Artificial Intelligence and Machine Learning Applications in Diabetic Retinopathy Screening and Glaucoma Detection: A Review Article

Gursimran Singh[1]

[1](PhD Scholar, M. Optom)

Publication Date: 2026/01/08

Abstract: Artificial intelligence (AI) and machine learning (ML) have become groundbreaking technologies in the field of Ophthalmology and especially in the screening of diabetic retinopathy (DR) and glaucoma. This review discusses how AI/ML systems are currently applied, methodologies used, performance metrics, and clinical implementation issues of AI/ML systems in the detection of these vision-threatening conditions. Very recent deep learning algorithms have achieved the diagnostic accuracy of human experts, or more, with a sensitivity and specificity rate over 90% in many of the studies. However, there are challenges such as dataset bias, regulatory approval, clinical integration and cost-effectiveness that need to be investigated further. This review is a synthesis of evidence obtained from recent literature and discusses future directions of AI powered ophthalmic diagnostics.

Keywords: *Artificial Intelligence, Machine Learning, Deep Learning, Diabetic Retinopathy, Glaucoma, Computer Aided Diagnosis, Retinal Imaging.*

How to Cite: Gursimran Singh (2026) Artificial Intelligence and Machine Learning Applications in Diabetic Retinopathy Screening and Glaucoma Detection: A Review Article. *International Journal of Innovative Science and Research Technology*, 11(1), 104-109. https://doi.org/10.38124/ijisrt/26jan012

I. INTRODUCTION

Diabetic Retinopathy and Glaucoma are two of the most common causes of preventable blindness in the world; they cause millions of people to go blind every year. Diabetic retinopathy (microvascular complications of diabetes mellitus) affects around 35% of diabetic patients and may lead to vision loss if not detected and treated early [1]. Similarly, glaucoma affects more than 76 million people worldwide with projected numbers of glaucoma rising to 111.8 million in 2040[2]. Early detection and timely intervention is very critical in preventing irreversible loss of vision in both conditions.

Traditional screening techniques involve extensive manual screenings conducted by ophthalmologists or trained personnel which are time consuming and are subjective in nature, often limited by access to specialized personnel especially in regions where there is a shortage. The worldwide dearth of eye care professionals is estimated at 4.3 million worldwide[3], resulting in great barriers to adequate screening coverage. This disparity is the greatest in low and middle-income countries where diabetes and glaucoma burden is rapidly increasing.

Artificial intelligence and especially the deep learning architectures based on convolutional neural networks (CNNs) has shown remarkable capabilities in the analysis of medical images. These technologies hold the promise of automating screening processes, enhancing diagnostic accuracy, cutting costs of healthcare services, and opening access to eye care services in resource-limited settings. The success of AI applications in ophthalmology has made a big leap since 2016, when the first FDA-approved autonomous AI diagnostic system for diabetic retinopathy was launched[4].

This review attempts to present a detailed analysis of current applications of AI and ML in diabetic retinopathy screening as well as glaucoma detection with a focus on methodologies used, performance characteristics, clinical validation studies, implementation challenges and future directions for this rapidly evolving field.

II. LITERATURE REVIEW

➢ *Diabetic Retinopathy Screening with AI/ML*

Diabetic Retinopathy Screening has become one of the most successful uses of AI in medicine. The disease develops in well-defined stages, namely mild, moderate and severe non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR), and each stage is

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan012

characterised by the presence of specific retinal lesions such as microaneurysms, haemorrhages, exudates, cotton wool spots and neovascularisation [5].

Early AI methods were based on traditional machine learning algorithms on hand-engineered features for lesion detection. Niemejer et al. used k-nearest neighbour classifiers and support vector machines (SVMs) to develop systems for detecting microaneurysms and haemorrhages with sensitivities of 85-90%[6]. However, these approaches involved a great deal of feature engineering and were limited by not being able to learn hierarchical representations.

The beginning of the deep learning revolutionization of DR screening, Gulshan et al. trained a deep CNN model using 128,175 retinal images and obtained a sensitivity of 97.5% and a specificity of 93.4% for referable diabetic retinopathy, which is as good as board-certified ophthalmologists [7]. This landmark study showed that deep learning was able to perform on the same level as an expert without human feature extraction.

Thereafter, a number of commercial artificial intelligence systems have been developed and clinically validated. IDx-DR was the first FDA-approved autonomous AI diagnostic system in 2018, which demonstrated 87.2% sensitivity and 90.7% specificity for the detection of more-than-mild DR in primary care settings [8]. EyeArt, another FDA-authorized system had 91.3% sensitivity and 91.1% specificity in diverse populations[9].

Recent developments have been made in order to enhance model interpretability, dataset bias, and the detection of diabetic macular oedema (DME), a major contributor to vision loss in diabetic patients. Ting et al. created a multi-ethnic deep learning system trained on almost 500,000 images from different ethnicities with excellent area under the curve (AUC) values of more than 0.93 for the detection of referable DR and vision-threatening DR for some ethnicities[10].

➤ *Detection of Glaucoma Using Artificial Intelligence/ Machine Learning*

Glaucoma is a progressive optic neuropathy in which there is a structural damage to the optic nerve head with corresponding visual defect of the field of vision. Early detection is very important since vision loss from glaucoma cannot be restored. Traditional diagnosis involves several examinations such as intraocular pressure measurement, visual field testing and evaluation of the optic disc by fundus photography or optical coherence tomography (OCT)[11].

AI applications in glaucoma diagnosis have been aimed at analysing the fundus photographs and OCT images that can detect the structural changes in the optic nerve head, including an increased cup-to-disc ratio (CDR), neuroretina rim thinning and retinal nerve fiber layer (RNFL) loss. Early ML methods involved the use of classical algorithms in order to segment the optic disc and cup in order to calculate CDR as a diagnostic marker. Bock et al. used SVMs for glaucoma classification using super pixel features, and it achieved 80% sensitivity and 80% specificity [12].Deep learning has made it much easier to detect glaucoma. Li et al. proposed a CNN with AUC of 0.986 for glaucoma detection based on fundus images, which was better than traditional methods[13]. Christopher et al. demonstrated the use of deep learning to detect glaucoma using OCT scans with 92.0% accuracy which is close to expert level[14].

Recent studies have been done that have investigated multi-modality approaches using a combination of fundus photography, OCT, and visual field data. Muhammad et al. created a hybrid deep learning framework using multiple imaging modalities, and achieved 98.8% accuracy classifying glaucoma[15]. These multi-modal systems make use of complementary information from different diagnostic tests, and may enhance diagnostic accuracy above the single modality approaches.

AI systems have also been created for glaucoma progression prediction. Medeiros et al designed deep learning models that were 85% accurate at visual field progression up to 5.5 years before conventional methods[16]. This predictive ability could help to allow earlier therapeutic intervention and disease management.

➤ *Comparison Analysis, Validation in the Clinic*

Several comparative studies have been done in which AI systems were directly compared to human experts. Kanagasingam et al. carried out a head-to-head comparison finding that automated DR screening systems had similar sensitivity (92.4% vs 91.8%) but greater specificity (90.3% vs 85.2%) as human graders[17]. For glaucoma detection, Shibata et al. demonstrated the detection of glaucoma by deep learning algorithms which matched or outperformed the glaucoma specialist in 90% of the cases [18].Real-world clinical validation studies have shown a feasibility and effectiveness in diverse healthcare settings. Rajalakshmi et al. implemented an AI screening system in 20 diabetic screening centers in India and were able to screen 8,000 patients with 96.7% sensitivity in the identification of referable DR[19]. Similarly, Ruamviboonsuk et al. used deep learning screening in Thailand and showed high accuracy and patient acceptance[20].

Validation studies have however also shown important limitations. Performance degradation has been noted when AI systems trained with high-quality research images are applied to low-quality real-world images [21]. Additionally, accuracy is reduced in most systems for less common DR manifestations as well as in populations underrepresented in training datasets[22].

III. METHODOLOGY

➤ *Search Strategy and Selection Criteria*

A thorough literature search was conducted using the PubMed, the IEEE Xplore, the Google Scholar and the Web of Science databases for articles published between 2015 and 2024. Search terms combined 'artificial intelligence,' 'machine learning,' 'deep learning,' 'convolutional neural networks,' 'diabetic retinopathy,' 'glaucoma,' 'screening,'

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan012

'detection,' 'diagnosis' and 'fundus imaging.' Additional articles were found via the reference lists of selected papers and review articles.

Inclusion criteria included original research articles and clinical trials which were peer-reviewed and systematic reviews that focused on the application of AI/ML for DR screening and glaucoma detection. In the studies, quantitative performance measures (sensitivity, specificity, accuracy or AUC) had to be reported. Exclusion criteria were conference abstracts without publications, studies with fewer than 100 cases, and articles not in English.

➢ *AI/ML Methodologies of Ophthalmic Imaging*

- *Acquisition and Preprocessing of an Image*

Both the DR and glaucoma detection systems usually use color fundus photography, which is the most widely available modality of retinal imaging. Images are taken with the use of fundoscopic cameras at different fields of view (usually 45o to 60o) and resolution. For glaucoma, OCT imaging provides three-dimensional structural information of the optic nerve head and RNFL thickness.[23]

Preprocessing steps may often involve quality assessment of an image, detection of field of view, artifact removal, and standardisation. Contrast enhancement techniques such as adaptive histogram equalization provides good vessel and lesion visibility. For deep learning models the images are usually resized to standard sizes (e.g. 512x512 or 224x224 pixels) and normalized to make the training converge easier [24].

- *Deep Learning Architectures*

Convolutional neural networks are the basis of most of the AI diagnostic systems available today. Architectures such as the following are popular:ResNet (Residual Networks): Used due to their capability of training very deep networks using skip connections. ResNet-50 and ResNet-101 have been widely used in DR classification[25].Inception Networks: Use multiple scale convolutional filters to detect features at different spatial scales, useful for the detection of different sizes of lesions in DR[7].Dense Net: Dense connections are used between layers which helps in better propagating and parameter efficiency, achieved good performance in both DR and glaucoma detection [26].Efficient Net: Optimizes network depth, width and resolution all at once to achieve state-of-the-art performance with less parameters [27].Vision Transformers: Recently introduced attention-based architectures that have been shown to be promising in medical imaging, albeit with the need for larger training datasets[28].

- *Training Strategies*

Transfer learning is commonly used, in which models pre-trained with large datasets of natural images (e.g. ImageNet) are fine-tuned with medical images. This approach remedies the one issue of the lack of available labels of medical images and shortens the training time [29].

Data augmentation techniques such as rotation, flipping, scaling, brightness adjustment and elastic deformations help to artificially enlarge training datasets and increase generalization of models. For unbalanced datasets that are so prevalent in medical imaging, several methods are used to even out class distributions, including oversampling, under sampling and synthetic minority oversampling (SMote)[30].

Ensemble methods using the predictions of several models or architectures usually help to increase robustness and accuracy. Ensemble approaches based on 5-10 models with distinct architectures or training procedures have led to improvements of accuracy of 2-5% compared to single models[31].

➢ *Performance Evaluation*

The common standards for evaluating artificial intelligence diagnostic systems are sensitivity (recall), specificity, accuracy, precision, F1-score, and AUC-ROC. For screening purposes, a high sensitivity is desired to avoid false negatives, while acceptable specificity is desired to prevent unnecessary referrals[32].Cross-validation techniques especially k-fold cross-validation check the generalization of any model on the training data. However, external validation on independent datasets from different institutions and populations and imaging devices gives the most reliable performance estimates [33].

Confusion matrices and classification reports show how the model is performing for the different severity of the disease, and therefore identify weaknesses. Gradient-weighted Class Activation Mapping (Grad-CAM) and visualization methods that offer interpretability by suggesting image regions that contribute to model decisions[34]

IV. DISCUSSION

➢ *Implications for Performance and Clinical Practice*

Current AI systems for DR screening achieve consistently high sensitivity and specificity over 90% for the detection of DR referrable disease, and can meet or exceed performance benchmarks set for performance by human graders. The autonomous nature of systems such as IDx-DR is a paradigm shift that allows the screening of non-specialist personnel in primary care settings, pharmacies and remote locations.[35] .For glaucoma, AI detection accuracy is 85-98% depending on the modality of imaging and severity of the disease. The ability to predict years of progression in advance has the potential to completely change how we treat these diseases because we would be able to intervene earlier and more individually. However, detecting glaucoma is still more difficult to do than detecting DR because of the heterogeneity of the disease and also requires the integration of functional and structural information [36].Cost-effectiveness analyses indicate that the use of AI screening may be very cost-effective for healthcare. Xie et al. estimated that in the United States, auditing using artificial intelligence to screen for DR could save $1.2 billion per year while screening 50% more patients[37] Implementation of this in low-resource settings has even greater potential impact where the other option is often no screening at all.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan012

➤ *Challenges and Limitations*

Despite the unbelievable performance, there are still a number of challenges that restrict the use of AI in clinical practice:

- *Dataset Bias and Generalisability:*

Most of the data for training is from developed countries and might not represent the diversity of the global population. Performance degradation when used on new populations, imaging devices or image quality conditions is a substantial barrier [38]. Raumviboonsuk et al found that an AI system trained mostly in Caucasian patients exhibited 8% lower sensitivity when used in Thai patients[20].Regulatory and Liability Concerns Regulatory pathways for AI medical devices are still evolving. Questions about liability in cases of diagnostic errors - who is responsible, the AI developer, the implementation institution, or the ordering physician - are not yet solved in many jurisdictions[39]. The regulatory landscape for AI/ML-based software as a medical device developed by the FDA is an attempt to address these issues and is faced with challenges with continuously learning systems. Clinical Integration and Workflow: Integrating AI systems into the existing healthcare workflows requires much technical infrastructure, staff training, and redesigning of the workflow. Many electronic health record systems do not have seamless capabilities for integrating proper AI. Physician skepticism and concerns about autonomous systems being able to make clinical decisions without the oversight of humans persists[40].

- *Interpretability and Trust:*

Deep learning models are "black boxes" that make decisions using complex non-linear transformations that are hard to interpret. While visualization techniques, such as Grad-CAM, do give some insight, they may not fully explain the reasoning of the model. This opacity can prevent clinician trust as well as patient acceptance[41].Image Quality and Artifact Sensitivity Real world clinical images are usually subject to quality problems such as poor focus, non-uniform illumination, non-uniformity of medium, and artifact. While some AI systems include automatic quality assessment, others experience serious performance deterioration using suboptimal images [42].Rare Manifestations and Edge Cases: Most of the AI systems are geared towards common presentations of disease. Performance for rare manifestations, early stages of the disease or atypical cases may be substantially lower. This limitation may lead to missed diagnosis of patients with unusual presentations of the disease [43].

➤ *Ethical Considerations*

There are important ethical questions raised by AI implementation in healthcare. Algorithmic bias has the potential to further widen healthcare disparities if systems are too bad for underrepresented populations. Making sure that all patient groups are able to access and perform equally requires conscious dataset curation and validation Patient autonomy and informed consent need to be considered. Patients need to know when AI systems are part of their diagnosis and should be offered a choice to request evaluation by humans only if they are so inclined. Transparency in the use of AI to make clinical decisions honours patient autonomy and has the potential to promote trust [45].Data privacy concerns are an issue because of the big data sets needed for training AI. Ensuring that data is properly de-identified, securely stored, and properly consented to is the key to ensuring that data is used appropriately. Federated learning strategies of model training without centralization of patient data are promising privacy-preserving alternatives[46].

➤ *Future Directions*

There are a number of promising developments that will potentially address the limitations and increase AI capabilities:

- *Multi-Modal and Multi-Disease Detection:*

The next-generation systems that can identify multiple diseases from a single fundus image could increase the screening efficiency. Grassmann et al developed a deep learning system detecting 10 retinal diseases and cardiovascular risk factors from fundus photographs [47]. Such comprehensive screening tools may represent more clinical value than single-disease systems.

- *Federated Learning:*

This is a method of model training that is performed across many institutions without exchanging patient data, overcoming privacy issues while using larger and more diverse datasets. There are several initiatives to federated learning in the ophthalmic AI development[48].

- *Explainable AI (XAI):*

Techniques that bring human-interpretable explanations for AI decisions could make clinicians more trusting and find it easier to detect errors. Attention mechanisms, layer-wise relevance propagation and counterfactual explanations are promising XAI approaches[49].

- *Integration with Mobile and Portable Devices:*

Fundus imaging from smartphones and AI analysis integration could help in the screening of the remotest locations. Several studies have proven the feasibility of smartphone-based DR screening with similar accuracy as traditional fundus cameras[50]

- *Continuous Learning Systems:*

AI models that are able to continuously learn from new data and expert feedback could ensure that the performance of the AI model is maintained as the imaging technologies, patient populations, and disease patterns evolve over time. However, such systems need powerful validation systems in order to guarantee safety[51].

V. CONCLUSION

Artificial intelligence and machine learning have shown revolutionary prospects in the screening of diabetic retinopathy and glaucoma, as this technology has shown similar or higher diagnostic accuracy compared to human experts in many studies. These technologies provide solutions to important issues in eye care delivery, such as workforce shortages, geographical barriers and variable coverage of screening. FDA-approved autonomous systems such as IDx-

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan012

DR is a milestone in the AI-enabled healthcare field for the feasibility of AI to work alone in a clinical setting.

However, there are still major challenges to overcome before AI can bring about widespread clinical adoption. Dataset bias, regulatory uncertainty, integration challenges, interpretability issues, and questions about generalisability to different populations and imaging conditions mean that further research and development are needed. Addressing these limitations will require collaborative efforts between AI developers, clinicians, regulators, and healthcare organizations.

The future of AI in ophthalmology is probably for hybrid systems, which combine the efficiency of AI with the knowledge of human professionals, rather than for AI to completely automate the system. Multi-modal, multi-disease detection systems that are embedded in comprehensive healthcare platforms have the potential to offer the most clinical value. Continued validation in varied real-life scenarios, development of explainable AI techniques and having a regulatory framework for it will be crucial to realizing the full potential of AI in preventing blindness due to diabetic retinopathy and glaucoma. As these technologies mature, they hold great promise of democratizing access to high-quality eye care that would allow us to better screen and treat millions of people in the world who, currently, are not able to access adequate screening services. The next decade will be crucial in determining whether this promise will be realised in terms of measurable reductions in population levels of preventable vision loss.

REFERENCES

[1]. Yau JW, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care. 2012;35(3):556-564.
[2]. Tham YC, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. Ophthalmology. 2014;121(11):2081-2090.
[3]. Resnikoff S, Lansingh VC, Washburn L, et al. Estimated number of ophthalmologists worldwide (International Council of Ophthalmology update): will we meet the needs? Br J Ophthalmol. 2020;104(4):588-592.
[4]. Abràmoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digit Med. 2018;1:39.
[5]. Wilkinson CP, Ferris FL, Klein RE, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology. 2003;110(9):1677-1682.
[6]. Niemeijer M, van Ginneken B, Cree MJ, et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans Med Imaging. 2010;29(1):185-195.
[7]. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA. 2016;316(22):2402-2410.
[8]. Abràmoff MD, Lou Y, Erginay A, 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):5200-5206.
[9]. Bhaskaranand M, Ramachandra C, Bhat S, et al. The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. Diabetes Technol Ther. 2019;21(11):635-643.
[10]. Ting DSW, Cheung CY, Lim G, 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):2211-2223.
[11]. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. JAMA. 2014;311(18):1901-1911.
[12]. Bock R, Meier J, Nyúl LG, et al. Glaucoma risk index: automated glaucoma detection from color fundus images. Med Image Anal. 2010;14(3):471-481.
[13]. Li Z, He Y, Keel S, et al. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. Ophthalmology. 2018;125(8):1199-1206.
[14]. Christopher M, Belghith A, Bowd C, et al. Performance of deep learning architectures and transfer learning for detecting glaucomatous optic neuropathy in fundus photographs. Sci Rep. 2018;8(1):16685.
[15]. Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects. J Glaucoma. 2017;26(12):1086-1094.
[16]. Medeiros FA, Jammal AA, Thompson AC. From machine to machine: an OCT-trained deep learning algorithm for objective quantification of glaucomatous damage in fundus photographs. Ophthalmology. 2019;126(4):513-521.
[17]. Kanagasingam Y, Xiao D, Vignarajan J, et al. Evaluation of artificial intelligence-based grading of diabetic retinopathy in primary care. JAMA Netw Open. 2018;1(5):e182665.
[18]. Shibata N, Tanito M, Mitsuhashi K, et al. Development of a deep residual learning algorithm to screen for glaucoma from fundus photography. Sci Rep. 2018;8(1):14665.
[19]. Rajalakshmi R, Subashini R, Anjana RM, et al. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye (Lond). 2018;32(6):1138-1144.
[20]. Ruamviboonsuk P, Krause J, Chotcomwongse P, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. NPJ Digit Med. 2019;2:25.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jan012

[21]. Gulshan V, Rajan RP, Widner K, et al. Performance of a deep-learning algorithm vs manual grading for detecting diabetic retinopathy in India. JAMA Ophthalmol. 2019;137(9):987-993.
[22]. Peng Y, Dharssi S, Chen Q, et al. DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs. Ophthalmology. 2019;126(4):565-575.
[23]. Schuman JS, Hee MR, Puliafito CA, et al. Quantification of nerve fiber layer thickness in normal and glaucomatous eyes using optical coherence tomography. Arch Ophthalmol. 1995;113(5):586-596.
[24]. Zhou Y, Wang B, Huang L, et al. A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability. IEEE Trans Med Imaging. 2021;40(3):818-828.
[25]. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
[26]. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:4700-4708.
[27]. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. 2019:6105-6114.
[28]. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. In: International Conference on Learning Representations. 2021.
[29]. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging. 2016;35(5):1299-1312.
[30]. Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321-357.
[31]. Ju L, Wang X, Wang L, et al. Improving medical images classification with label noise using dual-uncertainty estimation. IEEE Trans Med Imaging. 2022;41(6):1533-1546.
[32]. Heydon P, Egan C, Bolter L, 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):723-728.
[33]. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digit Med. 2022;5(1):48.
[34]. Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. 2017:618-626.
[35]. Ipp E, Liljenquist D, Bode B, 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.
[36]. Elze T, Pasquale LR, Shen LQ, et al. Patterns of functional vision loss in glaucoma determined with archetypal analysis. J R Soc Interface. 2015;12(103):20141118.
[37]. Xie Y, Nguyen QD, Hamzah H, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. Lancet Digit Health. 2020;2(5):e240-e249.
[38]. Beede E, Baylor E, Hersch F, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020:1-12.
[39]. Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. JAMA. 2019;322(18):1765-1766.
[40]. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. NPJ Digit Med. 2018;1:40.
[41]. Amann J, Blasimme A, Vayena E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310.
[42]. Bellemo V, Lim ZW, Lim G, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. Lancet Digit Health. 2019;1(1):e35-e44.
[43]. Scheetz J, Rothschild P, McGuinness M, et al. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep. 2021;11(1):5193.
[44]. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-453.
[45]. Morley J, Machado CC, Burr C, et al. The ethics of AI in health care: a mapping review. Soc Sci Med. 2020;260:113172.
[46]. Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. NPJ Digit Med. 2020;3:119.
[47]. Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology. 2018;125(9):1410-1420.
[48]. Kaissis GA, Makowski MR, Rückert D, et al. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell. 2020;2(6):305-311.
[49]. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Netw Learn Syst. 2021;32(11):4793-4813.
[50]. Russo A, Morescalchi F, Costagliola C, et al. Comparison of smartphone ophthalmoscopy with slit-lamp biomicroscopy for grading diabetic retinopathy. Am J Ophthalmol. 2015;159(2):360-364.
[51]. Chen RJ, Wang JJ, Williamson DF, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng. 2023;7(6):719-742.