

# Validación de cribado automatizado de retinopatía diabética derivable con un sistema de inteligencia artificial de diagnóstico autónomo en una población española

Abhay Shaha, Warren Clarida<sup>a</sup>, Ryan Amelona, María C. Hernáez-Ortega<sup>b</sup>, Amparo Naveac,<sup>d,e</sup> Jesús Morales-Olivasc, Rosa Dolz-Marcoc, Frank Verbraak<sup>f</sup>, Pablo Pérez Jordá<sup>c</sup>, Amber A van der Heijde<sup>g,h</sup>, Cristina Peris Martínez<sup>c,i</sup>

<sup>a</sup>*IDx Technologies Inc, Coralville, IA, EE. UU.*

<sup>b</sup>*European Innovative Biomedicine Institute (EIBI), Cantabria, España*

<sup>c</sup>*FISABIO OFTALMOLOGÍA MÉDICA (FOM) Valencia, España*

<sup>d</sup>*Instituto de la retina, Valencia, España*

<sup>e</sup>*Universidad Cardenal Herrera CEU*

<sup>f</sup>*Departamento de Oftalmología, VUmc, Centros Médicos de la Universidad de Ámsterdam,*

*Ámsterdam, Países Bajos*

<sup>g</sup>*Departamento de Medicina General y Geriátrica, Centro Médico de la Universidad VU, Ámsterdam, Países Bajos*

<sup>h</sup>*Instituto de Investigación en Salud Pública de Ámsterdam, Centro Médico de la Universidad VU,*

*Ámsterdam, Países Bajos*

<sup>i</sup>*Universidad de Valencia, España*

---

## Resumen

**Objetivo:** comparar el rendimiento diagnóstico de un sistema de inteligencia artificial (IA) de diagnóstico autónomo para el diagnóstico de retinopatía diabética derivable (RDR, por sus siglas en inglés) con la clasificación manual efectuada por oftalmólogos españoles.

**Métodos:** sujetos con diabetes tipo 1 y 2 participaron en un programa de cribado de retinopatía diabética (RD) en 2011-2012 en Valencia (España). Se recogieron 2 imágenes de cada ojo de acuerdo con el protocolo estándar. Se administraron gotas de acción midriática a todos los pacientes. Con la aprobación del Comité Ético de Investigación Clínica se obtuvieron imágenes retinianas no identificables, una centrada en el disco y otra en la fovea. Los exámenes se clasificaron tanto con el sistema de IA autónomo (IDx-DR, Coralville, Iowa, EE. UU.) como manualmente por parte de oftalmólogos anónimos a través de un proceso de adjudicación. Los resultados del sistema de IA y de la clasificación manual adjudicada se compararon en cuanto a sensibilidad y especificidad para el diagnóstico

tanto de retinopatía diabética derivable (RDR) como de retinopatía diabética con disminución de la visión (VTDR, por sus siglas en inglés).

Resultados: se incluyeron 2680 sujetos en el estudio. De acuerdo con la clasificación manual, la prevalencia de RDR fue de 111/2680 (4,14 %) y la de VTDR de 69/2680 (2,57 %). En contraste con la clasificación manual, el sistema de IA registró un 100 % (95 % CI: 97 % - 100 %) de sensibilidad y un 81,82 % (95 % CI: 80 % - 83 %) de especificidad para RDR, y un 100 % (95 % CI: 95 % - 100 %) de sensibilidad y un 94,64 % (95 % CI: 94 % - 95 %) de especificidad para VTDR.

Conclusiones: en comparación con la clasificación manual realizada por oftalmólogos, el sistema de IA de diagnóstico autónomo registró una sensibilidad (100 %) y una especificidad (82 %) elevadas en el diagnóstico de retinopatía diabética derivable 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 RDR en contextos de atención primaria.

---

## **1. Introducción**

A pesar de los avances en la atención a 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 laboral [1]. La prevalencia de RD en personas diabéticas oscila entre el 5 y el 30 % [2, 3, 4, 5, 6]. El cribado de RD ha demostrado su efectividad en la prevención de la ceguera, y muchas organizaciones profesionales han adoptado recomendaciones para un cribado de RD regular [7, 8, 9].

Los sistemas de inteligencia artificial (IA) para la clasificación y el cribado de RD pueden aumentar la accesibilidad al cribado de 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 RD [10, 11, 12, 13, 14, 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 RDR analizando imágenes

retinianas para detectar signos de RDR y edema macular diabético (EMD) sin supervisión humana o de un médico [16]. El sistema de IA cuenta con un marcado CE de clase IIa [17] y es actualmente el único sistema de IA de diagnóstico autónomo aprobado por la FDA en EE. UU. [18]. La elevada precisión diagnóstica de este sistema de IA se ha demostrado

previamente en conjuntos de datos disponibles al público [16], en datos retrospectivos [19, 20] y en un ensayo clínico prospectivo en contraste con un patrón de resultados de pacientes [21].

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, por sus siglas en inglés) [22] en una población real de pacientes en España.

## **2. Métodos**

### *2.1. 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 (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 1983 [24]. La aprobación desde el punto de vista ético se obtuvo del Comité Ético de Investigación Clínica de FOM para el uso de imágenes y datos anónimos de las clasificaciones.

### *2.2. 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 estandarizado [6]. Enfermeros cualificados capturaron 45 imágenes retinianas en color de dos campos (un campo centrado en la fovea 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, por sus siglas en inglés) 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 3531 exámenes de pacientes únicos no identificables seleccionados aleatoriamente.

#### *2.2.1. 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 retinopatía diabética y la presencia de EMD se clasificaron utilizando la escala ICDR [20]. Posteriormente, esa clasificación se sistematizó para clasificar 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 RDNP, RDP, según la escala ICDR), o presencia de edema macular diabético significativo desde el punto de vista clínico (CSDME, según sus siglas en inglés).
- Retinopatía diabética moderada (RDM) definida según el ETDRS [25], nivel de gravedad 35-47 (o RD moderada, según la escala ICDR), sin CSDME.
- Retinopatía diabética más que leve negativa definida según el ETDRS [25], 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 año [26].
- Cicatrices por láser. El clasificador registró la presencia de cicatrices de fotocoagulación panretiniana 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 RDM o 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.

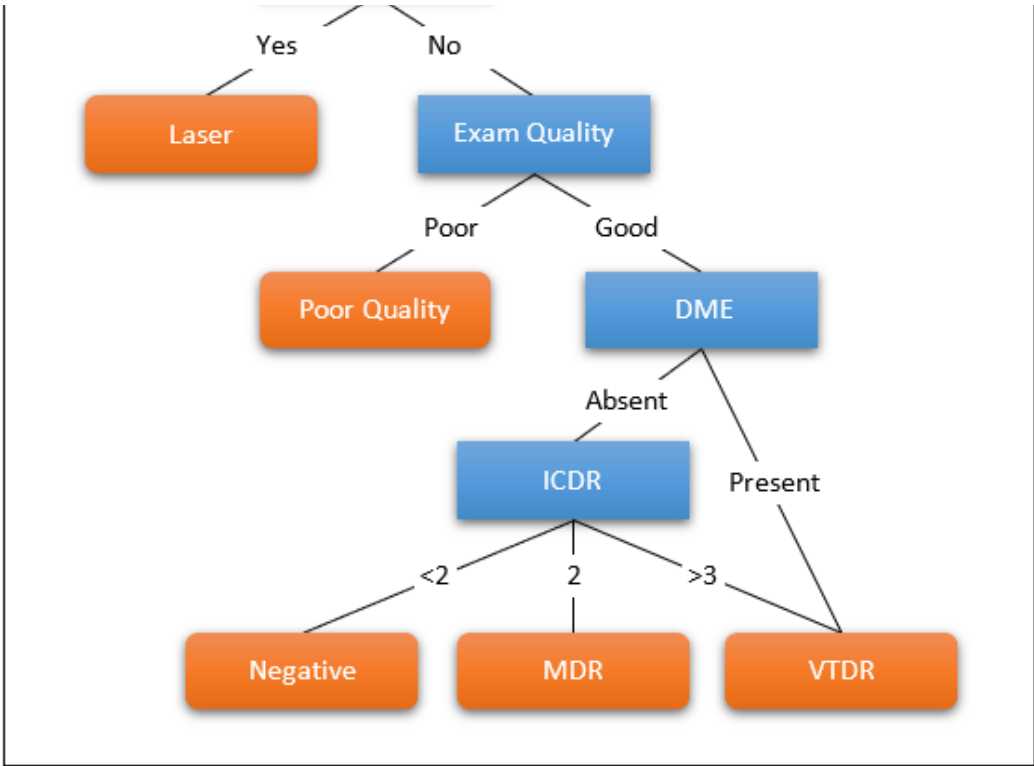


Figura 1: árbol de decisión de cada clasificador. RDM significa retinopatía diabética moderada, VTDR significa retinopatía diabética con disminución de la visión. El color azul indica las etapas de decisión y el naranja los resultados de cada una.



### 2.2.2. 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 dependientes [16]. 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], 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 (solo) 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 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.

### 2.2.3. 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, por sus siglas en inglés) para VTDR y RDM, y se calculó el área situada bajo la ROC (ABC). Se indicó la

sensibilidad y la especificidad, el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN), la 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 chi-cuadrado 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 la ABC obtenida en este estudio y aquella obtenida a partir del conjunto de datos francés [16] utilizando el mismo algoritmo, donde un valor p de 0,05 se consideró 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.

### **3. Resultados**

De los 3531 exámenes 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.

En los 3531 sujetos, la media de edad 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 manual, la prevalencia de RDR fue de 111/2680 (4,14 %) y la de VTDR de 69/2680 (2,57 %).

Los cuadros 1 y 2 tabulan 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 el cuadro 3. El sistema de IA registró 0 falsos negativos tanto para RDR como para VTDR, lo que resultó en una sensibilidad de  $111/111 = 100\%$  (95 % CI: 97 % - 100 %) para RDR y una sensibilidad de  $69/69 = 100\%$  (95 % CI: 95 % - 100 %) para VTDR. Las especificidades correspondientes fueron de  $2102/2569 = 82\%$  (95 % CI: 80 % - 83 %) para RDR y  $2471/2611 = 95\%$  (95 % CI: 94 % - 95 %) para VTDR. El VPP correspondiente fue de 19,2 % (95 % CI: 18 % - 21 %) para RDR y 33 % (95 % CI: 30 % - 37 %) para VTDR, mientras que el VPN fue de 100 % tanto para RDR como para VTDR.

La ABC fue de 0,984 (95 % CI: 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, la ABC fue de 0,998 (95 % CI: 0,997 - 0,999). La ROC se muestra en la figura 1. Para RDR, la 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 95 % CI de la ABC registrada

Cuadro 1: clasificación del sistema de IA comparada con el patrón oro para RDR

Patrón oro	Sistema de IA		
	Sin RDR	RDR	Total
Sin RDR	2102	467	2569
RDR	0	111	111
Total	2102	578	2680

Cuadro 2: clasificación del sistema de IA comparada con el patrón oro para VTDR

Patrón oro	Sistema de IA		
	Sin RDR	RDR	Total
Sin RDR	2471	140	2611
RDR	0	69	69
Total	2471	209	2680

Cuadro 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	$\frac{111}{111} = 100\% (96,73\% - 100\%)$	$\frac{69}{69} = 100\% (94,79\% - 100\%)$
Especificidad	$\frac{2102}{2569} = 81,82\% (80,27\% - 83,30\%)$	$\frac{2471}{2611} = 94,64\% (93,70\% - 95,47\%)$
VPN	100 %	100 %
VPP	19,20 % (17,96 % - 20,51 %)	33,01 % (29,55 % - 36,67 %)

para RDR del sistema de IA tanto en este estudio (95 % CI: 0,97 - 0,99) como en el conjunto de datos francés (95 % CI: 0,968 - 0,992) [16] se solapa con la ABC máxima teórica de 0,99 mensurable para un sistema de detección perfecto con la intervención de varios clasificadores [28].

El sistema de IA dio como resultado 111 verdaderos positivos (VP) y 467 falsos positivos (FP) para RDR de los 2680 exámenes analizados (EA), lo que resultó en un aumento de la eficiencia (AE) del 78,43 % ( $AE = (1 \cdot (VP + FP) / EA) \times 100$ ).

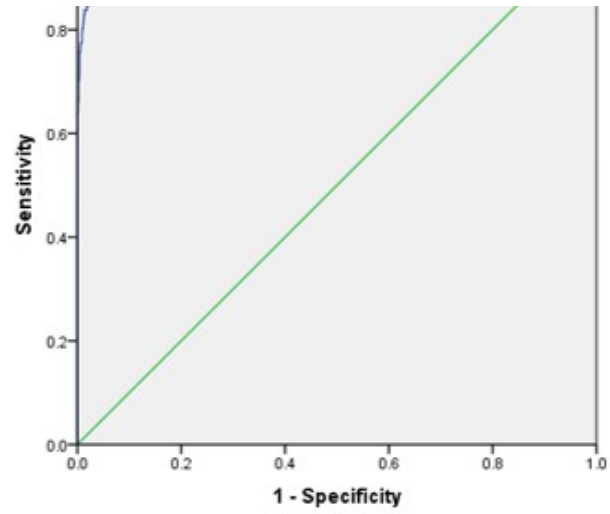
Se estratificaron la sensibilidad y la especificidad para las siguientes subcategorías: sexo, edad superior a 65, duración de la diabetes > 10 años. No hubo efectos significativos de ninguna de esas categorías en la sensibilidad ( $p > 0,05 / 3$ ). En cuanto a la especificidad, no hubo un efecto significativo 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 años, comparada con el 79 % en sujetos mayores de 65 años ( $p < 0,0001$ ).

#### **4. Discusión**

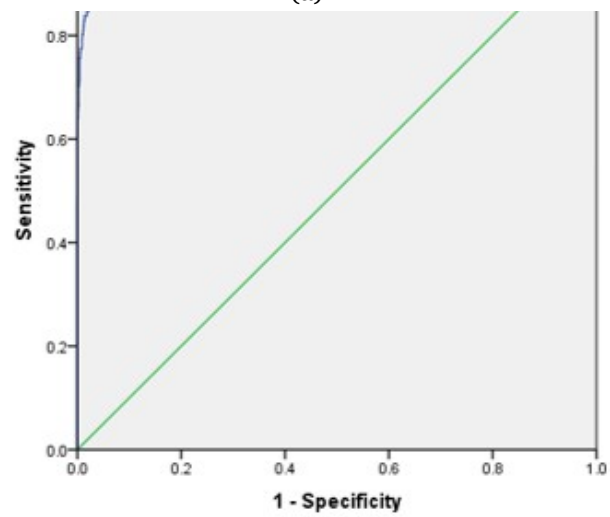
Este estudio compara el rendimiento de un sistema de IA de diagnóstico autónomo para la detección de RDR y VTDR en comparación con un patrón de clasificación manual adjudicada establecido por cuatro oftalmólogos en una población real de pacientes en España. El sistema de IA estudiado en este artículo se ha estudiado anteriormente utilizando conjuntos de datos retrospectivos en la población francesa [16], así como de forma retrospectiva en una muestra estadounidense clínica diversa registrada previamente [21]. Este estudio en una población española debe considerarse confirmatorio en el contexto de la demostración prospectiva del rendimiento de los sistemas, reflejando el rendimiento del sistema de IA en una población española con un patrón de referencia establecido por cuatro oftalmólogos.

Se han realizado estudios retrospectivos similares en diversas poblaciones utilizando otros algoritmos de IA y protocolos de estudio [10, 11, 12, 13, 14, 15, 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.



(a)



(b)

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



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 [29, 30]. En un ensayo prospectivo clave publicado recientemente [21] se utilizó un patrón de referencia con el nivel más elevado posible; un patrón oro o patrón de referencia basado en resultados clínicos a nivel del paciente, donde diagnósticos precisos a nivel de la enfermedad se correlacionan con la probabilidad real de pérdida de visión [25]. Para la retinopatía diabética, esto se determina a través de la fotografía digital estéreo de 4 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), de acuerdo con una escala de gravedad correlacionada con el resultado de los pacientes, la del ETDRS [25]. 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 anteriormente [31, 32, 33].

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 retinopatía diabética (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 del nivel de gravedad 35 del ETDRS en su definición de la retinopatía diabética moderada. 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 superior (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 blandos [22]. Cabría esperar que la especificidad aumente en personas de mayor edad, ya que tienen

una membrana limitante interna menos reflectora que puede indicar falsos positivos [21]. 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 (figuras 3, 4, 5).

El sistema de IA de diagnóstico autónomo proporciona un resultado a nivel del paciente, no una evaluación a nivel 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 86,8 % de los pacientes y un resultado de calidad insuficiente para el 13,2 % de los pacientes 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. La figura 6 muestra ejemplos de pares de imágenes consideradas de calidad insuficiente por el sistema de IA (que, por tanto, no se clasificaron para RD) que se clasificaron como de buena calidad y recibieron una clasificación final del patrón de referencia de gravedad de la enfermedad de RD. 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 referencia [21].

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 riguroso [21]. En resumen, comparado con la clasificación manual realizada por oftalmólogos, el sistema de IA registró alta sensibilidad (100 %) y especificidad (82 %) en el diagnóstico de retinopatía diabética 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.

## **5. 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

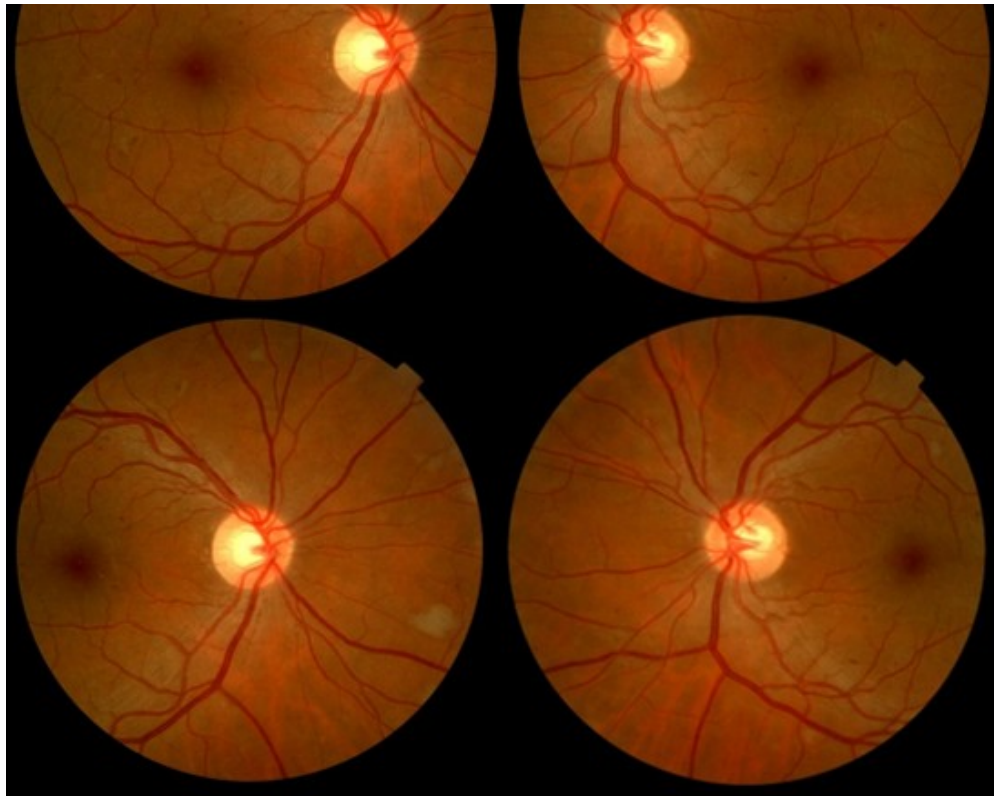


Figura 3: ilustración de ejemplo de examen clasificado como negativo para RDR por clasificadores humanos y como positivo para RDM por el sistema de IA.

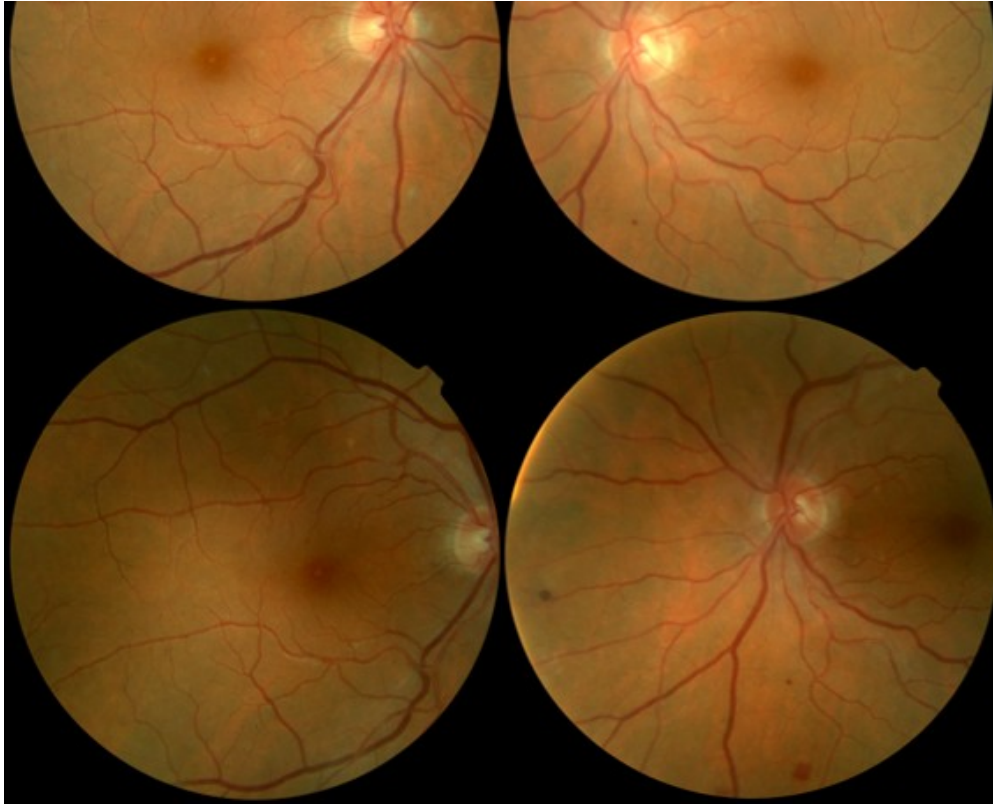


Figura 4: ilustración de ejemplo de examen clasificado como negativo para RDR por clasificadores humanos y como positivo para RDM por el sistema de IA.

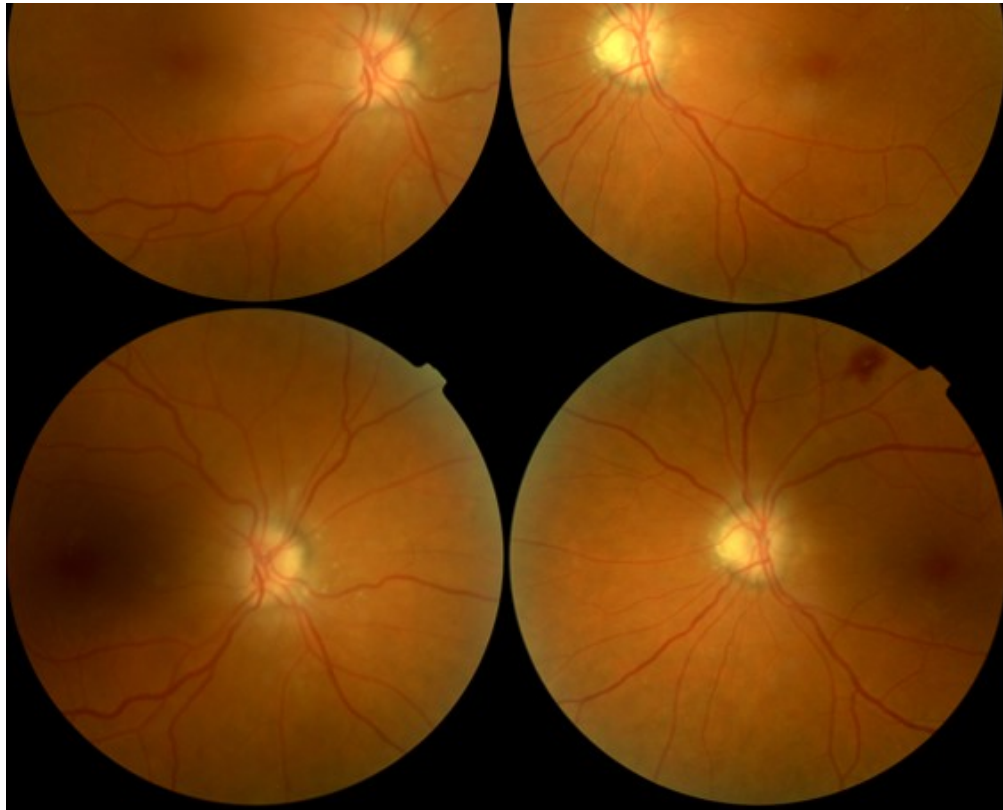


Figura 5: ilustración de ejemplo de examen clasificado como negativo para RDR por clasificadores humanos y como positivo para RDM por el sistema de IA.

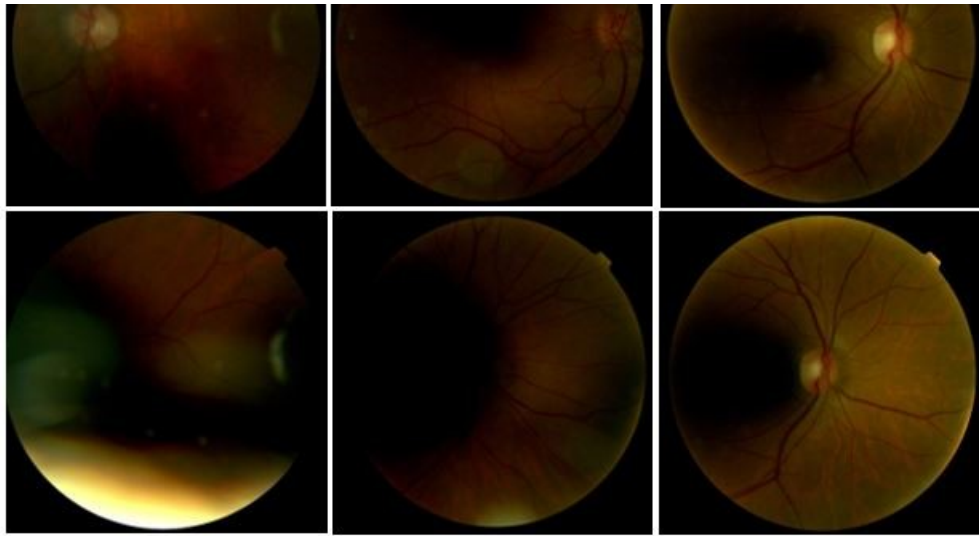


Figura 6: fotografía de fondo de ojo de varios pacientes realizada con una cámara retiniana no midriática (Topcon TRC-NW200). Cada columna muestra pares de imágenes del mismo examen. El sistema de IA consideró que la calidad de imagen de los pares de imágenes de la ilustración era insuficiente y, por tanto, no se clasificaron para RD. En contraste, los lectores humanos consideraron que la calidad de las imágenes era suficiente y las clasificaron para RD.



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.

## **6. Financiación**

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

## **7. Contribución de los autores**

A. S. y F. V. redactaron el manuscrito. F. V. supervisó el estudio. A. N. y EIBI se ocuparon del concepto y el diseño del estudio. EIBI, A. N., C. P. M. y P. P. J. se ocuparon de la recogida de datos. W. C. y A. A. v. d. H. analizaron los datos y realizaron el análisis estadístico. Todos los autores realizaron una revisión crítica del manuscrito en cuanto al contenido intelectual importante, y prestaron apoyo administrativo, técnico o material. F. V. es el responsable de este trabajo.

Correspondencia - Abhay Shah, 2300 Oakdale Blvd, Coralville, IA 52241, EE. UU. Contacto: (319) 248-5620. Dirección de correo electrónico: shah@eyediagnosis.net

## **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.

## **Referencia**

**s**

- [1] I. Kocur, S. Resnikoff, Visual impairment and blindness in Europe and their prevention, *British Journal of Ophthalmology* 86 (7) (2002) 716–722.
- [2] M. López, F. X. Cos, F. Álvarez-Guisasola, E. Fuster, 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, *Journal of clinical & translational endocrinology* 9 (2017) 61–65.
- [3] E. Alonso-Morán, J. F. Orueta, J. I. F. Esteban, J. M. A. Axpe, M. L. M. González, N. T. Polanco, P. E. Loiola, S. Gaztambide, R. Nuño-Solinis, The prevalence of diabetes-related complications and multimorbidity in the population with type 2 diabetes mellitus in the Basque country, *BMC public health* 14 (1) (2014) 1059.
- [4] M. I. López, A. Díez, S. Velilla, A. Rueda, A. Álvarez, J. C. Pastor, Prevalence of diabetic retinopathy and eye care in a rural area of Spain, *Ophthalmic epidemiology* 9 (3) (2002) 205–214.
- [5] R.-A. Pedro, S.-A. Ramon, B.-B. Marc, F.-B. Juan, M.-M. Isabel, Prevalence and relationship between diabetic retinopathy and nephropathy, and its risk factors in the north-east of Spain, a population-based study, *Ophthalmic epidemiology* 17 (4) (2010) 251–265.
- [6] M. D. Abramoff, M. S. Suttorp-Schulten, Web-based screening for diabetic retinopathy in a primary care population: the eyecheck project, *Telemedicine Journal & e-Health* 11 (6) (2005) 668–674.
- [7] E. Olafsdottir, E. Stefansson, Biennial eye screening in patients with diabetes without retinopathy: 10-year experience, *British journal of ophthalmology* 91 (12) (2007) 1599–1601.
- [8] P. H. Scanlon, The English national screening programme for diabetic retinopathy 2003–2016, *Acta diabetologica* 54 (6) (2017) 515–525.
- [9] D. Chalk, M. Pitt, B. Vaidya, K. Stein, Can the retinal screening interval be safely increased to 2 years for type 2 diabetic patients without retinopathy? *Diabetes care* 35 (8) (2012) 1663–1668.
- [10] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, D. Usher, Automated detection of diabetic retinopathy on digital fundus images, *Diabetic medicine* 19 (2) (2002) 105–112.
- [11] D. Usher, M. Dumskyj, M. Himaga, T. H. Williamson, S. Nussey, J. Boyce, Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening, *Diabetic Medicine* 21 (1) (2004) 84–90.

- [12] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *Jama* 316 (22) (2016) 2402–2410.
- [13] D. S. W. Ting, C. Y.-L. Cheung, G. Lim, G. S. W. Tan, N. D. Quang, A. Gan, H. Hamzah, R. Garcia-Franco, I. Y. San Yeo, S. Y. Lee, 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* 318 (22) (2017) 2211–2223.
- [14] J. Grauslund, M. E. Jørgensen, H. Lund-Andersen, Screening for diabetic retinopathy in Europe. Impact of new technologies: Denmark, in: National Representatives Meeting, 2016.
- [15] L. A. Helmchen, H. P. Lehmann, M. D. Abramoff, Automated detection of retinal disease, *Am J Manag Care* 11 (2014) 20.
- [16] M. D. Abramoff, Y. Lou, A. Erginay, W. Clarida, R. Amelon, J. C. Folk, M. Niemeijer, Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning, *Investigative ophthalmology & visual science* 57 (13) (2016) 5200–5206.
- [17] Cision pr newswire - fully automated diagnostic device receives ce certification; idx llc planning for rollout across Europe, <https://www.prnewswire.com/news-releases/fully-automated-diagnostic-device-receives-ce-certification-idx-llc-planning-for-rollout-across-europe-206263101.html>, acceso: 15-3-2019.
- [18] Fda news release - fda permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems, <https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm604357.htm>, acceso: 15-3-2019.
- [19] A. A. Van Der Heijden, M. D. Abramoff, F. Verbraak, M. V. van Hecke, A. Liem, G. Nijpels, Validation of automated screening for referable diabetic retinopathy with the idx-dr device in the hoorn diabetes care system, *Acta ophthalmologica* 96 (1) (2018) 63–68.

- [20] F. D. Verbraak, M. D. Abramoff, G. C. Bausch, C. Klaver, G. Nijpels, R. O. Schlingemann, A. A. van der Heijden, Diagnostic accuracy of a device for the automated detection of diabetic retinopathy in a primary care setting, *Diabetes care* 42 (4) (2019) 651–656.
- [21] M. D. Abramoff, P. T. Lavin, M. Birch, N. Shah, J. C. Folk, Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices, *Npj Digital Medicine* 1 (1) (2018) 39.
- [22] C. Wilkinson, F. L. Ferris III, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, J. T. Verdager, et al., Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales, *Ophthalmology* 110 (9) (2003) 1677–1682.
- [23] E. Soto-Pedre, A. Navea, S. Millán, M. C. Hernández-Ortega, J. Morales, M. C. Desco, P. Pérez, Evaluation of automated image analysis software for the detection of diabetic retinopathy to reduce the ophthalmologists' workload, *Acta ophthalmologica* 93 (1) (2015) e52–e56.
- [24] Declaration of Helsinki: ethical principles for medical research involving human subjects, <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>, acceso: 15-3-2019.
- [25] E. T. D. R. S. R. Group, et al., Grading diabetic retinopathy from stereoscopic color fundus photographs -- an extension of the modified Airlie House classification: ETDRS report number 10, *Ophthalmology* 98 (5) (1991) 786-806.
- [26] E. T. D. R. S. R. Group, et al., Fundus photographic risk factors for progression of diabetic retinopathy: Etdrs report number 12, *Ophthalmology* 98 (5) (1991) 823–833.
- [27] J. A. Hanley, B. J. McNeil, The meaning and use of the area under a receiver operating characteristic (roc) curve., *Radiology* 143 (1) (1982) 29–36.
- [28] G. Quellec, M. D. Abramoff, Estimating maximal measurable performance for automated decision systems from the characteristics of the

- reference standard. Application to diabetic retinopathy screening, in: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2014, pp. 154–157.
- [29] J. H. Chen, M. K. Goldstein, S. M. Asch, R. B. Altman, Dynamically evolving clinical practices and implications for predicting medical decisions, in: *Biocomputing 2016: Proceedings of the Pacific Symposium*, World Scientific, 2016, pp. 195–206.
- [30] Y. Arbel, F. Qiu, M. C. Bennell, P. C. Austin, I. Roifman, M. R. Rezai, J. V. Tu, D. T. Ko, H. C. Wijeyesundera, Association between publication of appropriate use criteria and the temporal trends in diagnostic angiography in stable coronary artery disease: a population-based study, *American heart journal* 175 (2016) 153–159.
- [31] M. G. Lawrence, 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, *Transactions of the American Ophthalmological Society* 102 (2004) 321.
- [32] D. Y. Lin, M. S. Blumenkranz, R. J. Brothers, D. M. Grosvenor, 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, *American journal of ophthalmology* 134 (2) (2002) 204–213.
- [33] J. A. Pugh, J. M. Jacobson, W. Van Heuven, J. A. Watters, M. R. Tuley, D. R. Lairson, R. J. Lorimor, A. S. Kapadia, R. Velez, Screening for diabetic retinopathy: the wide-angle retinal camera, *Diabetes care* 16 (6) (1993) 889–895.