Moannaei *et al.*
BioMedical Engineering OnLine (2025) 24:34
https://doi.org/10.1186/s12938-025-01336-1

BioMedical Engineering OnLine

RESEARCH Open Access

Performance and limitation of machine learning algorithms for diabetic retinopathy screening and its application in health management: a meta-analysis

Mehrsa Moannaei[1†], Faezeh Jadidian[10†], Tahereh Doustmohammadi[2†], Amir Mohammad Kiapasha[3†], Romina Bayani[3], Mohammadreza Rahmani[4†], Mohammad Reza Jahanbazy[9†], Fereshteh Sohrabivafa[5], Mahsa Asadi Anar[6*], Amin Magsudy[7], Seyyed Kiarash Sadat Rafiei[6] and Yaser khakpour[8]

[†]Mehrsa Moannaei and Faezeh Jadidian are contributed equally to the article and are co-first authors.

[†]Tahereh Doustmohammadi and Amir Mohammad Kiapasha have contributed to this work equally.

[†]Mohammadreza Rahmani and Mohammadreza Jahanbazy are contributed equally to the article and are co-fourth authors.

*Correspondence: Mahsa.boz@gmail.com

[6] Student Research Committee, Shahid Beheshti University of Medical Science, Arabi Ave, Daneshjoo Blvd, Velenjak, Tehran 19839-63113, Iran
Full list of author information is available at the end of the article

Abstract

Background: In recent years, artificial intelligence and machine learning algorithms have been used more extensively to diagnose diabetic retinopathy and other diseases. Still, the effectiveness of these methods has not been thoroughly investigated. This study aimed to evaluate the performance and limitations of machine learning and deep learning algorithms in detecting diabetic retinopathy.

Methods: This study was conducted based on the PRISMA checklist. We searched online databases, including PubMed, Scopus, and Google Scholar, for relevant articles up to September 30, 2023. After the title, abstract, and full-text screening, data extraction and quality assessment were done for the included studies. Finally, a meta-analysis was performed.

Results: We included 76 studies with a total of 1,371,517 retinal images, of which 51 were used for meta-analysis. Our meta-analysis showed a significant sensitivity and specificity with a percentage of 90.54 (95%CI [90.42, 90.66], $P<0.001$) and 78.33% (95%CI [78.21, 78.45], $P<0.001$). However, the AUC (area under curvature) did not statistically differ across studies, but had a significant figure of 0.94 (95% CI [– 46.71, 48.60], $P=1$).

Conclusions: Although machine learning and deep learning algorithms can properly diagnose diabetic retinopathy, their discriminating capacity is limited. However, they could simplify the diagnosing process. Further studies are required to improve algorithms.

Keywords: Machine learning algorithms, Artificial intelligence, Diabetic retinopathy, Meta-analysis, Deep learning

Introduction

Diabetic retinopathy (DR) has a significant role in vision problems and blindness among individuals in middle age and older populations worldwide [1, 2]. According to estimates, close to 103 million adults worldwide were diagnosed with DR in 2020. Experts project that this number will increase to approximately 160 million by 2045 [3]. It is believed that nearly one-third of people who have diabetes may eventually experience different forms of DR [4], and 10% of those with diabetes are at risk of vision loss [5]. DR consists of several stages: its initial stage includes non-proliferative DR (mild, moderate, and severe), which can progress to diabetic macular oedema or pre-proliferative DR. Pre-proliferative DR itself can lead to proliferative DR. There may be no symptoms in the early stages, or the severity of symptoms is very mild. In asymptomatic patients, the cotton wool and microaneurysms could be found incidentally in fundoscopy. Initial changes can be controlled with proper management. In contrast, the lack of appropriate treatment leads to the progression of DR to the final stages and symptoms, such as blurred vision, blurred vision, flutter, and partial or complete loss of vision [6, 7].

Vision problems resulting from DR have great potential to be considerably improved if detected at the initial stages and cured appropriately [8]. However, less than sixty percent of diabetic patients receive the recommended regular eye examinations due to the high cost and limited access to ophthalmological services [9]. As DR is a sneaky disease, many patients are not consciously aware that they have it, particularly those living in locations with inadequate medical resources. This makes it difficult for ophthalmologists to accurately determine a patient's condition based on fundus pictures [10]. DR's gold standard screening method involves clinical assessments by human clinicians or evaluating color fundus photographs remotely via telehealth services [11]. One of the primary modalities used is optical coherence tomography (OCT). OCT can generate sectional three-dimensional images of the retina's thickness and structure by measuring the amount of light reflection. This technique mainly detects macular oedema in DR [12]. However, these methods require a significant investment of time and effort. In addition, the likelihood of achieving irregular and inconsistent outcomes rises due to natural human subjectivity [13].

Hence, highly sensitive and specific automated systems are essential for widespread implementation of DR screening using color fundus photographs. The advancement of artificial intelligence (AI), particularly machine learning (ML), has made it possible to develop such automated approaches. Machine learning utilizes pre-existing data to instruct a computer on identifying a particular pattern or making predictions about a specific event in a novel data set [11]. The emergence of deep learning (DL), a subset of machine learning (ML), has significantly revolutionized the domain of automated image analysis [14]. In essence, DL approaches refer to representation learning techniques that utilize neural networks with multiple layers. These networks can achieve better performance by iteratively adjusting their internal parameters [11, 12]. Unlike other machine learning algorithms, deep learning does not necessitate image preprocessing or manipulation. After being provided with raw data, the system creates its own representations necessary for pattern identification. It has demonstrated higher accuracy compared to other traditional machine learning methods [11, 13].

Although the use of AI in the medical domain is being steadily developed [15], the impact and performance of AI diagnostic tools are not consistent across all research [16, 17], so questions remain regarding whether ML techniques have enough reliability to be used for actual medical needs, specifically in DR screening. Due to this heterogeneity among the studies, the present meta-analysis is conducted to update the performance and limitations of ML algorithms for DR screening and its application in health management.

Methods

In this systematic review, we intend to scrutinize the performance and limitations of a machine learning algorithm for diabetic retinopathy screening and its application in health management. The design protocol of this review was registered in the Open Science Framework (OSF: osf.io/3prs8).

Search strategy

Our methodology follows the PRISMA (preferred reporting items for systematic reviews and meta-analysis). It is critical to state that this manuscript's search strategy, screening, and data selection were all checklist-based. Databases, including PubMed (Medline), Scopus, and Google Scholar, were searched up to September 30, 2023, without any time restriction. The search strategy for each database is defined in Table 1. By searching the references of the found studies, we manually added studies with relevant titles that were not discovered during the database search. The studies were manually filtered to exclude non-diabetic studies. Duplicated studies were found automatically and manually by two independent researchers. After complete screening, studies compatible with our inclusion criteria were included.

Inclusion and exclusion criteria

In this study, all articles that explore a machine-learning or deep learning algorithm for diabetic retinopathy screening using color fundus photographs were included. In addition, we only used English papers. It is worth noting that the inclusion was not limited to specific types of diabetes (type 1 or 2) or age groups. However, all letters to editors, case reports, case series, posters, and abstracts were excluded.

Table 1 Search strategy for online databases

Database	Search strategy	Date
PubMed	(retinopathy[Title]) AND ((artificial intelligence [Title]) OR (AI[Title]) OR (machine learning [Title]) OR (ML[Title]) OR (deep learning [Title]) OR (DL[Title]))	September 30, 2023
Scopus	TITLE ((retinopathy AND ("artificial intelligence" OR "machine learning" OR "deep Learning")))	September 30, 2023
Web of science	TITLE ((retinopathy AND ("artificial intelligence" OR "machine learning" OR "deep Learning")))	September 30, 2023

Quality assessment and data extraction

The Joanna Briggs Institute's (JBI) checklist was used in our manuscript study quality assessment. Two reviewers evaluated the full text of the papers to exclude improper studies. If there was any dispute, it was resolved by consultation. Two other researchers extracted data, including author, year, country, study design, total image number, sensitivity, specificity, and area under the curve (AUC).

Statistical analysis

Data analysis was conducted using STATA 13.1 software, developed by StataCorp LP in College Station, TX, USA. Results were presented as pooled sensitivity and specificity with a 95% confidence interval (CI), visualized in a forest plot. Heterogeneity among included studies was assessed using the I^2 statistic, and the random effects model was applied in the presence of significant heterogeneity ($I^2 > 50\%$). In addition, we calculated the AUC. Eventually, the publication assessed using a funnel plot and Egger's regression test.

Result

Study selection and study characteristics

A total of 1861 studies were obtained by searching online databases, including PubMed, Scopus, Google Scholar, and manual search. Duplicated reports were removed, of whom 576 cases were done automatically. After screening the remaining articles, 969 irrelevant studies were excluded. Finally, 76 studies with a total of 1,371,517 images were included in the systematic review (Fig. 1). Of which, 51 studies were used for meta-analysis. The remaining 25 studies were excluded from the meta-analysis due to insufficient data. These studies were published between 2013 and 2023 (Table 2). Both private and public data sets were used in these studies. Our findings showed that studies suggested a range of approaches used in diabetic retinopathy screening, from traditional manual grading to advanced machine learning methods like convolutional neural networks (CNNs) and deep learning models. Some studies used smartphone-based imaging for more accessible screening, especially in resource-limited settings. Hybrid models combining different algorithms were also utilized, showing potential for improving screening processes (Table 2).

Meta-analysis

We performed a meta-analysis to assess the accuracy (including sensitivity and specificity) of machine learning detection in diabetic retinopathy in 51 studies. Our result revealed a significant sensitivity and specificity with a percentage of 90.54% (95%CI [90.42, 90.66], $P < 0.001$) and 78.33% (95%CI [78.21, 78.45], $P < 0.001$) (Figs. 2 and 3). We observed a severe heterogeneity between studies in both cases ($I^2 > 99\%$). In addition, we observed a substantial AUC with an amount of 0.94 (95% CI [− 46.71, 48.60]); however, it was not statistically significant ($P = 1$) (Fig. 4).

Publication bias

The funnel plot and Egger's test were performed to investigate the possible publication bias. The funnel plot showed a symmetrical pattern, indicating no publication bias (not shown). In addition, Egger's test supported this result.

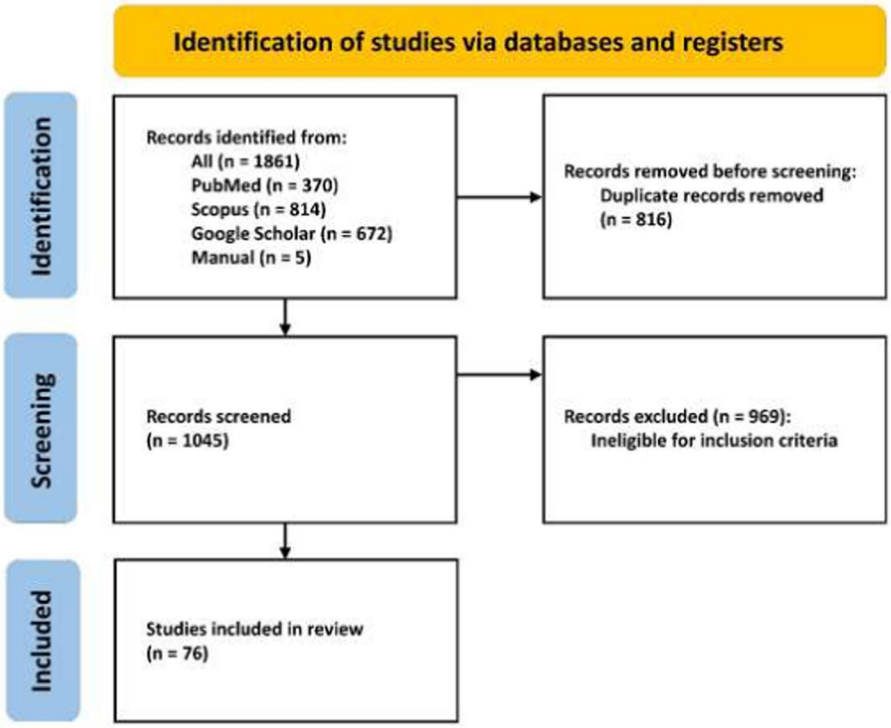

Fig. 1 PRISMA diagram for present study

Discussion

Incorporating ML algorithms into healthcare, particularly in DR screening, represents a significant advancement in medical diagnostics. Diabetic retinopathy stands as a major cause of blindness among adults globally, necessitating early detection and prompt management to prevent irreversible vision loss. With the emergence of machine learning and AI, the approach to DR screening and diagnosis has undergone a transformative shift, capitalizing on these technologies' immense potential to bolster accuracy, efficiency, and accessibility in healthcare [10, 18].

This meta-analysis consolidates existing evidence and assesses the diagnostic accuracy of ML algorithms in detecting DR using color fundus photographs.

Having completed a comprehensive meta-analysis, we have investigated the diagnostic accuracy of ML algorithms in detecting DR using color fundus photographs. Our analysis, which involved data from 76 studies encompassing 1,357,517 images, has concluded. The results, showing a sensitivity of 90.54% (95% CI [90.42, 90.66]), specificity of 78.33% (95% CI [78.21, 78.45]), and an area under the receiver operating characteristic curve (AUC) of 0.94, underscore the high diagnostic accuracy of ML models in identifying DR from retinal images. These findings suggest the capability of ML algorithms to detect the presence of DR accurately, a critical step toward preventing vision loss in diabetic patients.

The high sensitivity rate indicates that ML algorithms are proficient at identifying those individuals with DR, minimizing the risk of missed diagnoses. Meanwhile, the specificity rate reflects the algorithms' ability to correctly identify those without the

Table 2 Baseline characteristics of included studies and performance of color fundus image screening

Code	Author (year) [Ref.]	Total images	Sensitivity	Specificity	AUC	Approach for screening
2	Jain (2021) [26]	1294	100	89.55		Kowa VX-10a mydriatic camera & Remidio FOP NM-10
3	Keel (2018) [27]	96	92.3	93.7	0.95	DLA
4	Jiang (2020) [28]	3228	93.9	94.4	0.94	Grad-CAM
5	Ipp (2020) [29]	893	95.5	85		ETDRS scale
6	Ibáñez-Bruron (2021) [30]	89	100	55.4		DART
7	Yo-Ping Huang (2020) [31]	52	96.6	95.2	0.99	VGG16, VGG19, MobileNet, InceptionV3, DenseNet
8	Hsu (2021) [32]	13,410	96.84	89.44	0.97	DLA
9	Yi-Ting Hsieh (2019) [33]	7524	92.2	97.5	0.95	CNN (VeriSee)
10	Heydon (2020) [34]	30,405	95.7	68		EyeArt v2.1
11	He (2019) [35]	889	90.79	98.5	0.94	Airdoc
12	Hao (2022) [36]	6146	79.2	87.1		VoxelCloud
13	Guo (2021) [37]	978	54	95	0.88	ResNet & U-Net
14	Gulshan (2019) [38]	3049	83.5	98.7	0.96	ICDR scale
15	Gulshan (2016) [39]	4997	90.3	98.1	0.99	DLA
16	Grzybowski (2021) [40]	60	93.33	94.45	0.94	Retinalyze
17	González-Briceño (2020) [41]	3368	89	92		Cross-industry standard process for data mining
18	Gargeya (2017) [42]	75,137	94	98	0.97	DLA
19	Glinton (2022) [43]	597	91	95	0.93	Python (version 3.6.9)
20	Gadekallu (2020) [44]	1151	90.4	94.3		DLA
21	Fleming (2023) [45]	179,944	89.19	77.41	0.99	DLA
22	M. Al-hazaimeh (2022) [46]	88,702	99.2	96.4	0.98	SVMGA
23	TamoorAziz (2023) [47]	219	94.21	97.46	0.98	DLA
24	Ghadah Alwakid (2023) [48]	12,522	89	99		CLAHE, ESRGAN
25	Eman AbdelMaksoud (2022) [49]	3662	96	69	0.99	CNN
26	Marc Baget-Bernaldiz (2021) [50]	1200	97.92	99.91	0.99	DLA
27	Anas Bilal (2022) [51]	98	96.9	96.9	0.97	U-NET, CNN-SVD
28	Usharani Bhimavarapu (2023) [52]	88,702	96.34	96.74	0.89	CNN
29	Wejdan L. Alyoubi (2021) [53]	13,673	89	97.3	0.95	CNN512, YOLOv3
30	Miao (2022) [54]	35,126	79.01	89.07	0.79	DLA
31	Penha (2023) [21]	686	93.6	71.7	0.86	EyerMaps
32	Lee (2021) [55]	311,604	80.47	81.28		VA HCS
33	Lam (2018) [56]	1346	95	96		DIGITS
34	Nugroho (2021) [57]	200	95	81		DLA
35	Nneji (2022) [58]	35,126	98.9	98	0.99	WFDLN
36	Zhang (2022) [22]	1089	98.23	74.45	0.95	EyeWisdom V1
37	Yang (2021) [59]	1418	79.6	79.9	0.81	XGBoost, RF, naïve Bayes, KNN, AdaBoost, Light GBM, ANN, LR
38	Zhao (2022) [60]	7943	88.9	74	0.8	RF, XGBoost, LR, SVM, KNN
39	Pinedo-Diaz (2022) [61]	420	97.66	98.33	0.98	DLA
40	Surya (2023) [62]	1085	83.33	98.86	0.83	Dr Noon
41	Zhang (2020) [63]	47,269	83.3	92.5		DLA

Table 2 (continued)

Code	Author (year) [Ref.]	Total images	Sensitivity	Specificity	AUC	Approach for screening
42	Sosale (2020) [64]	922	93	92.5	0.9	Medios AI
43	Mehboob (2022) [65]	25,600	78	44		DLA
44	Mujeeb Rahman (2022) [66]	560	93.65	95.13	0.97	DNN, SVM
45	Abramoff (2016) [67]	1748	96.7	87	0.98	IDx-DR X2.1
46	Palaniswamy (2023) [68]	813	94.28	99.34	0.96	DLA
47	Ting (2017) [69]	71,896	90.5	91.6	0.93	DLA
48	Jebaseeli (2019) [70]	201	80.61	99.54		DLBSVM
49	Jena (2022) [71]	100	99.2	99.4	0.99	2-branch CNN
50	Jiang (2019) [72]	30,244	85.57	90.85	0.946	DLA
51	Khan (2023) [73]	45	79.63	98.63	0.98	Inception v3 & DenseNet-121
52	Shankar (2020) [74]	541	98.54	99.38		SDL
53	Kuna (2023) [75]	1200	98.9	99.7		DL-DRDC
54	Ludwig (2020) [76]	92,364	89	89	0.89	CNN
55	Sosale (2020) [77]	297	98.84	86.73	0.92	ICDRS scale
56	Li (2022) [23]	1674	95	85.1	0.94	Deep learning algorithm
57	Roy (2020) [78]	1330	94	95		DLA
58	Romero-Aroca (2020) [79]	1748	96.7	97.6		DLA
59	Pei (2022) [80]	324	91	81.3	0.86	EyeWisdom
60	Rayave (2022) [81]	650	65.54	100		CNN, RNN, SVM, FKM, DA
61	Paradisa (2020) [82]	89	99.3	98		CNN, SVM, KNN, RF, XGBoost
62	Li (2022) [83]	950	97.96	93.88	0.99	NNs, SVM, XGBoost, DT, LR, GNB, KNN
63	Roychowdhury (2013) [84]	1200	100	53.16	0.87	GMM, SVM, KNN, AdaBoost
64	Sarao (2020) [85]	165	90.8	75.3	0.07	EyeArt
65	Li (2021) [86]	32,452	70	90	0.9	LR, XGBoost, RF, SVM
66	Wu (2022) [87]	7033	100	37.8	0.9	OC-Net, SE-Net
67	Ruamviboonsuk (2022) [88]	138	91.4	95.4		DLA
68	Ruamviboonsuk (2019) [89]	25,326	97	96		DLA
69	Saxena (2020) [90]	56,839	81.02	86.09	0.92	CNN
70	Sayres (2019) [91]	1612	79.4	96.6		DLA
71	A. Shah (2021) [92]	2680	100	81.82	0.98	IDx-DR
72	P. Shah (2020) [93]	1533	99.7	98.5	0.99	CNN
73	Rajalakshmi (2018) [94]	296	95.8	80.2		ICDR
74	Reddy (2022) [95]	89	90.2	95.2	0.88	DLA
75	Rom (2022) [96]	363	45	94	0.81	CNN
76	Rogers (2021) [97]	22,180	81.6	81.7	0.98	Pegasus
77	Ryu (2022) [98]	918	67.5	94.4		CNN

DLA: Deep-learning algorithm; VA: Veterans Affairs; HCS: Puget Sound Health Care System; DIGITS: Deep Learning GPU Training System; WFDLN: weighted fusion deep learning network; XGBoost: extreme gradient boosting; KNN: k-nearest neighbour; ANN: artificial neural network; RF: Random Forest; LR: Logistic Regression; SVM: Support Vector Machine; DNN: deep neural network; DLBSVM: Deep Learning Based Support Vector Machine; SDL: Synergic deep learning; CNN: convolutional neural networks; ICDRS: International Clinical Diabetic Retinopathy Severity Scale System; RNN: Recurrent Neural Network; FKM: Fuzzy K-means cluster; DA: Discriminant Analysis; GNB: Gaussian Naive Bayes; LR: Logistics Regression; DT: Decision Tree; NNs: Neural Networks; GMM: Gaussian Mixture model; OC-Net: occurrence network; SE-Net: a severity network; DL-DRDC: w deep learning empowered diabetic retinopathy detection and classification

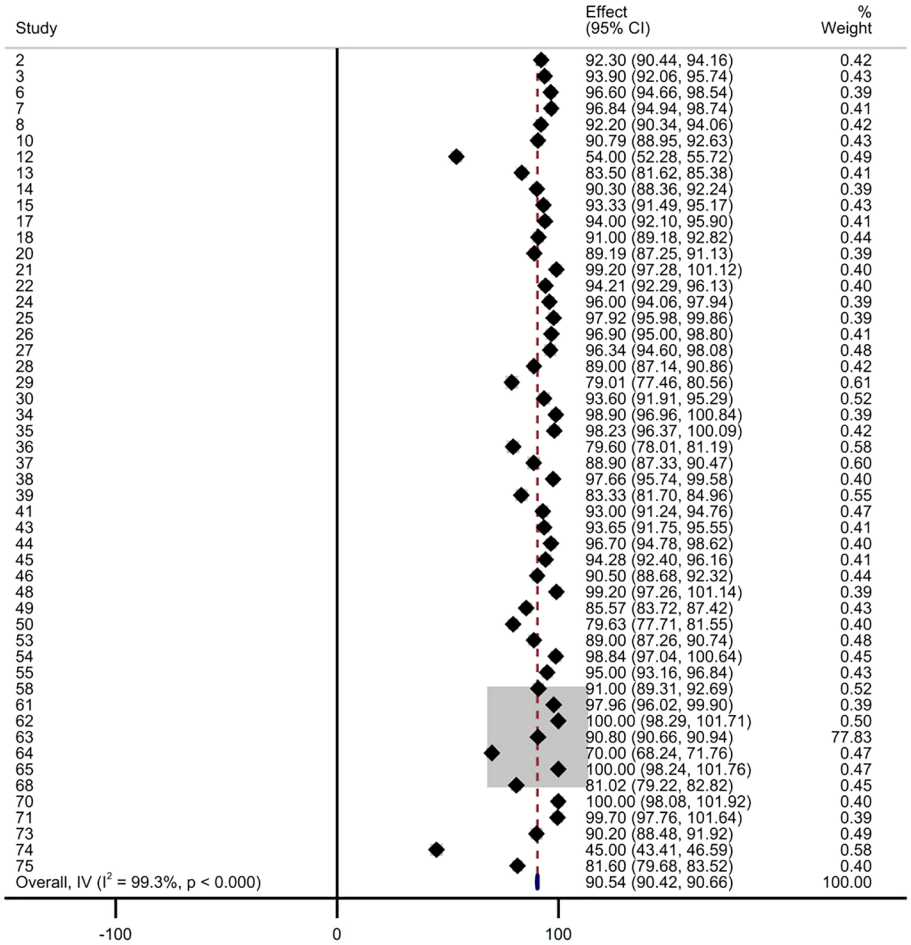

Fig. 2 Forest plot for sensitivity showed a significant amount of 90.54%

condition, although there is room for improvement to reduce false positives. The AUC, a measure of the algorithm's overall diagnostic ability, further confirms the efficacy of ML in DR screening, suggesting that these technologies can reliably distinguish between affected and unaffected individuals.

However, the implementation of ML in DR screening is not without challenges. The variation in specificity rates points towards the need for further refinement of algorithms to enhance their discriminative capacity, minimizing the occurrence of false positives that could lead to unnecessary anxiety or interventions for patients. In addition, the effectiveness of ML algorithms can vary based on factors such as image quality, the diversity of the data sets on which the algorithms are trained, and the prevalence of DR in the screened population. Ensuring that ML models are trained on diverse, high-quality data sets is crucial to enhancing their generalizability and accuracy across different populations and settings.

Moreover, integrating ML algorithms into clinical practice necessitates a multidisciplinary approach involving technologists, data scientists, clinicians, patients, and policymakers. The meta-analysis of ML algorithms for DR screening represents a significant step forward in applying AI in health management. The high diagnostic accuracy of

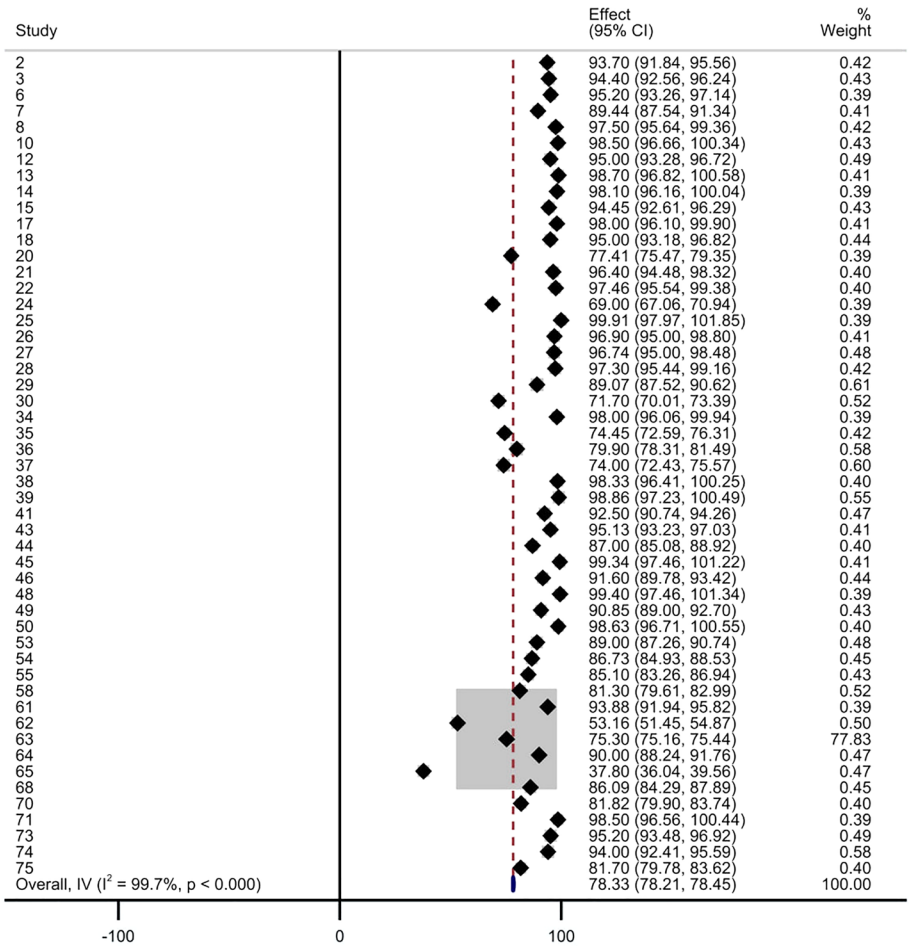

Fig. 3 Forest plot for specificity showed a significant amount of 78.33%

these models holds the promise of revolutionizing DR screening, making it more accessible, efficient, and effective. However, realizing this potential requires addressing the limitations and challenges associated with the deployment of ML in healthcare. ML can play a significant role in transforming healthcare delivery, and improving patient outcomes through continuous refinement of algorithms, adherence to ethical standards, and collaboration across disciplines.

The results show diverse techniques utilized across the studies, reflecting the evolving landscape of approaches to screening diabetic retinopathy, where different algorithms cater to specific needs in clinical settings. Traditional methods, such as manual grading by trained professionals, are contrasted with advanced machine learning techniques, including CNNs and deep learning models, which have demonstrated remarkable improvements in sensitivity and specificity, underscoring the potential of artificial intelligence to enhance diagnostic accuracy. Several studies adopted smartphone-based imaging systems, highlighting a shift towards more accessible screening methods, particularly significant in resource-limited settings where conventional imaging equipment may not be available. The integration of AI with portable devices facilitates rapid screening and timely referrals, potentially improving patient outcomes. Furthermore, the table

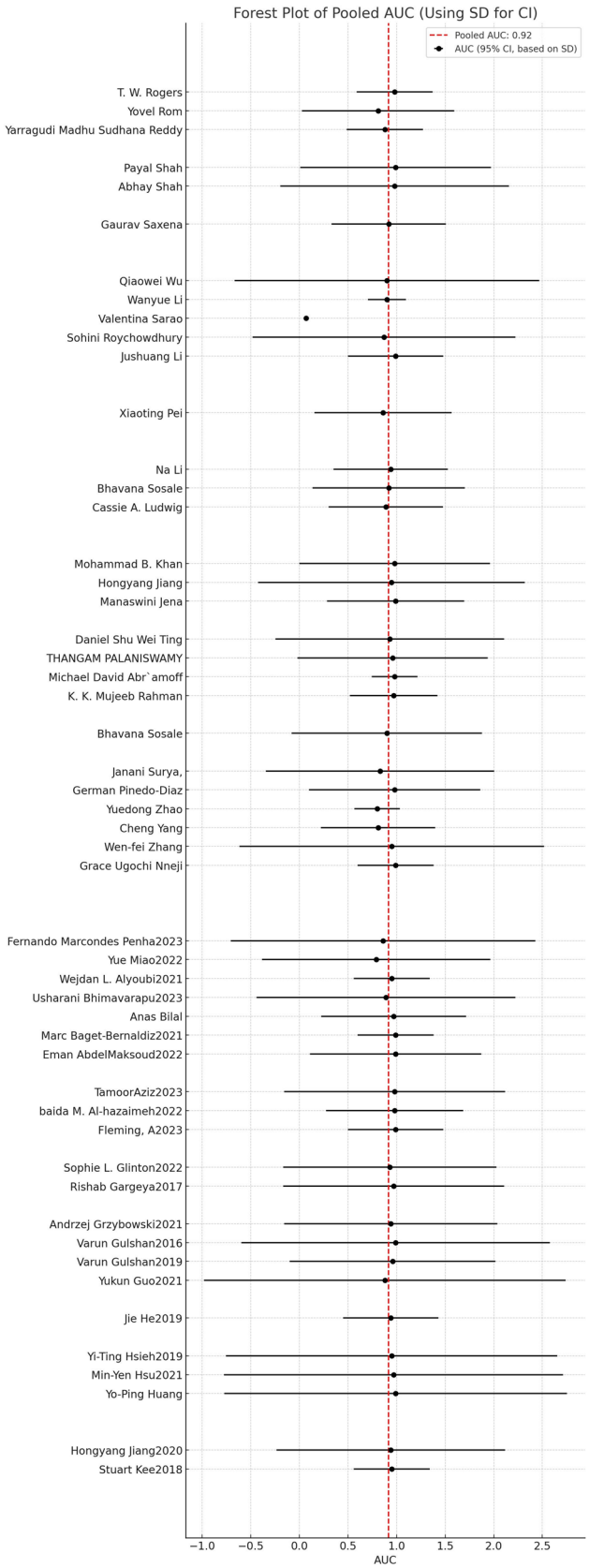

Fig. 4 Forest plot for area under the curve (AUC) showed the amount of 0.94; however, it was not significant

indicates that some studies utilized hybrid models, combining various algorithms to leverage their strengths, which may enhance the robustness of screening processes and address limitations often encountered with single-algorithm applications; the effectiveness of these hybrid models, as evidenced by their reported performance metrics, suggests a promising avenue for future research and implementation.

Furthermore, Wang et al. conducted a meta-analysis utilizing multiple algorithms to assess the diagnostic efficacy of DR. Their study included 21 original studies involving 129,759 eyes. The pooled sensitivity, specificity, and area under the curve (AUC) of the AI model for diagnosing DR were reported as 0.880 (0.875–0.884), 0.912 (0.99–0.913), and 0.9798, respectively [10]. Similarly, Wu et al. conducted a meta-analysis utilizing machine learning algorithms for DR screening, encompassing 60 color fundus photograph studies involving 445,175 interpretations. Their study reported high ML accuracy in diagnosing various categories of DR, with a pooled AUROC ranging from 0.97 (95% CI 0.96–0.99) to 0.99 (95% CI 0.98–1.00) and a pooled sensitivity and specificity ranging from 0.93 to 0.97 and 0.90 to 0.98, respectively. They concluded that ML algorithms' performance detecting DR based on color fundus photographs is likely comparable to human clinicians [19]. Ryu et al. developed an end-to-end deep learning-based classification system for DR and referable DR diagnoses using optical coherence tomography angiography (OCTA) images. They achieved high accuracy, sensitivity, and specificity for detecting the onset of DR and referable DR, further supporting the efficacy of ML algorithms in DR detection [20]. In their study, Penha et al. utilized an AI system, which included 686 individuals. Their findings demonstrated high sensitivity for DR screening using only one image per eye, suggesting a simpler protocol than the traditional approach [21].

In addition, Zhang et al. conducted a prospective, multi-center clinical trial study utilizing AI software to diagnose DR. Their study reported high sensitivity for DR detection compared to manual grading, highlighting the potential of AI in improving DR diagnosis [22]. Li et al. surveyed using deep learning algorithms (DLA). The DLA graded retinal fundus images; for all 1674 gradable images the AUC, sensitivity, and specificity of the DLA for referable DR were 0.942, 85.1%, and 95.6%, respectively [23]. Besides that, Li et al. conducted a study using DLA. They achieved an AUC, sensitivity, and specificity of 0.955, 92.5%, and 98.5%, respectively, for detecting referable DR in their independent multiethnic data sets [24]. Finally, Joseph et al. conducted a survey encompassing 34 studies utilizing AI algorithms for diagnosing DR based on real-world fundus images. Their study reported overall pooled accuracy, sensitivity, and specificity, further emphasizing the potential of AI in DR diagnosis [25].

One drawback of our analysis is that we had limited access to the data sets and the complete text of articles [18]. Furthermore, we exclusively incorporated materials written in English. The collected data on DR did not include the proliferative diabetic subtype or other categories of DR, which could affect the assessment of its diagnostic usefulness [10].

In addition, bias may have been incorporated by poor description of patient characteristics in the included studies. By the reason of spectrum bias, there is a possibility of ML's performance being overestimated in real-world scenarios, and this factor should be considered [19]

It is reasonable for another complementary study to initiate using this algorithm in primary care settings; other applications and research are necessary to improve the this algorithm's clinical validity [24].

In conclusion, machine learning algorithms can potentially diagnose diabetic retinopathy using retinal images. Although this capability was insignificant in discriminating true positives from false positives, it could be used for faster and better evaluation. However, further studies are required to develop improved algorithms.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12938-025-01336-1.

Supplementary Material 1.

Acknowledgements
We appreciate all the authors of included studies.

Author contributions
Study design and conception: M.AA; Search and study selection: M.M, F.J; Data extraction: T.D, AM.K; Quality assessment: R.B, M.R; Statistical analysis and interpretation: MR.J, F.S; Drafting the manuscript: M.M, F.J,A.M, SKSR, YK, T.D, AM.K, R.B, M.R, MR.J, F.S, M.AA; Critical revision: M.AA All authors were approved the submitted version.

Funding
None.

Availability of data and materials
The datasets generated and analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]School of Medicine, Hormozgan University of Medical Sciences, Bandar Abbas, Iran. [2]Department and Faculty of Health Education and Health Promotion, Student Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [3]Student Research Committee, School of Medicine, Shahid Beheshti University of Medical Science, Tehran, Iran. [4]Student Research Committee, Zanjan University of Medical Sciences, Zanjan, Iran. [5]Health Education and Promotion, Department of Community Medicine, School of Medicine, Dezful University of Medical Sciences, Dezful, Iran. [6]Student Research Committee, Shahid Beheshti University of Medical Science, Arabi Ave, Daneshjoo Blvd, Velenjak, Tehran 19839-63113, Iran. [7]Faculty of Medicine, Islamic Azad University Tabriz Branch, Tabriz, Iran. [8]Faculty of Medicine, Guilan University of Medical Sciences, Rasht, Iran. [9]Student Research Committee, Isfahan University of Medical Sciences, Isfahan, Iran. [10]School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Received: 18 April 2024 Accepted: 7 January 2025
Published online: 14 March 2025

References

1. Tan T-E, Wong TY. Diabetic retinopathy: looking forward to 2030. Front Endocrinol. 2023;13:1077669.
2. Nebbioso M, et al. Diabetic retinopathy, oxidative stress, and sirtuins: an in depth look in enzymatic patterns and new therapeutic horizons. Surv Ophthalmol. 2022;67(1):168–83.
3. Teo ZL, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. Ophthalmology. 2021;128(11):1580–91.
4. Association AD. 10. Microvascular complications and foot care: standards of medical care in diabetes—2018. Diabetes Care. 2018;41(Supplement_1):S105–18.
5. Pradeepa R, Mohan V. Prevalence of type 2 diabetes and its complications in India and economic costs to the nation. Eur J Clin Nutr. 2017;71(7):816–24.
6. Wong TY, et al. Diabetic retinopathy. Nat Rev Dis Primers. 2016;2:16012.

7. Shukla UV, Tripathy K. Diabetic retinopathy. Treasure Island: StatPearls Publishing; 2024.
8. Vashist P, et al. Role of early screening for diabetic retinopathy in patients with diabetes mellitus: an overview. Indian J Community Med. 2011;36(4):247.
9. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol. 2018;14(2):88–98.
10. Wang Z, et al. Performance of artificial intelligence in diabetic retinopathy screening: a systematic review and meta-analysis of prospective studies. Front Endocrinol. 2023;14:1197783.
11. Viswanath K, McGavin DM. Diabetic retinopathy: clinical findings and management. Community Eye Health. 2003;16(46):21.
12. Virgili G, et al. Optical coherence tomography (OCT) for detection of macular oedema in patients with diabetic retinopathy. Cochrane Database Syst Rev. 2015;1(1): CD008081.
13. Heitmar R, et al. Comparison of subjective and objective methods to determine the retinal arterio-venous ratio using fundus photography. J Optom. 2015;8(4):252–7.
14. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. JAMA. 2016;316(22):2366–7.
15. Quazi S. Artificial intelligence and machine learning in precision and genomic medicine. Med Oncol. 2022;39(8):120.
16. Ting DSW, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol. 2019;103(2):167–75.
17. Grauslund J. Diabetic retinopathy screening in the emerging era of artificial intelligence. Diabetologia. 2022;65(9):1415–23.
18. Shahriari MH, et al. Artificial intelligence in screening, diagnosis, and classification of diabetic macular edema: a systematic review. Surv Ophthalmol. 2023;68(1):42–53.
19. Wu J-H, et al. Performance and limitation of machine learning algorithms for diabetic retinopathy screening: meta-analysis. J Med Internet Res. 2021;23(7): e23863.
20. Ryu G, et al. A deep learning model for identifying diabetic retinopathy using optical coherence tomography angiography. Sci Rep. 2021;11(1):23024.
21. Penha FM, et al. Single retinal image for diabetic retinopathy screening: performance of a handheld device with embedded artificial intelligence. Int J Retina Vitreous. 2023;9(1):41.
22. Zhang W-F, et al. The validation of deep learning-based grading model for diabetic retinopathy. Front Med. 2022;9:839088.
23. Li N, et al. A stratified analysis of a deep learning algorithm in the diagnosis of diabetic retinopathy in a real-world study. J Diabetes. 2022;14(2):111–20.
24. Li Z, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs. Diabetes Care. 2018;41(12):2509–16.
25. Joseph S, et al. Diagnostic accuracy of artificial intelligence based automated diabetic retinopathy screening in real-world settings: a systematic review and meta-analysis. Am J Ophthalmol. 2024. https://doi.org/10.1016/j.ajo.2024.02.012.
26. Jain A, et al. Use of offline artificial intelligence in a smartphone-based fundus camera for community screening of diabetic retinopathy. Indian J Ophthalmol. 2021;69(11):3150–4.
27. Keel S, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. Sci Rep. 2018;8(1):4330.
28. Jiang H, et al. A multi-label deep learning model with interpretable Grad-CAM for diabetic retinopathy classification. in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2020.
29. Ipp E, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. JAMA Netw Open. 2021;4(11):e2134254–e2134254.
30. Ibáñez-Bruron MC, et al. Accuracy of artificial intelligence compared to trained medical technologists in diabetic retinopathy screening. Rev Med Chil. 2021;149(4):493–500.
31. Huang Y-P, et al. Deep learning models for automated diagnosis of retinopathy of prematurity in preterm infants. Electronics. 2020;9(9):1444.
32. Hsu M-Y, et al. Deep learning for automated diabetic retinopathy screening fused with heterogeneous data from EHRs can lead to earlier referral decisions. Transl Vis Sci Technol. 2021;10(9):18–18.
33. Hsieh Y-T, et al. Application of deep learning image assessment software VeriSee™ for diabetic retinopathy screening. J Formos Med Assoc. 2021;120(1):165–71.
34. Heydon P, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. Br J Ophthalmol. 2021;105(5):723–8.
35. He J, et al. Artificial intelligence-based screening for diabetic retinopathy at community hospital. Eye. 2020;34(3):572–6.
36. Hao Z, et al. Application and observation of artificial intelligence in clinical practice of fundus screening for diabetic retinopathy with non-mydriatic fundus photography: a retrospective observational study of T2DM patients in Tianjin, China. Ther Adv Chronic Dis. 2022;13:20406223221097336.
37. Guo Y, et al. Quantification of nonperfusion area in montaged widefield OCT angiography using deep learning in diabetic retinopathy. Ophthalmol Sci. 2021;1(2):100027.
38. Gulshan V, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. JAMA Ophthalmol. 2019;137(9):987–93.
39. Gulshan V, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402–10.
40. Grzybowski A, Brona P. Analysis and comparison of two artificial intelligence diabetic retinopathy screening algorithms in a pilot study: IDx-DR and retinalyze. J Clin Med. 2021;10(11):2352.
41. Gonzalez-Briceno G, et al. Artificial intelligence-based referral system for patients with diabetic retinopathy. Computer. 2020;53(10):77–87.

42. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology. 2017;124(7):962–9.
43. Glinton SL, et al. Phenotyping of ABCA4 retinopathy by machine learning analysis of full-field electroretinography. Transl Vis Sci Technol. 2022;11(9):34–34.
44. Gadekallu TR, et al. Early detection of diabetic retinopathy using PCA-firefly based deep learning model. Electronics. 2020;9(2):274.
45. Fleming AD, et al. Deep learning detection of diabetic retinopathy in Scotland's diabetic eye screening programme. Br J Ophthalmol. 2023. https://doi.org/10.1136/bjo-2023-323395.
46. Al-hazaimeh OM, et al. Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images. Int J Online Biomed Eng. 2022;18(13):131.
47. Aziz T, Charoenlarpnopparut C, Mahapakulchai S. Deep learning-based hemorrhage detection for diabetic retinopathy screening. Sci Rep. 2023;13(1):1479.
48. Alwakid G, et al. Deep learning-enhanced diabetic retinopathy image classification. Digit Health. 2023;9:20552076231194944.
49. AbdelMaksoud E, Barakat S, Elmogy M. A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. Med Biol Eng Comput. 2022;60(7):2015–38.
50. Baget-Bernaldiz M, et al. Testing a deep learning algorithm for detection of diabetic retinopathy in a Spanish diabetic population and with messidor database. Diagnostics. 2021;11(8):1385.
51. Bilal A, et al. AI-based automatic detection and classification of diabetic retinopathy using U-Net and deep learning. Symmetry. 2022;14(7):1427.
52. Bhimavarapu U, Battineni G. Deep learning for the detection and classification of diabetic retinopathy with an improved activation function. Healthcare. 2022. https://doi.org/10.3390/healthcare11010097.
53. Alyoubi WL, Abulkhair MF, Shalash WM. Diabetic retinopathy fundus image classification and lesions localization system using deep learning. Sensors. 2021;21(11):3704.
54. Miao Y, Tang S. Classification of diabetic retinopathy based on multiscale hybrid attention mechanism and residual algorithm. Wirel Commun Mob Comput. 2022. https://doi.org/10.1155/2022/5441366.
55. Lee AY, et al. Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. Diabetes Care. 2021;44(5):1168–75.
56. Lam C, et al. Automated detection of diabetic retinopathy using deep learning. In: AMIA Summits on Translational Science Proceedings, 2018. 2018; p. 147.
57. Nugroho HA, Frannita EL. Intelligent diabetic retinopathy detection using deep learning. In: 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). IEEE; 2021.
58. Nneji GU, et al. Identification of diabetic retinopathy using weighted fusion deep learning based on dual-channel fundus scans. Diagnostics. 2022;12(2):540.
59. Yang C, et al. Usefulness of machine learning for identification of referable diabetic retinopathy in a large-scale population-based study. Front Med. 2021;8:773881.
60. Zhao Y, et al. Using machine learning techniques to develop risk prediction models for the risk of incident diabetic retinopathy among patients with type 2 diabetes mellitus: a cohort study. Front Endocrinol. 2022;13:876559.
61. Pinedo-Diaz G, et al. Suitability classification of retinal fundus images for diabetic retinopathy using deep learning. Electronics. 2022;11(16):2564.
62. Surya J, et al. Efficacy of deep learning-based artificial intelligence models in screening and referring patients with diabetic retinopathy and glaucoma. Indian J Ophthalmol. 2023;71(8):3039–45.
63. Zhang Y, et al. Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multicenter and prospective study. BMJ Open Diabetes Res Care. 2020;8(1): e001596.
64. Sosale B, et al. Simple, mobile-based artificial intelligence algorithm in the detection of diabetic retinopathy (SMART) study. BMJ Open Diabetes Res Care. 2020;8(1): e000892.
65. Mehboob A, et al. A deep learning based approach for grading of diabetic retinopathy using large fundus image dataset. Diagnostics. 2022;12(12):3084.
66. Mujeeb Rahman K, Nasor M, Imran A. Automatic screening of diabetic retinopathy using fundus images and machine learning algorithms. Diagnostics. 2022;12(9):2262.
67. Abràmoff MD, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci. 2016;57(13):5200–6.
68. Palaniswamy T, Vellingiri M. Internet of things and deep learning enabled diabetic retinopathy diagnosis using retinal fundus images. IEEE Access. 2023;11:27590–601.
69. Ting DSW, 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(22):2211–23.
70. Jebaseeli TJ, Durai CAD, Peter JD. Retinal blood vessel segmentation from diabetic retinopathy images using tandem PCNN model and deep learning based SVM. Optik. 2019;199:163328.
71. Jena M, et al. Exploring the parametric impact on a deep learning model and proposal of a 2-branch CNN for diabetic retinopathy classification with case study in IoT-Blockchain based smart healthcare system. Informatica. 2022. https://doi.org/10.31449/inf.v46i2.3906.
72. Jiang H, et al. An interpretable ensemble deep learning model for diabetic retinopathy disease classification. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). 2019; IEEE.
73. Khan MB, et al. Automated diagnosis of diabetic retinopathy using deep learning: on the search of segmented retinal blood vessel images for better performance. Bioengineering. 2023;10(4):413.
74. Shankar K, et al. Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recogn Lett. 2020;133:210–6.
75. Kuna S, Prasad A. Deep learning empowered diabetic retinopathy detection and classification using retinal fundus images. Int J Recent Innov Trends Comput Commun. 2023;11(1):117–27.

76. Ludwig CA, et al. Automatic identification of referral-warranted diabetic retinopathy using deep learning on mobile phone images. Transl Vis Sci Technol. 2020;9(2):60–60.
77. Sosale B, et al. Medios—an offline, smartphone-based artificial intelligence algorithm for the diagnosis of diabetic retinopathy. Indian J Ophthalmol. 2020;68(2):391–5.
78. Roy ND, Biswas A. Deep learning-based early sign detection model for proliferative diabetic retinopathy in neovascularization at the disc. In: Mandal JK, Mukhopadhyay S, Dutta P, Dasgupta K, editors. Algorithms in machine learning paradigms. Berlin: Springer; 2020. p. 91–108.
79. Romero-Aroca P, et al. Validation of a deep learning algorithm for diabetic retinopathy. Telemed e-Health. 2020;26(8):1001–9.
80. Pei X, et al. Efficacy of artificial intelligence-based screening for diabetic retinopathy in type 2 diabetes mellitus patients. Diabetes Res Clin Pract. 2022;184:109190.
81. Rayavel P, Murukesh C. Comparative analysis of deep learning classifiers for diabetic retinopathy identification and detection. Imaging Sci J. 2022;70(6):358–70.
82. Paradisa RH, et al. Classification of diabetic retinopathy through deep feature extraction and classic machine learning approach. In: 2020 3rd International Conference on Information and Communications Technology (ICOIACT). IEEE; 2020.
83. Li J, et al. Interpretable machine learning-derived nomogram model for early detection of diabetic retinopathy in type 2 diabetes mellitus: a widely targeted metabolomics study. Nutr Diabetes. 2022;12(1):36.
84. Roychowdhury S, Koozekanani DD, Parhi KK. DREAM: diabetic retinopathy analysis using machine learning. IEEE J Biomed Health Inform. 2013;18(5):1717–28.
85. Sarao V, Veritti D, Lanzetta P. Automated diabetic retinopathy detection with two different retinal imaging devices using artificial intelligence: a comparison study. Graefes Arch Clin Exp Ophthalmol. 2020;258:2647–54.
86. Li W, et al. Predictive model and risk analysis for diabetic retinopathy using machine learning: a retrospective cohort study in China. BMJ Open. 2021;11(11): e050989.
87. Wu Q, et al. Development and validation of a deep learning model to predict the occurrence and severity of retinopathy of prematurity. JAMA Netw Open. 2022;5(6):e2217447–e2217447.
88. Ruamviboonsuk P, et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. Lancet Digital Health. 2022;4(4):e235–44.
89. Ruamviboonsuk P, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. NPJ Digit Med. 2019;2(1):25.
90. Saxena G, et al. Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public datasets. Intell-Based Med. 2020;3:100022.
91. Sayres R, et al. Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. Ophthalmology. 2019;126(4):552–64.
92. Shah A, et al. Validation of automated screening for referable diabetic retinopathy with an autonomous diagnostic artificial intelligence system in a Spanish population. J Diabetes Sci Technol. 2021;15(3):655–63.
93. Shah P, et al. Validation of deep convolutional neural network-based algorithm for detection of diabetic retinopathy—artificial intelligence versus clinician for screening. Indian J Ophthalmol. 2020;68(2):398–405.
94. Rajalakshmi R, et al. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye. 2018;32(6):1138–44.
95. Reddy YMS, Ravindran RSE. Retinal image lesions assisted diabetic retinopathy screening system through machine learning. Int J Intell Eng Syst. 2022;15(2):175.
96. Rom Y, et al. Predicting the future development of diabetic retinopathy using a deep learning algorithm for the analysis of non-invasive retinal imaging. BMJ Open Ophthalmol. 2022;7(1): e001140.
97. Rogers TW, et al. Evaluation of an AI system for the detection of diabetic retinopathy from images captured with a handheld portable fundus camera: the MAILOR AI study. Eye. 2021;35(2):632–8.
98. Ryu G, et al. A deep learning algorithm for classifying diabetic retinopathy using optical coherence tomography angiography. Transl Vis Sci Technol. 2022;11(2):39–39.

Publisher's Note