Diabetic retinopathy classification for supervised machine learning algorithms

Luis Filipe Nakayama1*, Lucas Zago Ribeiro1, Mariana Batista Gonçalves1,2,3, Daniel A. Ferraz1,2,3, Helen Nazareth Veloso dos Santos1, Fernando Korn Malerbi1, Paulo Henrique Morales1,2, Mauricio Maia1, Caio Vinicius Saito Regatieri1 and Rubens Belfort Mattos Jr1,2

Abstract

Background: Artificial intelligence and automated technology were first reported more than 70 years ago and nowadays provide unprecedented diagnostic accuracy, screening capacity, risk stratification, and workflow optimization. Diabetic retinopathy is an important cause of preventable blindness worldwide, and artificial intelligence technology provides precocious diagnosis, monitoring, and guide treatment. High-quality exams are fundamental in supervised artificial intelligence algorithms, but the lack of ground truth standards in retinal exams datasets is a problem.

Main body: In this article, ETDRS, NHS, ICDR, SDGS diabetic retinopathy grading, and manual annotation are described and compared in publicly available datasets. The various DR labeling systems generate a fundamental problem for AI datasets. Possible solutions are standardization of DR classification and direct retinal-finding identifications.

Conclusion: Reliable labeling methods also need to be considered in datasets with more trustworthy labeling.

Keywords: Diabetic retinopathy classifications, Artificial intelligence, Datasets

Background

Computers executing automated functions were first described in 1950, with the first publication in 1943. Since then, Artificial Intelligence capacity has evolved into deep learning and neural networks, technologies that could simulate interconnected neurons and provide outputs after multiple information layers [1, 2].

Automated technology provides unprecedented diagnostic accuracy, screening capacity, risk stratification, and workflow optimization with accuracy equivalent to healthcare professionals [3] and more cost-effective diseases screening [4].

In Machine Learning, supervised learning is the most applied method in disease screening and classification algorithms, corroborating the importance of data labeling quality [5, 6].

Diabetic retinopathy (DR) is the leading cause of preventable blindness in working-age adults worldwide [7, 8], responsible for more than 24,000 annual cases of blindness [9] and the main focus in Ophthalmological AI screening algorithms [10]. There is an increased blindness risk in patients with chronic diabetes mellitus, especially those with poor clinical control [11].

Telemedicine and automated screening programs could diagnose, monitor, and guide treatment. Precocious diagnosis and therapy could avoid severe vision loss in 90% of cases, but only 60% of diabetic patients have recommended yearly examinations [12].

There are many Diabetic Retinopathy classifications applied in distinct countries and screening programs, with the International Council of Ophthalmology Diabetic Retinopathy (ICDR) classification as the most applied in open-access ophthalmological datasets [13].
High-quality retinal exams are fundamental in the development of AI algorithms, but also standards in labeling protocols, classifications, and quality control. This article describes and compares the most commonly diabetic retinopathy classifications, referencing criteria, and their applications in datasets.

Main text
This study compared the most often-applied DR classification scales: Scottish diabetic retinopathy grading [14], Early Treatment Diabetic Retinopathy Grading [15], International Clinic Diabetic Retinopathy [16], National Health Service diabetic retinopathy classification grading [17], Modified Davis retinopathy staging [18], and direct findings identification.

The Early Treatment Diabetic Retinopathy Study
At an international consortium of ophthalmologists at Airlie House in 1968, internists and neurosurgeons standardized a diabetic retinopathy classification applied in the landmark Early Treatment Diabetic Retinopathy Study [15], designed to generate a more precise staging for DR and macular edema. The study screened for the presence of microaneurysms (MA), retinal hemorrhages, cotton-wool spots, intraretinal microvascular abnormalities (IRMA), venous beading, and neovessels in 35-mm photographs. The consortium provided standard photos of microaneurysms, hemorrhages, and neovessels.

The ETDRS defined microaneurysms as red spots of less than 125 microns in its longest dimension with well-delimited margins and defined hemorrhage as a red spot with irregular margins with more than 125 microns. Punctate lesions, blots, linear hemorrhages, and microaneurysms were classified as red spots when they were not distinguished in ETDRS charts [19].

ETDRS defined clinically significant macular edema as retinal edema seen in retinal stereo photographs at or within 500 microns of the center of the macula or hard exudates at or within 500 microns of the foveal center and retina thickening or retinal thickening larger than one disc diameter area within one disc diameter of the center of the macula. In 2006, Rudnisky compared modified ETDRS protocols with one or two fields and 16:1 JPEG images and showed good reproducibility compared to standard ETDRS stereoscopic photos [20]. (Table 1).

National Health Service diabetic retinopathy classification
The National Health Service (NHS) was a diabetic retinopathy classification system applied in England, Scotland, Wales, and Northern Ireland between 2002 and 2007. It applied an ETDRS modified diabetic retinopathy scale classified in four severity stages [17, 21]. This program evaluated and classified DR using macula-centered and optic disc-centered images [22]. The NHS screening program provided guidelines for grading and lesions classifications [23].

This DR classification considered macular exudates sign of macular edema because the images were non-stereoscopic; it also added a photocoagulation classification (Table 1).

International Clinic Diabetic Retinopathy
The International Clinic Diabetic Retinopathy (ICDR) classification was published in 2003 after a consensus of 31 retina specialists, endocrinologists, and epidemiologists from 16 countries and sponsored by the American Academy of Ophthalmology [16]. ICDR classified DR on a five-stage severity scale and classified diabetic macular edema as apparently absent or present. The classification was created to simplify the ETDR and Wisconsin Epidemiologic Study scale and make it more applicable in daily practice studies [16].

ICDR is applied in the EYEPACS dataset [24], Asian Pacific Tele-Ophthalmology Society dataset [25], Indian Diabetic Retinopathy Image Dataset [26], Messidor 1 and 2 datasets [27] (Table 1).

The Scottish Diabetic Retinopathy Grading Scheme, 2004
In 2003, the National Scotland Eye Screening for Diabetic Retinopathy Program was created [28]. This grading system classified DR in all patients aged 12 years and older. Retinal digital photos were analyzed, and the re-screening period or ophthalmologist referral was established. The Scottish diabetic retinopathy grade (SDRG) is divided into four DR severities in a single fovea-centered image with at least two disc diameters temporal to the fovea and one disc diameter nasal to the disc [14] (Table 1).

Modified Davis retinopathy staging
The ICDR score simplifies DR in three stages: simple diabetic retinopathy, pre-proliferative retinopathy, and proliferative retinopathy using 45-degree photographs of the posterior pole applied in the Jichi DR dataset [18] (Table 1).

Direct findings identification
In AI datasets, findings such as microaneurysms, hemorrhages, hard exudates, and retinal detachment could be identified through direct identification. Applications such as SuperAnnotate [29], VGG Image annotation Tool [30], Supervise.ly [31], Labelbox [32], and Visual Object Tagging Tool [33] are available as labeling tools.

In ODIR [34], DIARETDB 0 and 1 [35], DR 1 and 2 [36], E-Ophtha [37], and HEI-MED [38], retinal findings are manually annotated (Fig. 1).
Table 1. Comparison of ETDRS, NHS, ICDR, SDRGS, Modified Davis diabetic retinopathy scales

| Grade         | ETDRS                  | NHS                     | ICDR | Scottish Diabetic Retinopathy Grading Scheme | Modified Davis |
|---------------|------------------------|-------------------------|------|---------------------------------------------|----------------|
| 0             | No red dot             | R0 - One or more isolated Cotton Wool Spots in the absence of microaneurysms or hemorrhage | No disease | No disease | No disease |
| 1             | Mild Non Proliferative DR | Mild Background - At least one microaneurysm, flame exudate, blot edema, and more than three small soft exudates | No disease | No disease | No disease |
| 2             | Moderate NPDR          | Moderate background - more than 4 blot hemorrhages in one hemifield | No disease | No disease | No disease |
| 3             | Severe NPDR           | Severe non-proliferative or pre proliferative DR - >4 blot hemorrhage in both hemifields, intraretinal microvascular anomalies, venous beading, and vitreous hemorrhage | No disease | No disease | No disease |
| 4             | Proliferative DR       | Early PDR - Neovessels and less than high-risk proliferative DR | No disease | No disease | No disease |
| 5             | Macular edema          | High-risk PDR- disc neovessels greater than photograph 10A or disc neovessels and vitreous hemorrhage / pre retinal hemorrhage | No disease | No disease | No disease |

Conclusions

Artificial intelligence and automated technology were first reported more than 70 years ago and nowadays provide unprecedented diagnostic accuracy, screening, risk stratification, and workflow optimization [3]. Reliable datasets are fundamental in supervised Machine Learning development; however, labeling process standardization, quality control, and homogenization remain challenging [39].

In diabetic retinopathy, there are distinct DR classifications, with different numbers of DR gradings and methods such as the Scottish Diabetic Retinopathy Grading [14], Early Treatment Retinopathy Grading [15], ICDR [16], NHS Diabetic Retinopathy Classification grading [17], and Modified Davis Retinopathy staging [18] that are described in this review. Still, direct retinal findings annotation is valuable in neural networks training.

The Scottish Diabetic Retinopathy Grading is a valuable classification through retinal photographs due to a single macular centered retinal evaluation and is more sensitive for grading moderate and severe cases than ICDR classification.

When choosing the classification method applied in the dataset, the image field of view and the number of images must be considered. Classical ETDRS and ICDR classifications tend to underestimate DR classification in retinal photographic images due to limited image view areas compared to retinal fundus examinations.

The various DR labeling systems generate a fundamental problem for AI datasets, and it is fundamental to standardize DR grading in datasets to develop algorithms and ensure proper patient referral. Reliable labeling methods also need to be considered in datasets with more trustworthy labeling.
Fig. 1 Direct retinal findings manual annotation example, in Labelbox software

Abbreviations
ETDRS: Early Treatment Diabetic Retinopathy Grading; NHS: National Health Service Diabetic Retinopathy Classification grading; IDC: International Council of Ophthalmology Diabetic Retinopathy; SDRG: Scottish Diabetic Retinopathy Grading; AI: Artificial intelligence.

Acknowledgements
We would like to thank Ophthalmology sectors from Escola Paulista de Medicina - São Paulo federal University and Instituto Paulista de Estudos e Pesquisas em Oftalmologia, IPEPO, Vision Instituto, São Paulo (SP), Brazil.

Authors’ contributions
LFN: Conceptualization, data curation, formal analysis, investigation, methodology, Project administration, writing—original draft, writing—review and editing. LZR: Conceptualization, formal analysis, writing—original draft, writing—review and editing. DF, HS, FKM, PHM, MM, CVSR: Conceptualization, methodology, writing-review and editing. RBJ: conceptualization, methodology, writing—review and editing. MBG: Conceptualization, methodology, writing—original draft, writing—review and editing. LZR: Conceptualization, formal analysis, investigation, methodology, supervision, writing—review and editing. All authors have read and approved the final manuscript.

Funding
M. B. Gonçalves is a researcher supported by Lemann Foundation, Instituto da Visão-IPEPO, São Paulo, Brazil and CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil.

Availability of data and materials
Not applicable.

Declarations
Ethics approval and consent to participate
UNIFESP Ethics Institutional Review Board Number: CAAE 33842220.7.0000.55 05/n0698/2020.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1 Physician, Department of Ophthalmology, Universidade Federal de São Paulo - EPM, Botucatu Street, 821, Vila Clementino, São Paulo, SP 04023-062, Brazil. 2 Instituto Paulista de Estudos e Pesquisas em Oftalmologia, IPEPO, Vision Instituto, São Paulo, SP, Brazil. 3 NIHR Biomedical Research Centre for Ophthalmology, Moorfield Eye Hospital, NHS Foundation Trust, and UCL Institute of Ophthalmology, London, UK.

Received: 28 September 2021   Accepted: 17 December 2021
Published online: 03 January 2022

References
1. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointest Endosc. 2020;92:807–12. https://doi.org/10.1016/j.gie.2020.06.040.
2. Muthukrishnan N, Maleki F, Ovens K, Reinhold C, Forghani B, Forghani R. Brief history of artificial intelligence. Neuroimaging Clin N Am. 2020;30:393–9. https://doi.org/10.1016/j.nic.2020.07.004.
3. Liu X, Faes J, Alayli M, Wagner SK, Di UV, Brynskens A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digital Health. 2019;1:e271–97. https://doi.org/10.1016/S2589-7500(19)30123-2.
4. Abràmoff 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 Digit Med. 2018;1:39. https://doi.org/10.1038/s41746-018-0040-6.
5. Khan SW, Liu X, Nath S, Koept E, Faes J, Wagner SK, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. Lancet Digital Health. 2020. https://doi.org/10.1016/S2589-7500(20)30240-5.
6. Lin W-C, Chen JS, Chang MF, Hribar MR. Applications of artificial intelligence to electronic health record data in ophthalmology. Transl Vis Sci Technol. 2020;9:13. https://doi.org/10.1177/2047988119892286.
7. Ophthalmology IC. Updated 2017 ICO guidelines for diabetic eye care. 2016.
8. Sadda SR. Assessing the severity of diabetic retinopathy: early treatment diabetic retinopathy study report number 10. Ophthalmology. 2020;127:597–8. https://doi.org/10.1016/j.ophtal.2019.11.028.
9. Md A, Abramoff 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. Yearb Paediatr Endocrinol. 2019. https://doi.org/10.1530/ypaendo-19-0021.
10. Kras A, Celi LA, Miller JB. Accelerating ophthalmic artificial intelligence research: the role of an open access data repository. Curr Opin Ophthalmol. 2020;31:337–50. https://doi.org/10.1097/ICO.0000000000000678.
11. Antonetti DA, Klein R, Gardner TW. Diabetic retinopathy. N Engl J Med. 2012;366:1227–39. https://doi.org/10.1056/NEJMr a1005 073.
12. Floris CI, Aidelman RA, Bailey ST, Fawzi A, Lim J, Vemulakonda GA, et al. Diabetic retinopathy preferred practice pattern®. Ophthalmology. 2020;127:P66–145. https://doi.org/10.1016/j.opth.2019.09.025.
13. Nakayama LF, Gonçalves MB, Ferraz DA, Santos HM, Malerbi FK, Morales PH, et al. The Challenge of Diabetic Retinopathy Standardization in an Ophthalmological Dataset. J Diabetes Sci Technol. 2021. https://doi.org/10.1177/19322968211029943.
14. Zachariah S, Wykes W, Ortonson D. Grading diabetic retinopathy (DR) using the Scottish grading protocol. Community Eye Health. 2015;28:72–3. https://www.ncbi.nlm.nih.gov/pubmed/27418727. Accessed 6 Sept 2021.
15. Solomon SD, Goldberg MF. ETDRS grading of diabetic retinopathy: still the gold standard? Ophthalmic Res. 2019;62:190–5. https://doi.org/10.1159/00050 1372.
16. Wilkinson CP, Ferris FL 3rd, Klein RE, Agardh CD, Davis M, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology. 2003;110:1677–82. https://doi.org/10.1016/j. ophtha.2003.04.075-5.
17. Scanlon PH. The english national screening programme for diabetic retinopathy 2003–2016. Acta Diabetol. 2015;52:15–26. https://doi.org/10.1007/ s00592-017-0974-1.
18. Takahashi H, Tampo H, Ara Y, Iwase Y, Kawashima H. Applying artificial intelligence to disease staging: deep learning for improved staging of diabetic
ETDRS Group. Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified Airlie house classification. ETDRS report number 10. Ophthalmology. 1991;98(8):796–806. https://doi.org/10.1016/S0161-6420(13)38012-9.

20. Rudnisky CJ, Tennant MTS, Weis E, Ting A, Hinz BJ, Greve MDJ. Web-based grading of compressed stereoscopic digital photography versus standard slide film photography for the diagnosis of diabetic retinopathy. Ophthalmology. 2007;114:1748–54. https://doi.org/10.1016/j.jophtha.2006.12.010.

21. Peate I. The NHS diabetic eye screening programme. Br J Healthc Assist. 2019;13:596–9. https://doi.org/10.12968/bjha.2019.13.12.596.

22. Diabetic eye screening: guidance when adequate images cannot be taken. 2021. https://www.gov.uk/government/publications/diabetic-eye-screening-programme-for-images-and-where-images-cannot-be-taken/diabetic-eye-screening-guidance-when-adequate-images-cannot-be-taken. Accessed 6 Dec 2021.

23. England PH. NHS Diabetic Eye Screening Programme grading definitions for referable disease. 2017. https://www.gov.uk/government/publications/diabetic-eye-screening-grading-guidance-referrable-disease. Accessed 6 Sept 2021.

24. Korot E, Guan Z, Ferraz D, Wagner SK, Zhang G, Liu X, et al. Code-free deep learning for multi-modality medical image classification. Nat Mach Intell. 2021;3:288–98. https://doi.org/10.1038/s42256-021-00305-2.

25. Khalifa NEM, Loey M, Taha MHN, Mohamed HNET. Deep transfer learning models for medical diabetic retinopathy detection. Acta Inform Med. 2019;27:327–32. https://doi.org/10.5345/aim.2019.27.327-332.

26. Pires R, Jelinek HF, Wainer J, Valle E, Rocha A. Advancing bag-of-visual-words independent ophthalmic disease recognition from fundus image using neural network. In: 2019 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON). 2019. doi:https://doi.org/10.1109/spicscon48833.2019.9065162.

27. Decencièrea E, Zhang X, Cazuguel G, Thibault G, Zhang Y, Garg S, Tobin K, et al. Indian diabetic retinopathy image dataset (idrid): a database for diabetic retinopathy screening research. Data. 2018;3:25. https://doi.org/10.3390/data3030025.

28. SuperAnnotate. 2020. https://superannotate.com/. Accessed 15 Jun 2021.

29. Labelbox: The leading training data platform for data labeling. 2021. https://labelbox.com/. Accessed 15 Jun 2021.

30. Visual object tagging tool (VoTT) v2.2.0. 2021. https://vott.z22.web.core.windows.net/. Accessed 15 Jun 2021.

31. Visual geometry group—University of oxford. 2021. https://www.robots.ox.ac.uk/~vgg/software/via/. Accessed 15 Jun 2021.

32. Supervise—Web platform for computer vision. Annotation, training and deploy. 2021. https://supervise.ly/. Accessed 15 Jun 2021.

33. Labelbox: The leading training data platform for data labeling. 2021. https://labelbox.com/. Accessed 15 Jun 2021.

34. Visual object tagging tool (VoTT) v2.2.0. 2021. https://vott.z22.web.core.windows.net/. Accessed 15 Jun 2021.

35. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Med Image Anal. 2012;16:216–26. https://doi.org/10.1016/j.media.2011.07.004.

36. Pires R, Jelinek HF, Wainer J, Valle E, Rocha A. Advancing bag-of-visual-words representations for lesion classification in retinal images. 2014. PLoS ONE. https://doi.org/10.6084/m9.figshare.953671.v1.

37. Decencièrea E, Cazuguel G, Zhang X, Thibault G, Klein J-C, Meyer F, et al. TeleOphta: machine learning and image processing methods for teleophthalmology. IRBM. 2013;34:196–203. https://doi.org/10.1016/j.irbm.2013.01.010.

38. Siancuardo L, Meriaudeau F, Kamovski TP, Ly, Garg S, Tobin KW Jr, et al. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. Med Image Anal. 2012;16:216–26. https://doi.org/10.1016/j.media.2011.07.004.

39. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AT, Raman R, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019;103:167–75. https://doi.org/10.1136/bjophthalmol-2018-313173.