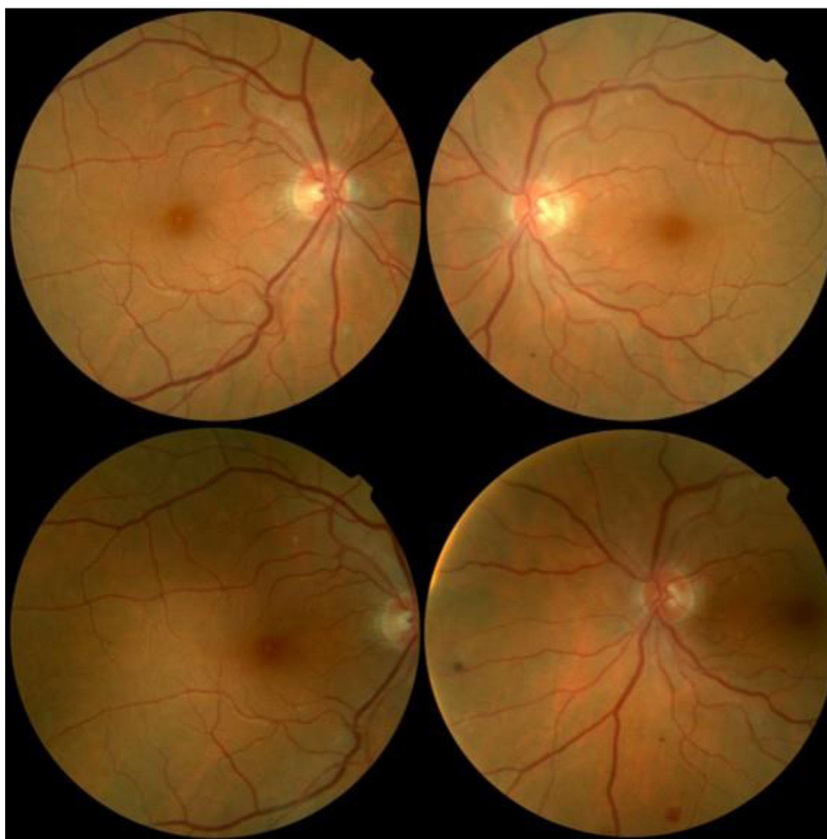


Figura 4 – Imagen de fondo de ojo de ejemplo de examen clasificado como negativo para RDR por clasificadores humanos y como positivo para RDM por el sistema de IA.

de acuerdo con una escala de gravedad correlacionada con el resultado de los pacientes, la del ETDRS²⁵. Según este protocolo, el sistema de IA de diagnóstico autónomo mostró un rendimiento con 87% de sensibilidad y 91% de especificidad. El rendimiento aparentemente mejor del sistema de IA idéntico presentado en el conjunto de datos español se debe a su comparación con el patrón de atención (clasificación manual adjudicada realizada por oftalmólogos) en lugar del patrón oro descrito anteriormente³⁵⁻³⁷.

En la escala ICDR, la RDM incluye los niveles de gravedad del 35 al 47 del *Early Treatment Diabetic Retinopathy Study* (ETDRS)²⁵. Se incluye en la definición de un nivel 35 del ETDRS la presencia de cualquier lesión característica de la RD (por ejemplo, hemorragias y exudados duros y blandos), así como uno o más microaneurismas. No todos los sistemas de clasificación alternativos basados en resultados independientes de los pacientes incluyen la totalidad de la definición de la gravedad 35 del ETDRS en su definición de la RDM. Como resultado de esa confusión, se produce una desviación diagnóstica, en función de la preparación y la experiencia previas de los clasificadores. Por tanto, la sensibilidad diagnóstica del sistema para RDR parece ser superior (se aproxima al 100%, en comparación con 87%) en este estudio, aunque, por supuesto, el rendimiento subyacente es el mismo. Las figuras 3, 4 y 5 muestran un ejemplo de examen clasificado como negativo para RDR por los oftalmólogos que, sin embargo, el sistema de IA consideró positivo. Claramente, cada examen

presenta lesiones que determinan una clasificación moderada en la escala ICDR, ya que hay hemorragias además de otras lesiones, microaneurismas o exudados duros y blandos²². Cabría esperar que la especificidad aumente en personas de mayor edad, ya que tienen una membrana limitante interna menos reflectora que puede indicar falsos positivos²¹. Sin embargo, este estudio demostró, por el contrario, que la especificidad tiende a ser menor en personas de mayor edad. Nuestra hipótesis es que esto se debe a los exámenes clasificados como negativos por los lectores en este estudio que se podrían considerar positivos de acuerdo con el ETDRS (figs. 3, 4, 5).

El sistema de IA de diagnóstico autónomo proporciona un resultado a escala del paciente, no una evaluación de la imagen. Por tanto, el algoritmo se diseñó para ofrecer un resultado con una calidad de imagen insuficiente, en vez de un diagnóstico de la enfermedad en casos en los que los datos de imagen eran insuficientes para emitir un diagnóstico. Esto se hace por motivos de seguridad de los pacientes y es un componente esencial de cualquier sistema de diagnóstico totalmente automatizado. En este estudio, el sistema de IA proporcionó un resultado diagnóstico para el 86,80% de los pacientes y un resultado de calidad insuficiente para el 13,20% de aquellos a los que se asignó una clasificación del patrón de referencia. Queda a discreción de oftalmólogos autorizados tomar una decisión médica con respecto a RDR en imágenes de calidad insuficiente. Las imágenes consideradas de calidad

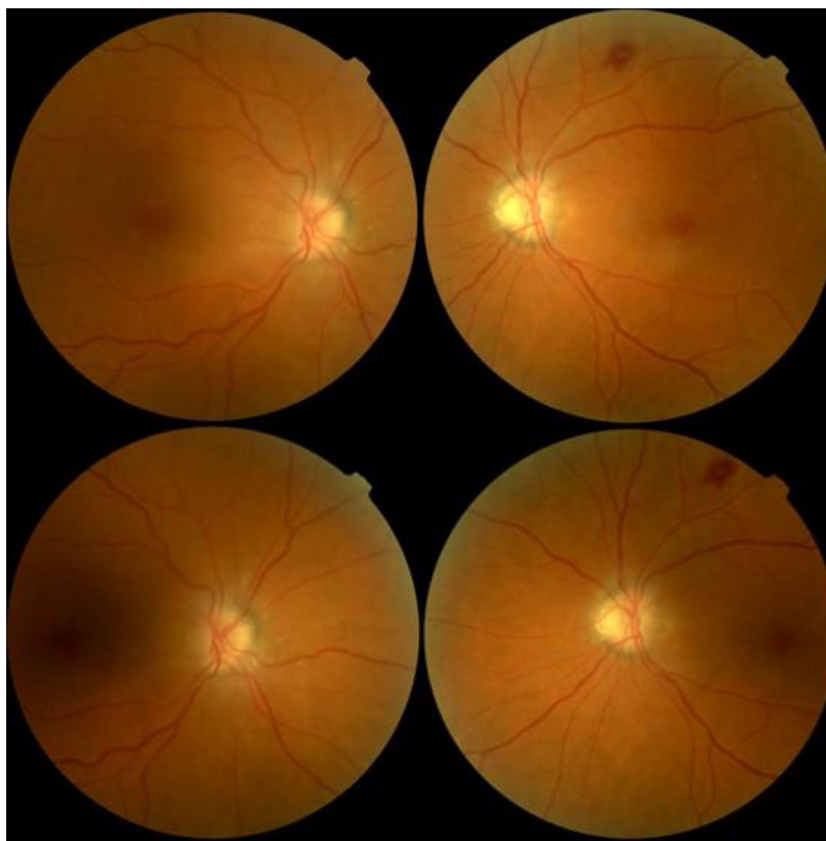


Figura 5 – Imagen de fondo de ojo de ejemplo de examen clasificado como negativo para RDR por clasificadores humanos y como positivo para RDM por el sistema de IA.

insuficiente por el sistema de IA, no se clasificaron y se eliminaron del estudio. Dado que este es un estudio retrospectivo, no fue posible volver a obtener imágenes de los pacientes. En un ensayo clínico prospectivo del mismo sistema de IA que permitió que se volvieran a obtener imágenes de los exámenes de calidad insuficiente, y en el que se utilizó una dilatación farmacológica cuando fue necesario, el sistema de IA proporcionó un diagnóstico de la enfermedad en el 96% de los pacientes que también contaban con una clasificación del patrón de referencia²¹.

Conclusiones

En general, en el contexto del estudio más amplio del sistema de IA de diagnóstico autónomo estudiado aquí, esta evaluación confirma el excelente rendimiento de los dispositivos en poblaciones clínicas, como se observó originalmente en un estudio clave riguroso²¹. En resumen, comparado con la clasificación manual realizada por oftalmólogos, el sistema de IA registró alta sensibilidad (se aproxima al 100%) y especificidad (82%) en el diagnóstico de RD y edema macular en personas con diabetes en un programa de cribado. Debido a su diagnóstico inmediato en el lugar en que se presta la atención médica, el sistema de IA de diagnóstico autónomo puede aumentar la accesibilidad al cribado de RD en contextos de atención primaria.

Financiación

Este estudio ha sido financiado por IDx Technologies Inc., Coralville, EE. UU. El contenido es de responsabilidad exclusiva de los autores.

Conflicto de intereses

El estudio fue financiado por IDx Technologies Inc. R.A. y W.C. son directores de investigación y desarrollo de IDx Technologies Inc. A.S. es ingeniero de software de investigación y desarrollo de IDx Technologies Inc. W.C., R.A. y A.S. son accionistas de IDx Technologies Inc. Todos los autores, FOM y EIBI, a excepción de A.N., P.P.J., A.A. v.d.H. y C.P.M. recibieron apoyo económico de IDx Technologies Inc. No se informó de otros posibles conflictos de intereses relevantes para este artículo.

Agradecimientos

Los autores agradecen a Enrique Soto-Pedre (*European Innovative Biomedicine Institute (EIBI)*, Cantabria, España, el diseño del protocolo utilizado en este estudio, así como la búsqueda de los oftalmólogos, a los que dio instrucciones para la clasificación manual de los exámenes.

BIBLIOGRAFÍA

1. Kocer I, Resnikoff S. Visual impairment and blindness in Europe and their prevention. *Br J Ophthalmol*. 2002;86:716–22, <http://dx.doi.org/10.1136/bjo.86.7.716>.
2. López M, Cos FX, Álvarez-Guisasola F, Fuster E. Prevalence of diabetic retinopathy and its relationship with glomerular filtration rate and other risk factors in patients with type 2 diabetes mellitus in Spain. DM2 HOPE study. *J Clin Transl Endocrinol*. 2017;9:61–5, <http://dx.doi.org/10.1016/j.jcte.2017.07.004>.
3. Alonso-Morán E, Orueta JF, Fraile Esteban JI, Arteagoitia Axpe JM, Marqués González ML, Toro Polanco N, et al. The prevalence of diabetes-related complications and multimorbidity in the population with type 2 diabetes mellitus in the Basque country. *BMC Public Health*. 2014;14:1059, <http://dx.doi.org/10.1186/1471-2458-14-1059>.
4. López IM, Díez A, Velilla S, Rueda A, Álvarez A, Pastor CJ. Prevalence of diabetic retinopathy and eye care in a rural area of Spain. *Ophthalmic Epidemiol*. 2002;9:205–14, <http://dx.doi.org/10.1076/opep.9.3.205.1516>.
5. Romero-Aroca P, Sagarra-Alamo R, Baget-Bernaldiz M, Fernández-Ballart J, Méndez-Marin I. Prevalence and relationship between diabetic retinopathy and nephropathy, and its risk factors in the north-east of Spain, a population-based study. *Ophthalmic Epidemiol*. 2010;17:251–65, <http://dx.doi.org/10.3109/09286586.2010.498661>.
6. Abramoff MD, Suttrop-Schulten MSA. Web-based screening for diabetic retinopathy in a primary care population: the EyeCheck project. *Telemed J E Health*. 2005;11:668–74, <http://dx.doi.org/10.1089/tmj.2005.11.668>.
7. Ólafsdóttir E, Stefánsson E. Biennial eye screening in patients with diabetes without retinopathy: 10-year experience. *Br J Ophthalmol*. 2007;91:1599–601, <http://dx.doi.org/10.1136/bjo.2007.123810>.
8. Scanlon PH. The English national screening programme for diabetic retinopathy 2003-2016. *Acta Diabetol*. 2017;54:515–25, <http://dx.doi.org/10.1007/s00592-017-0974-1>.
9. Chalk D, Pitt M, Vaidya B, Stein K. Can the retinal screening interval be safely increased to 2 years for type 2 diabetic patients without retinopathy? *Diabetes care*. 2012;35:1663–8, <http://dx.doi.org/10.2337/dc11-2282>.
10. Sinthanayothin C, Boyce JF, Williamson TH, Cook HL, Mensah E, Lal S, et al. Automated detection of diabetic retinopathy on digital fundus images. *Diabet Med*. 2002;19:105–12, <http://dx.doi.org/10.1046/j.1464-5491.2002.00613.x>.
11. Usher D, Dumskyj M, Himaga M, Williamson TH, Nussey S, Boyce J. Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. *Diabet Med*. 2004;21:84–90, <http://dx.doi.org/10.1046/j.1464-5491.2003.01085.x>.
12. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402–10, <http://dx.doi.org/10.1001/jama.2016.17216>.
13. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318:2211–23, <http://dx.doi.org/10.1001/jama.2017.18152>. PMID: 29234807.
14. Grauslund J, Jørgensen ME, Lund-Andersen H. Screening for Diabetic Retinopathy in Europe. Impact of New Technologies: Denmark. 2016. Disponible en: <https://portal.findresearcher.sdu.dk/en/publications/screening-for-diabetic-retinopathy-in-europe-impact-of-new-techno>.
15. Helmchen LA, Lehmann HP, Abramoff MD. Automated detection of retinal disease. *Am J Manag Care*. 2014;20:eSP48–52.
16. Abramoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57:5200–6, <http://dx.doi.org/10.1167/iovs.16-19964>.
17. IDx LLC. Fully automated diagnostic device receives CE certification; IDx LLC planning for rollout across Europe. 2013. Disponible en: <https://www.prnewswire.com/news-releases/fully-automated-diagnostic-device-receives-ce-certification-idx-llc-planning-for-rollout-across-europe-206263101.html>.
18. U.S. Food and Drug Administration. FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. 2018. Disponible en: <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye>.
19. van der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem A, Nijpels G. Validation of automated screening for referable diabetic retinopathy with the idx-dr device in the hoorn diabetes care system. *Acta Ophthalmol*. 2018;96:63–8, <http://dx.doi.org/10.1111/aos.13613>.
20. Verbraak FD, Abramoff MD, Bausch GCF, Klaver C, Nijpels G, Schlingemann RO, et al. Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting. *Diabetes care*. 2019;42:651–6, <http://dx.doi.org/10.2337/dc18-0148>.
21. Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *Npj Digital Medicine*. 2018;1:39.
22. Wilkinson CP, Ferris FL III, Klein RE, Lee PP, Agardh CD, Davis M-ET-AL. Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110:1677–82, [http://dx.doi.org/10.1016/S0161-6420\(03\)00475-5](http://dx.doi.org/10.1016/S0161-6420(03)00475-5).
23. Soto-Pedre E, Navea A, Millán S, Hernaez-Ortega MC, Morales J, Desco MC, et al. Evaluation of automated image analysis software for the detection of diabetic retinopathy to reduce the ophthalmologists' workload. *Acta Ophthalmol*. 2015;93:e52–6, <http://dx.doi.org/10.1111/aos.12481>.
24. World Medical Association. WMA declaration of Helsinki – ethical principles for medical research involving human subjects. Disponible en: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>.
25. Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs –an extension of the modified Airlie House classification. ETDRS report number 10. *Ophthalmology*. 1991;98:786–806.
26. Early Treatment Diabetic Retinopathy Study Research Group. Fundus photographic risk factors for progression of diabetic retinopathy. ETDRS report number 12. *Ophthalmology*. 1991;98:823–33.
27. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36, <http://dx.doi.org/10.1148/radiology.143.1.7063747>.
28. Quellec G; Abramoff MD. Estimating maximal measurable performance for automated decision systems from the

- characteristics of the reference standard. Application to diabetic retinopathy screening. Disponible en: <https://ieeexplore.ieee.org/document/6943552>.
29. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016;316:2402-10, <http://dx.doi.org/10.1001/jama.2016.17216>.
 30. Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology*. 2017;124:962-9, <http://dx.doi.org/10.1016/j.ophtha.2017.02.008>.
 31. Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *J Diabetes Sci Technol*. 2009;3:509-16, <http://dx.doi.org/10.1177/193229680900300315>.
 32. Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *PLoS One*. 2019;14:e0217541, <http://dx.doi.org/10.1371/journal.pone.0217541>.
 33. Chen JH, Goldstein MK, Asch SM, Altman RB. Dynamically evolving clinical practices and implications for predicting medical decisions. *Pac Symp Biocomput*. 2016:195-206.
 34. Arbel Y, Qiu F, Bennell MC, Austin PC, Roifman I, Rezaei MR, et al. Association between publication of appropriate use criteria and the temporal trends in diagnostic angiography in stable coronary artery disease: a population-based study. *Am Heart J*. 2016;175:153-9, <http://dx.doi.org/10.1016/j.ahj.2016.02.014>.
 35. Lawrence MG. The accuracy of digital-video retinal imaging to screen for diabetic retinopathy: an analysis of two digital-video retinal imaging systems using standard stereoscopic seven-field photography and dilated clinical examination as reference standards. *Trans Am Ophthalmol Soc*. 2004;102:321-40.
 36. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM. T.D. D. S. Group. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol*. 2002;134:204-13, [http://dx.doi.org/10.1016/S0002-9394\(02\)01522-2](http://dx.doi.org/10.1016/S0002-9394(02)01522-2).
 37. Pugh JA, Jacobson JM, Van Heuven WA, Watters JA, Tuley MR, Lairson DR, et al. Screening for diabetic retinopathy: the wide-angle retinal camera. *Diabetes Care*. 1993;16:889-95, <http://dx.doi.org/10.2337/diacare.16.6.889>.