ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

Rev Med Chile 2025; 153(10): 686-694. DOI: 10.4067/s0034-98872025001000686

Amending AI Software Accuracy for Diabetic Retinopathy Detection Using Conditional Probability and the Appropriate Reference Standard

Christian Segovia[1,2,3,a,b,], Daniela Salinas-Toro[3,4,b], Carla Moraga[3,b], Maylyn Sepúlveda[5,b].*

[1]Programa de Doctorado en Salud Ecosistémica, Centro de Investigaciones y Estudios Avanzados del Maule, Universidad Católica del Maule. Talca, Chile.
[2]Laboratorio de Microbiología y Parasitología, Departamento de Ciencias Preclínicas, Facultad de Medicina, Universidad Católica del Maule. Talca, Chile.
[3]Fundación de Investigación y desarrollo Científico en Salud (FICSA). Santiago, Chile.
[4]Servicio de Oftalmología, Clínica Las Condes. Santiago, Chile.
[5]Departamento de Tecnología Médica, Facultad de Ciencias de la Salud, Universidad Católica del Maule. Talca, Chile.
[a]PhD(c) en Salud Ecosistémica. Universidad Católica del Maule.
[b]Tecnólogo Médico con mención en Oftalmología y Optometría.

Mejorando la precisión de software de IA para detección de retinopatía diabética mediante probabilidad condicional y estándar de referencia adecuado

ABSTRACT

Aim: *To evaluate and correct the reported sensitivity and specificity values of DART (Diagnóstico Automatizado de Retinografías Telemáticas), an automated Artificial Intelligence based (AI-based) screening tool used for diabetic retinopathy (DR) detection in the Chilean public healthcare system, by employing the appropriate gold standard and conditional probability.* ***Methods:*** *Data were obtained from the clinical validation of DART. DR detection capabilities were assessed for three methods 1) Fundoscopy, 2) Retinography and 3) AI- DART. To estimate the true sensitivity and specificity of DART, conditional probability was applied using three hypothetical sensitivities level for method 2: A) Optimistic (90%), B) Moderate (80%), and C) Pessimistic (70%). Based on these scenarios, corrected sensitivity and specificity values for DART were calculated, along with false negative/positive rates (%FN/%FP), and predictive values (NPV/PPV).* ***Results:*** *In all scenarios, corrected sensitivity and specificity values for DART were significantly lower than those reported in the original validation study. Compared to method 3 (AI-based), method 2 (retinography by and ophthalmologist) consistently demonstrated superior performance across all metrics, including FN%, FP%, NPV and PPV values.* ***Conclusion:*** *While the integration of new AI-based technologies like DART in healthcare offer promise for enhancing patient care, their implementation must be preceded by validation using the correct gold standard. Reliable clinical decision-making depends on trustworthy diagnostic parameters.*

Keywords: *Artificial Intelligence; Diabetic retinopathy; Telemedicine.*

**Corresponding Author: Christian Tomás Segovia Cabello / christian.segovia.c@gmail.com*

Funding: This work received no funding.

The authors declare no conflict of interest.

Received: March 30, 2025.
Accepted: July 1, 2025.

RESUMEN

Objetivo: *Evaluar y corregir los valores de sensibilidad y especificidad de la herramienta automatizada de Inteligencia Artificial DART (Diagnóstico Automatizado de Retinografías Telemáticas) empleada en la detección de Retinopatía Diabética en el sistema público de salud en Chile, utilizando un estándar de referencia adecuado.* ***Métodos:*** *Los casos fueron calculados a partir de los datos utilizados en la validación de DART. Se evaluó la capacidad de detección de RD para: 1) Fundoscopía, 2) Retinografía, 3) IA DART. Para el cálculo de la especificidad real o corregida de DART, se aplicó probabilidad condicional considerando las siguientes sensibilidades para el método 2: A) Optimista (90%), B) Moderado (80%) y C) Pesimista (70%). De acuerdo a estos supuestos, se estimaron los valores reales (corregidos) de sensibilidad y especificidad de DART. Además, se calcularon el porcentaje de falsos negativos y positivos (%FN, %FP), así como el Valor Predictivo Negativo (VPN) y Valor Predictivo Positivo (VPP).* ***Resultados:*** *En general, los valores corregidos de sensibilidad y especificidad para los tres escenarios (A, B y C) fueron significativamente menores en todos los casos en comparación con los reportados en la validación de la herramienta IA. Además, al comparar el método 2 con el método 3, el método 2 presento siempre mejores valores de sensibilidad y especificidad, %FN, %FP, VPN y VPP.* ***Conclusión:*** *Si bien la incorporación de IA como DART mejora potencialmente la atención de pacientes, es imprescindible validar estas herramientas con estándares de referencia adecuados antes de su implementación. Esto garantiza decisiones clínicas basadas en parámetros confiables.*

Palabras clave: *Inteligencia Artificial; Retinopatía diabética; Telemedicina.*

Diabetes Mellitus (DM) is a systemic chronic disease characterised by sustained high blood sugar levels due to alterations in insulin action or secretion[1]. According to the World Health Organisation (WHO) and the International Diabetes Federation (IDF), the prevalence of DM is constantly increasing, with an estimated 642 million diabetics worldwide by the year 2040[2]. In Chile, the prevalence of DM in the population over 15 years old was 12.3% according to the 2016-2017 National Health Survey, showing a rising trend from 2.3% in 2003 to 9.4% in 2010[1,2].

Diabetic retinopathy (DR) is a complication of DM and is the third leading cause of blindness worldwide, particularly affecting individuals in their active working life[2,3,4]. International studies have shown that the prevalence of DR is directly related to the duration of diabetes and metabolic control. A study in 2012 involving 12,620 diabetic patients found a DR prevalence of 35.4%, with 11.7% at high risk of blindness[4,5]. In Latin America, the estimated prevalence of DR among diabetic patients is between 20% and 30%, with 5% at high risk of blindness[4]. In Chile, the estimated prevalence of DR ranges from 14.9% to 24.8% among diabetics[6]. DR is a progressive, bilateral, asymmetric microangiopathy that is asymptomatic in its early stages, highlighting the importance of early detection and treatment. Diagnostic methods for DR include direct fundoscopy, indirect fundoscopy, and retinography[1,7].

In Chile, there are three methods for diagnosing DR. Method 1 involves direct observation of the retina through a slit lamp by an ophthalmologist,

considered the gold standard for diagnosis. Method 2 utilizes retinal photography for evaluation and diagnosis by the ophthalmologist using retinography taken with a non-mydriatic camera. Method 3 involves the classification of retinographies by an artificial intelligence (AI) system called DART (Diagnóstico Automatizado de Retinografías Telemáticas). This program classifies retinographies as altered or non-altered, with only the altered ones being further evaluated by the ophthalmologist.

In recent years, artificial intelligence and deep learning have shown promising performance in various diagnostic applications, including medical image processing. In ophthalmology, AI has been a valuable tool for improving diagnostic accuracy, enhancing patient access to DR detection, and reducing costs. While several AI software for DR detection are FDA-approved for commercial use, challenges remain in accessing large image datasets for training and validation, as well as ensuring data confidentiality and compliance with regulations. Continuous improvements in AI performance, interpretation, and reliability are necessary[8,9]. Chile's implementation of the DART AI software in the public healthcare network in 2018 is considered a pioneering program in the country. According to information released by the Ministry of Health, the DART AI platform has been operating in the public health system since 2018, and its operational goal is to filter out up to 80% of retinographies without findings, alleviating the specialist's reading burden. However, the studies validated to date have shown variable results (sensitivity 40-100%, specificity 55-90%), indicating that its actual performance is still under evaluation and depends on the clinical scenario[11].

The validation of the DART software reported sensitivity of 94.6% and specificity of 74.3%[12] compared to retinographies classified by ophthalmologists, without direct fundoscopy, the gold standard method. This evaluation may lead to inaccuracies in sensitivity and specificity values due to the absence of the correct reference standard. Incorrect gold standards can result in patient misclassifications, affecting sensitivity and specificity values. By comparing the test with the correct gold standard, deviations can be corrected to determine accurate sensitivity and specificity values[13]. This study aims to correct the sensitivity and specificity of the DART software for diagnosing diabetic retinopathy by addressing misclassifications using the appropriate reference standard.

1. Methods

For the analysis, the number of cases was obtained from the information reported by DART in "Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system"[12]. The sensitivity and specificity values self-reported by the developers of DART (TeleDx, Chile), published in their clinical validation study (sensitivity 94.6%; specificity 74.3%), were used. Since these results were obtained by the team that created and markets the tool and have not been independently replicated, we consider them to be self-reported and subject to potential authorship bias. Considering self-reported DART sensitivity and specificity for the IA retinography method, the conditional probability method was applied for three scenarios (optimistic, moderate, and pessimistic). Then, the corrected sensitivity and specificity for each case were calculated.

1.1 Diabetic Retinopathy Diagnosis

For the case evaluation, the three procedures currently established in Chile to diagnose diabetic retinopathy were utilised:

Method 1: Fundoscopy by slit lamp observation, considered the gold standard method, consists of the medical evaluation by an ophthalmologist who asses and defines the diagnosis by evaluating the patient`s retina directly.

Method 2: Non-mydriatic retinography: The ophthalmologist assesses the diagnosis by evaluating the patient`s retina photography taken with a non-mydriatic retinography, an examination performed by a medical technologist in ophthalmology trained in this technique.

Method 3: Non-mydriatic retinography analysed by DART. The AI software evaluates the retinographies and classifies them as normal or altered. Only the retinographies considered by the software as altered are further analysed, in a second instance, by an ophthalmologist.

1.2 Simulated case scenarios

The sensitivity for Method 2 versus the gold standard was evaluated in the following three scenarios:

A. Optimistic: 90% sensitivity.
B. Moderate: 80% sensitivity.
C. Pessimistic: 70% sensitivity.

In Chile, currently, there is no information regarding sensitivity for DR detection with the mentioned methods. Therefore, the sensitivity for each scenario was selected based on reports from studies made elsewhere. Overall, those studies reported sensitivity for method 2 vs method 1 as high as 90% to 95%; meanwhile, the worst sensitivities were between 65% to 75%[14,15].

1.3 Combined sensitivity estimation

The combined sensitivity (S) for DART software was calculated based on the DART-reported sensitivity and each proposed scenario. Additionally, for the estimation, 18% DR prevalence was considered. As a result, the corrected DART sensitivity calculation was as follows:

$$S\ (3vs1) = S\ (2vs1) \times S(3vs2)$$

Where:

- S (3vs1)= S (2vs1)x S(3vs2)
- S (2 vs 3)= Data from DART publication
- S (2 vs 1)= sensitivity assumption for each scenario
- S (3 vs 1)= corrected DART sensitivity.

The contingency tables were constructed with the estimated number for each category (Table 1). Based on the calculated data, the specificity, false negative percentage (FN%), false positive percentage (FP%), negative predictive value (NPV) and positive predictive value (PPV) were estimated with the following formula:

$$Specificity = [B / (B + D)] * 10$$

$$FN\ \% = [C / (A + C)] * 100$$
$$FP\% = [B / (B + D)] * 100$$
$$NPV = [D / (C + D)] * 100$$
$$PPV = [A / (A + B)] * 100$$

Results

Utilising the data from DART validation, it calculated the values for the three mentioned scenarios. Thus, a 1,123 diabetic eye sample was obtained, with a 94.6% sensitivity and a 74.3% specificity. These values corresponded to the comparison of AI software versus the non-mydriatic retinography (method 2).

The DART clinical validation (Table 2) considered 1.123 subjects and reported 226 individuals as altered; meanwhile, method 2 identified 239. Therefore, DART presented 13 false negatives and 227 false positives. On the other hand, the AI classified without DR 657 subjects from 884 that were identified by method 2 (retinography) with a 94.6% sensitivity and 74.3% specificity. With these values, the three sensitivity scenarios were evaluated.

Table 1. Contingence table utilized to estimate de parameters evaluated (Sensitivity, Specificity, FN% (false negative percentage), FP% (false positive percentage), PPV (positive predictive value) and NPV (negative predictive value).

Test result		**Method 2**		
		DR +	**DR -**	**Total**
Method 3	DR +	A	B	A + B
	DR -	C	D	C + D
	Total	A + C	B + D	A + B + C + D

Table 2. DART results. Data reported by DART validation study published in "Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system"[12].

Test result		Method 2		
		DR +	DR -	Total
Method 3	DR + DR -	226 13	227 657	453 670
	Total	239	884	1123

Optimistic simulated scenario

Considering the sensitivity and prevalence previously established for the simulation, the values for methods 2 and 3 compare with the gold standard. Thus, method 2 appropriately classified 182 subjects from 202 with DR, according to the gold standard, and presented 20 false negatives and 57 false positives. Additionally, method 2 adequately identifies 864 of 921 patients as having no DR, resulting in an estimated specificity for method 2 of 93.8% compared to method 1. Moreover, the method 3 versus 1 comparison obtained an 85.14% sensitivity and 69.7% specificity. From the previous calculations, 172 of 202 patients were correctly determined by the AI, having 30 false negatives and 281 false positives. In addition, DART classifies accurately as without DR 640 from 921 subjects (Table 3).

Moderated simulated scenario

For the moderate scenario (Table 3), considering 80% sensitivity, it was calculated the values for method 2 versus method 1 were calculated. Thus, method 2 correctly classified 162 altered patients, having 40 false negatives and 77 false positives. Moreover, this procedure accurately classified as no DR 844 subjects from 921, estimating a specificity of 91.6% for method 2. Afterwards, method 3 was analysed against method 1, obtaining a 75.7% sensitivity and 68.0% specificity estimation. With this information, from 202 patients altered, the AI classified correctly just 153 subjects, reporting 49 false negatives and 300 false positives.

Pessimistic simulated scenario

In the calculations, a sensitivity of 70% is assumed according to the previously established. Under these circumstances, method 2 accurately classified 141 from 202, with 61 false negatives and 98 false positives. Moreover, method 2 correctly classifies 823 subjects as no DR from 921, reporting a specificity of 89.4% (Table 3). Finally, method 3 versus method 1 were analysed under a pessimistic scenario, obtaining a sensitivity and specificity of 66.2% and 66.4%, respectively. Therefore, the AI correctly classified as altered 134 subjects from 202, presenting 68 false negatives and 319 false positives (Table 3).

Comparison between estimated values and DART report

In the optimistic scenario (A), method 2 obtained a sensitivity and specificity of 90% and 93.8%, respectively. Though the false negatives and positives were 9.9% and 6.2%, and negative and positive predicted values were 97.7% and 76.15%, respectively. Although method 3 presented 85.14% sensitivity, 69.7% specificity, 14.85% false negatives, 30.51% false positives, 95.5% NPV and 37.96% PPV.

In the moderate scenario (B), method 2 presented 80.0% sensitivity and 91.6% specificity, with false negatives and positives of 19.80% and 8.36%, respectively. Additionally, it had 95.47% NPV and 67.78% PPV. On the other hand, method 3 has 75.7% sensitivity, 68.0% specificity, 24.25% FN, 32.57% FP, 92.68% NPV and 33.77% PPV.

Finally, in a pessimistic scenario (C), method 2 obtained a sensitivity of 70.0% and 89.4% specificity. Moreover, the false negatives and positives were 30.19% and 10.64%, respectively; meanwhile, the NPV and PPV were 93.09% and 58.99%, respectively. In contrast, method 3 presented 66.2% sensitivity, 66.4% specificity, 33.66% FN, 34.54% FP, 89.85% NPV and 29.58% PPV (Table 4).

Comparing the previous results (Table 4), retinography (method 2) showed better indicators than

Table 3. Results for each method and scenario. The table presents the values obtained for each of the three analyzed methods, considering the assumptions of each of the evaluated scenarios.

Test result		**Scenario A (Optimistic) Method 1**			**Scenario B (Moderated) Method 1**			**Scenario C (Pessimistic) Method 1**		
		DR +	**DR -**	**Total**	**DR +**	**DR -**	**Total**	**DR +**	**DR -**	**Total**
Method 2	DR +	182	57	239	162	77	239	141	98	239
	DR -	20	864	884	40	844	884	61	823	884
	Total	202	921	1123	202	921	1123	202	921	1123
Method 3	DR +	172	281	453	153	300	453	134	319	453
	DR -	30	640	670	49	621	670	68	602	670
	Total	202	921	1123	202	921	1123	202	921	1123

Table 4. Method 2 and 3 comparison versus the gold standard. The table presents the Sensitivity, Specificity, FN% (false negative percentage), FP% (false positive percentage), NPV (negative predictive value), and PPV (positive predictive value) of methods 2 and 3 in relation to the appropriate reference standard (method 1). The McNemar test was used to assess whether there were statistically significant differences, considering a p-value less than 0.05 as significant.

Scenario	**Indicator**	**Method 2**	**Method 3**	**p-value**
A (Optimistic)	Sensitivity	90.0%	85.14%	p= 0.000*
	Specificidad	93.8%	69.7%	p= 0.000*
	FN %	9.9%	14.85%	p= 0.000*
	FP %	6.2%	30.51%	p= 0.000*
	NPV	97.73%	95.5%	p= 0.000*
	PPV	76.15%	37.96%	p= 0.000*
B (Moderated)	Sensitivity	80.0%	75.7%	p= 0.000*
	Specificidad	91.6%	68.0%	p= 0.000*
	FN %	19.80%	24.25%	p= 0.000*
	FP %	8.36%	32.57%	p= 0.000*
	NPV	95.47%	92.68%	p= 0.000*
	PPV	67.78%	33,77%	p= 0.000*
C (Pessimistic)	Sensitivity	70.0%	66.2%	p= 0.000*
	Specificidad	89.4%	66.4%	p= 0.000*
	FN %	30.19%	33.66%	p= 0.000*
	FP %	10.64%	34.63%	p= 0.000*
	NPV	93.09%	89.85%	p= 0.000*
	PPV	58.99%	29.58%	p= 0.018*

DART (method 3), independently of the simulated scenario, with values that were significant in all the parameters evaluated. In addition, comparing method 3 reported by DART validation versus method 3 values from the present study, there were significant differences in sensitivity, specificity, false negative percentage, false positive percentage, NPV and PPV in all parameters for all the scenarios (Table 5).

Discussion

The DART platform has revolutionised DR detection in Chilean primary healthcare since its incorporation in 2018. This AI software allows the evaluation of DR through the image processing of retinographies taken with a non-mydriatic camera. The UAPOS are the units in charge of DR screening in primary healthcare in Chile, where a medical technologist in ophthalmology is responsible for taking the retinal image and submitting it to the platform for further analysis.

The DART validation paper reported parameters with 94.6% sensitivity and 74.3% specificity, which supported its incorporation as a useful DR study tool[12]. However, these parameters did not correspond to a proper test validation because they were not calculated by contrasting against the gold standard. Hence, in the present study, we re-evaluate the DART publication data, correcting the information and obtaining lower values for both metrics, sensitivity and specificity, in all the proposed scenarios, even in the optimistic one. These results could explain the problems that DART has presented in its implementation, where the direct users manifest dissatisfaction with the platform's performance due to the high frequency of false positives and negatives. According to data collected by ATEMOOCH (Medical Technologist in Ophthalmology and Optometry Association), 63.2% of medical technologists consider DART information unclear, while 79.4% of them think DART needs to provide more data. These notable differences with the manufacturer numbers rely on the reference standard used for sensitivity and specificity calculation. Thus, while the original paper compared the software results versus the

Table 5. Method 3 reported by DART compared to the gold standard for each scenario. The table presents the Sensitivity, Specificity, FN% (false negative percentage), FP% (false positive percentage), NPV (negative predictive value), and PPV (positive predictive value) obtained by method 3 for scenarios A, B, and C and compares them with those reported by DART. The McNemar test was used to assess whether there were statistically significant differences, considering a p-value less than 0.05 as significant.

	Sensitivity	Specificity	FN%	FP%	NPV	PPV
DART (2 vs 3)	94.6%	74.3%	5.44%	25.67%	98.05%	49.88
A (1 vs 3)	85.14% (p= 0.000)	69.7% (p= 0.000)	14.85% (p= 0.000)	30.51% (p= 0.000)	95.5% (p= 0.000)	37.96% (p= 0.006)
B (1 vs 3)	75.7% (p= 0.000)	68.0% (p= 0.000)	24.25% (p= 0.000)	32.57% (p= 0.000)	92.68% (p= 0.000)	33.77% (p= 0.000)
C (1 vs 3)	66.2% (p= 0.000)	66.4% (p= 0.000)	33.66% (p= 0.000)	34.63% (p= 0.000)	89.85% (p= 0.000)	29.58% (p= 0.000)

retinography evaluation by an ophthalmologist, in this study, we compared the software`s results with the gold standard method, which is the direct retina evaluation by the ophthalmologist.

In light of our results, the DART technique seems similar to the previous procedure for the DR study. Therefore, if DART is not better at detecting RD, it seems to be no justification for using the new method. Accordingly, method 2 in clinical practice appears to be more appropriate and agrees with the statement recommended by Ibáñez-Bruron et al. They found significantly better sensitivity for medical technologists than DART in the retinography analysis[3]. Even though the DART software did not present better sensitivity and specificity than method 2, it was proposed to use it to improve the DR coverage. However, no improvement has been observed, as Silva-Jorque and Zett reported in their research, since its implementation in 2018, the fundus coverage with DART`s implementation just changed from 36.03% in 2017 to 35.74% and 36.45% in 2018 and 2019, respectively[6]. One possible explanation for this phenomenon could be the necessity of someone who takes the images. Hence, even when the ophthalmologist is not needed, this technique is still operator-dependent because it is limited by the medical technologist's capacity. Therefore, DART does not dismiss the operator-dependency factor; it only changes the subject needed. As a result, the dilemma remains because AI incorporation constitutes an evolution in Chilean medicine; nonetheless, method 2 improvement seems more cost-effective to improve DR coverage.

As was mentioned, the coverage of retinal evaluation with DART depends on the number of hours dedicated to retinography acquisitions. Even though the AI algorithm classifies the patients, the platform needs photography. Currently in Chile, the majority of retinographies are performed at the primary healthcare under the UAPO program. There, refraction evaluations, glaucoma exams and other clinical evaluations are realised, among other responsibilities such as administration and management[16], all limiting the time destined for retinal evaluation. Another DART limitation is related to economic factors, where the bidding process needed to work with the software limits the number of photographs to a certain amount, which is insufficient to evaluate 100% of diabetic patients. A 2021 study established that 334,185 of 916,952 diabetic patients had valid actualised fundoscopy to date during 2019, which implies just 36.45% of all diabetic patients[6]. From a clinical point of view, DART setting analysis focuses on identifying retinal alterations characteristic of diabetic retinopathy, being unable to detect other retinal alterations or pathologies. In addition, DART classifies the patients as altered and not altered only under DR criteria. For instance, if an image presents an optic nerve head with a 0.7 excavation, the patient is considered a glaucoma suspect; in this case, the patient needs a thorough examination and an evaluation by the ophthalmologist. However, DART is not programmed to detect that kind of alteration, and as a result, that patient would be considered normal and sent home. Thus, these misdiagnoses can cause terrible damage to people's visual health, causing even dramatic repercussions such as blindness[17,18].

Finally, even though incorporating new technologies in healthcare benefits the patient's attention and clinicians, it is crucial to set clear goals. Before incorporating a new procedure, an appropriate evaluation must be performed to determine the grade of improvement offered to the health system. In the same order, addressing the implementation topics before the use could guarantee the results reported by the creators. For example, in the retinography case, the use of pupil dilatation, as was previously reported by our research group. Other examples are technical issues such as internet access, a national standardised image acquisition use and the operator-dependency problem, resolved with more medical technologists and ophthalmologists. Finally, by addressing the implementation issues in conjunction with the software application, telemedicine would result in improvement and, as a result, would enhance the healthcare system.

To conclude, in the present study, the sensitivity and specificity were estimated, assessing

more realistic values because it was considered the real gold standard. Even when these values were calculated based on assumptions, the results seem to agree more with the performance reported by the clinician's users than those reported by the software validation. Nonetheless, the present study intended to solve a possible miscalculation due to an incorrect standard. However, the correct procedure to obtain the true values is to perform an investigation in which the methodology incorporates the simultaneous evaluation made by the three mentioned methods. Therefore, to validate correctly, adequate standards must be considered in the methodology during the software evaluation process to guarantee the applicability reported.

Acknowledgements

The authors appreciate the support provided by Cristian Cartes, MD, Ophthalmologist, for his recommendations and support in this research and to Beatriz Cancino-Faure, PhD, researcher in Microbiology and Parasitology Laboratory, Medicine Faculty, Universidad Católica del Maule, for her unconditional support.

References

1. *Segovia C, Salinas D, Bassi R, Villagra V, Moraga C, Veas R. Implementación de cámara no midriática en tamizaje de pacientes diabéticos en una unidad de atención primaria oftalmológica. Iris: The International Journal of Research in Eye Sciences. 2022; 1(3): 17-22.*
2. *Flores R, Donoso R, Arenas-Cavalli JT, Anguita R. Modelo de manejo en red y por telemedicina de la retinopatía diabética en dos comunas del Servicio de Salud Metropolitano Oriente. Revista Médica de Chile. 2019; 147(4): 444-450.*
3. *Cruzat A, Órdenes-Cavieres G, Coria M. Exactitud de tamizaje de retinopatía diabética: Inteligencia artificial versus tecnólogos médicos entrenados. Revista Médica de Chile. 2021; 149(4): 493-500.*
4. *Covarrubias T, Delgado I, Rojas D, Coria M. Tamizaje en el diagnóstico y prevalencia de retinopatía diabética en atención primaria. Revista Médica de Chile. 2017; 145(5): 564-571.*
5. *Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, et al. Meta-Analysis for Eye Disease (META-EYE) Study Group. Global prevalence and major risk factors of diabetic retinopathy. Diabetes care. 2012; 35(3): 556-564.*
6. *Silva-Jorquera R, Zett C. Retinopatía Diabética en Chile: Un reporte cuantitativo de la proporción de diabéticos con fondo de ojo anual. Revista Médica de Chile. 2021; 149(7): 971-979.*
7. *Hernández Salazar L, Martínez Castellanos M, Saucedo Castillo A, Solís Vivanco A. Capítulo 2. Técnicas de Exploración del Polo Posterior. Retina y Vítreo. 2012; 2° Edición. El Manual Moderno S.A., México.*
8. *Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan GSW, Abramoff M, et al. Artificial intelligence for diabetic retinopathy screening: A review. Eye (Lond). 2020; 34(3): 451-460.*
9. *Straňák Z, Penčák M, Veith M. Artificial Intelligence in Diabetic Retinopathy Screening. A Review. Cesk Slov Oftalmol. 2021; 77(5): 224-231.*
10. *Tsiknakis N, Theodoropoulos D, Manikis G, Ktistakis E, Boutsora O, Berto A, et al. Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. Computers in biology and medicine. 2021; 135: 104599.*
11. *Santelices E. Ministro de Salud presenta software que permitirá triplicar la cantidad de exámenes para prevenir la ceguera diabética. 2018. https://www.minsal.cl/ministro-de-salud-presenta-software-que-permitira-triplicar-la-cantidad-de-examenes-para-prevenir-la-ceguera-diabetica/*
12. *Arenas-Cavalli JT, Abarca I, Rojas-Contreras M, Bernuy F, Donoso R. Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system. Eye (Lond). 2022; 36(1): 78-85.*
13. *Sherwin PF. Correcting Diagnostic Test Sensitivity and Specificity for Patient Misclassifications Resulting from Use of an Imperfect Reference Standard. Diagnostics. 2022; 13(1): 90.*
14. *Pérez-de-Arcelus M, Andonegui J, Serrano L, Eguzkiza A, Ramon Maya J. Diabetic retinopathy screening by general practitioners using non-mydriatic retinography. Current diabetes reviews. 2013; 9(1): 2-6.*
15. *Rosses AP, Ben ÂJ, Souza CFD, Skortika A, Araújo ALD, Carvalho GD, et al. Diagnostic performance of retinal digital photography for diabetic retinopathy screening in primary care. Family Practice. 2017; 34(5): 546-551.*
16. *Ministerio de Salud (MINSAL). Programa de Resolutividad. Orientaciones técnico - administrativas Unidad de Atención Primaria Oftalmológica (UAPO) y procedimientos quirurgicos de baja complejidad. Ministerio de Salud de Chile, Subsecretaria de Redes Asistenciales, División de Atención Primaria, Depto. de Modelo de APS. 2017.*
17. *Jayaram H, Kolko M, Friedman DS, Gazzard G. Glaucoma: Now and beyond. The Lancet. 2023; 402(10414): 1788-1801.*
18. *Geevarghese A, Wollstein G, Ishikawa H, Schuman JS. Optical coherence tomography and glaucoma. Annual review of vision science. 2021; 7: 693-726.*