https://doi.org/10.1007/s00417-023-06245-4

MINI REVIEW

Behind the mask: a critical perspective on the ethical, moral, and legal implications of AI in ophthalmology

Daniele Veritti[1] · **Leopoldo Rubinato**[1] · **Valentina Sarao**[1,2] · **Axel De Nardin**[3] · **Gian Luca Foresti**[3] · **Paolo Lanzetta**[1,2]

Received: 19 January 2023 / Revised: 24 July 2023 / Accepted: 15 September 2023 / Published online: 25 September 2023

Abstract

Purpose This narrative review aims to provide an overview of the dangers, controversial aspects, and implications of artificial intelligence (AI) use in ophthalmology and other medical-related fields.

Methods We conducted a decade-long comprehensive search (January 2013–May 2023) of both academic and grey literature, focusing on the application of AI in ophthalmology and healthcare. This search included key web-based academic databases, non-traditional sources, and targeted searches of specific organizations and institutions. We reviewed and selected documents for relevance to AI, healthcare, ethics, and guidelines, aiming for a critical analysis of ethical, moral, and legal implications of AI in healthcare.

Results Six main issues were identified, analyzed, and discussed. These include bias and clinical safety, cybersecurity, health data and AI algorithm ownership, the "black-box" problem, medical liability, and the risk of widening inequality in healthcare.

Conclusion Solutions to address these issues include collecting high-quality data of the target population, incorporating stronger security measures, using explainable AI algorithms and ensemble methods, and making AI-based solutions accessible to everyone. With careful oversight and regulation, AI-based systems can be used to supplement physician decision-making and improve patient care and outcomes.

Key messages

What is known:

- Artificial intelligence (AI) in ophthalmology has the potential to improve patient care and outcomes, but it is not without controversy.

New Information:

- Safety, cybersecurity, health data and AI algorithm ownership, the "black-box" problem, medical liability, and the risk of widening inequality in healthcare are the primary concerns associated with the application of artificial intelligence in ophthalmology and healthcare.
- To address these issues, researchers should collect high-quality data of the target population, incorporate stronger security measures and real-life validation, employ explainable AI algorithms and ensemble methods, and strive to make AI-based solutions accessible to all.

Keywords Artificial intelligence · Bias · Black-box · Cybersecurity · Data ownership · Deep learning · Distributional shift · Ethics · Liability · Machine learning · Medical liability · Ophthalmology

Extended author information available on the last page of the article

Introduction

Artificial intelligence (AI) has the potential to revolutionize ophthalmology by making the diagnostic and decision-making process easier and faster through the analysis of large amounts of data [1]. AI methodologies have been applied to the analysis of ocular imaging, such as fundus photographs and optical coherence tomography, in the screening, diagnosis, and treatment of common ophthalmic diseases. While the use of AI in ophthalmology holds great promise, there are also a number of dangers and controversies that must be considered, including errors resulting from incorrect data synthesis, legal liability implications, cybersecurity and data ownership concerns, and the challenge of interpreting the output provided by AI software with "black-box" operation [2]. This narrative review aims to provide an overview of these issues and their implications for the widespread use of AI in ophthalmology.

Methods

In this narrative review, we conducted a comprehensive search spanning January 2013 to May 2023 of both academic and grey literature to analyze the application of AI in ophthalmology and healthcare. Our search strategy was guided by a carefully selected set of keywords: "AI accountability," "AI accessibility," "AI in diagnostics," "AI in medicine," "AI in medical research," "AI in patient monitoring," "AI in prognosis," "AI in treatment planning," "AI transparency," "artificial intelligence," "bias in AI," "clinical decision-making," "cybersecurity," "data privacy," "deep learning," "ethics," "guidelines," "healthcare," "healthcare inequality," "legal implications," "machine learning," "ophthalmology," "patient care," and "regulation." Our search spanned five academic databases: ACM, PubMed, Nature-SCI, IEEE Xplore, and the AI Ethics Guidelines Global Inventory. Given the dynamic nature of the AI field, we also included grey literature, following the search methods outlined by Godin et al [3]. These additional sources included grey literature databases, Google Scholar, targeted website searches of known organizations and institutions, and the results from a prior environmental scan conducted by a team member (VS). Two investigators (DV and LR) selected documents based on their relevance to artificial intelligence, healthcare, ethics, and guidelines. Our focus was on the ethical, moral, and legal implications of AI in healthcare, aiming for a critical analysis rather than a comprehensive review. During the code mapping process, we used an abductive methodology, incorporating both inductive and deductive approaches, to identify and rename ethical issues based on the content of the selected documents. This approach allowed us to categorize the information into six primary issues associated with the use of AI in ophthalmology.

Results

Artificial intelligence, bias, and clinical safety

Recent advancements in AI have promoted the application of machine learning (ML) in clinical practice as a way to use data from past experiences to make inferences about new patients [1]. Although this approach showed some success, the use of ML systems has been shown to be prone to bias due to the phenomenon of "distributional shift," where the ML system is inefficient at recognizing a relevant change in the actual context from the samples provided during its training [2]. This can lead to diagnostic errors due to unrepresentative training data, inadequate labeling of the patient's outcome, inadequate definition of a gold standard diagnosis for the disease, and differences in the disease stage [4–6]. For example, the epic sepsis model (ESM) is an AI-powered sepsis prediction model, trained on large datasets, which demonstrated strong performance in its initial evaluation. However, the model performance dramatically fell when it was applied in the real-world. In an external validation study, this model only generated sepsis alerts for 843 cases (33%) on 2552 patients with clinically confirmed sepsis, missing 67% of the people with sepsis. Out of 6971 ESM sepsis alerts, only 843 (12%) were correct, resulting in 88% of false alarms [6].

The accuracy of ML model predictions can be strongly affected by bias in the training data. This bias can be caused by a variety of factors, including differences between the population of the training data and the target population (e.g., ethnicity, sex, age, comorbidities), variations in data quality (e.g., image quality on the fundus camera used for the screening of diabetic retinopathy (DR)), and differences in the way the data was collected [7, 8]. In ophthalmology, bias in ML training is a significant concern, especially when using ML algorithms for screening of DR among underprivileged populations. Incorrect training of ML models can lead to inaccurate results and potentially lead to inadequate care for patients in need [8]. For instance, the performance of automated DR algorithms varies considerably due to limited training data, heterogeneity in disease presentations, and suboptimal image quality [9, 10]. Moreover, ophthalmological ML-ready datasets are only available in a few countries, leaving a large number of countries unrepresented in training and validation cohorts [11]. Additionally, AI systems cannot capture the emotional components of diagnosis, such as the impact of a diagnosis on the patient's quality of life. Esteva et al. compared ML systems to expert dermatologists on the ability to discriminate between benign and malignant skin lesions with the result that both humans and machines found it difficult to express a firm judgment, but because of the emotional aspect of a potentially life-threatening disease, human dermatologists were more prone to overdiagnose malignancy [12]. This aspect of human behavior is crucial for clinical safety and should be integrated into AI algorithms, such as in those used for high-risk DR screening. These algorithms should be trained not just with the end result, but also with the real-world impact of potential missed diagnoses.

Cybersecurity

Data security is a major concern in healthcare, particularly given the current development of AI-powered solutions. Inadequate data security in healthcare can cause serious negative consequences, including potential data breaches, loss of patient information, potential misuse of sensitive information, and potential effects on patient's health [13]. The security of digital medical devices, such as AI-based devices with diagnostic or therapeutic software, is becoming a critical issue in healthcare. The strongest motivation of healthcare systems hackers is the financial gain because the sale of health data is extremely valuable but should even be considered the political gain and the potential aim of controlling lives in a form of cyberwarfare [14]. In 2021, 45 million individuals in the US have been affected by healthcare attacks, a threefold increase from 2018 [15]. According to an annual report by "Comparitech", ransomware attacks cost the healthcare industry $20.8 billion in 2020 in the USA alone [15]. An emblematic example of how the intrusion into AI algorithms in healthcare can have serious consequences for patients was reported by Zhou et al. [16]. These authors performed a study to investigate the behaviors of an AI breast cancer diagnosis model under adversarial mammogram images. They found that the adversarial mammogram samples fooled the AI model to output a wrong diagnosis in 69.1% cases that were initially correctly classified [16]. Furthermore, the Food and Drug Administration has reported that the "MiniMed™ 508" and the "MiniMed™ Paradigm" implantable insulin pumps, manufactured by the American company Medtronic, were vulnerable to external manipulations on the insulin release settings [17]. CyberMDX, an American cybersecurity company, also discovered previously undocumented vulnerabilities in the Alaris Gateway workstation pumps used in intravenous therapies, allowing hackers to alter the administration of life-saving treatments [18]. These examples highlight the urgent need to improve the security of computerized medical devices and AI software, to protect patients from potential risks. In the field of retinal diseases, for example, unauthorized access to AI systems developed to form a plan of optimal photocoagulation arrangement for retinal laser coagulation for each case could lead to manipulation of data or algorithms, potentially leading to incorrect diagnoses or treatment plans [19].

Health data and AI algorithms ownership

In recent years, the healthcare industry has experienced the emergence of several corporate entrants, including digital technology companies, such as Google, Microsoft, IBM, and Apple. These companies have invested heavily in the development and use of AI technologies in healthcare, resulting in a number of innovative products and services. Google Health (Mountain View, California, USA) was launched in 2006, providing a repository of health records to connect doctors, hospitals, and pharmacies directly. Microsoft (Redmond, Washington, USA) followed in 2007 with "Microsoft-Healthvault," a web-based personal health record to store health and fitness information. In 2019, Apple Inc. (Cupertino, California, USA) developed an algorithm for an apple watch application which was validated in the "Apple Heart Study" conducted at Stanford University [20]. Deepmind (a British artificial intelligence subsidiary of Alphabet Inc.) collaborated with the Moorfield Eye Hospital in 2020 to develop an AI system to predict whether a patient with unilateral neovascular AMD will develop neovascular AMD in the second eye from the analysis of OCT scans [21]. In July 2021, Google announced the "Healthcare Data Engine," an end-to-end solution for healthcare and life sciences organizations to harmonize data from multiple sources for AI advanced analytics. However, these investments have raised some criticism. In July 2017, a deep-seated concern arose over the 2015 transfer of 1.6 million identifiable patient records from the London-based Royal Free NHS Trust to Google's DeepMind health unit, without the patients' explicit permission for their data to be shared [22]. This has opened a debate in the scientific community regarding the deliberate use of personal health data. The ownership of health data and responsibility for it are unresolved questions, with bioethics literature commonly asserting that patients should own their data and have the right to make decisions about its access [23–25]. The issue of health data and AI algorithms ownership is a critical concern in the field of ophthalmology. For instance, the development and application of AI algorithms, such as Eye2Gene, for diagnosing inherited retinal diseases (IRDs) have raised questions about who owns the data and the algorithm itself. Eye2Gene, an AI algorithm designed to accelerate the diagnosis of IRDs, utilizes patient retinal images to predict causative genes. The algorithm is trained on datasets from multiple hospitals and validated on external datasets. However, the ownership of the data used for training and validation, as well as the AI algorithm itself, is not explicitly stated. This situation underscores the need for clear policies and regulations regarding the ownership and use of health data and AI algorithms in ophthalmology [26]. Companies and institutions have started to consider patients as "data traders," who provide information about their health in exchange for remuneration [27]. Online platforms, such as "The Savvy platform," have been created to enable patients to sell and share their medical data. However, once the data has been shared, patients have no control over its secondary use.

The "black-box" problem

AI systems used to interpret data and provide decisions based on that data have recently been subject to much debate regarding the "black-box problem" [28]. This concept refers to the opacity of the AI system, meaning that it is difficult to understand, predict, or systematically influence the way in which an input is transformed to an output. As a result, end-users are less likely to trust and cede control to machines whose workings they do not understand [29]. Some authors propose that the issue of transparency might be subordinate to the proficiency of AI in augmenting patient outcomes and overall public health. In this regard, Evans and colleagues for instance underscore that the capability of AI to substantially enhance the results of retinopathy of prematurity screening and treatment, particularly in resource-constrained environments, accentuates the potential advantages of AI in the healthcare sector despite the prevailing concerns related to transparency [30]. In contrast, other authors support white-box AI [31]. White-box AI is more transparent in its decision-making process, since it is based on simpler models such as linear regression and decision trees that are significantly easier to interpret. However, these models provide less predictive capacity and cannot always model the inherent complexity of the dataset. "Black-box" AI systems, on the other hand, are based on more complex algorithms and are more efficient and accurate than "white-box" models and are particularly useful in the field of image analysis and processing (as in the ophthalmology field) [32]. To increase trust, solutions such as estimating uncertainty of predictions and mimicking ophthalmologists' clinical reasoning can be employed [33, 34]. Abràmoff et al. developed a two-stage ML system that first learns to detect lesions considered relevant by ophthalmologists (microaneurysms, exudates, etc.) and then it bases its predictions on these detections. Because these models mimic ophthalmologists' clinical reasoning, they are more likely to be adopted in the clinical practice [35]. Class activation mapping is a technique that can be used to understand how and why a model makes a particular decision. This technique produces a heatmap of the input image which highlights the areas of the image that most strongly contribute to the model's decision [36, 37]. While these techniques are instrumental for image-based models, there are also methods available to interpret the decision-making process for non-image data. Partial dependence plots (PDPs) and techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) provide robust ways of understanding how each feature influences predictions. PDPs show the effect of a feature on predictions, SHAP provides a consistent measure of feature importance, and LIME explains predictions by approximating the model locally with an interpretable model. In 2017, the European Union proposed the "Right for an Explanation" as an aim for AI applications utilizing personal data to make diagnostic or therapeutic decisions regarding individuals. This right is intended to ensure that patients and ophthalmologists comprehend the basis for such decisions, and to ensure the transparency and accountability of the AI system [38].

Medical liability

Ophthalmologists have expressed concerns regarding the potential medical liability arising from errors in AI-based decision-making, particularly in the diagnosis and management of major eye conditions such as diabetic retinopathy, glaucoma, age-related macular degeneration, and cataracts. The fear of medical malpractice acts as a significant deterrent to the adoption of AI in clinical practice, despite its potential to improve diagnostic accuracy and efficiency [39].

The legal implications of AI-based medical decision-making are complex and multi-faceted. Who is to be held responsible in case of errors? The clinician, the software developer, the software rights holder, or the hospitals and institutions that adopted the system? Currently, AI-based software is not a component of the common medical standard of care. If a clinician follows an incorrect treatment recommendation given by AI-based software, and this leads to harm for the patient, the clinician is likely liable for medical malpractice. This is because AI-based software is usually considered a tool under the control of the healthcare professional, who must make the final decision. Therefore, clinicians could be held accountable even when they had faith in the black-box algorithms. In the USA, initiatives have been made to apply "product liability" laws to healthcare AI software, but various US courts have been reluctant to enforce these laws to healthcare AI algorithm developers due to the current perception of AI systems as confirmatory tools to support clinicians in their final decision [40, 41]. Others support the "negligent credentialing" theory as a potential avenue of liability for hospitals, should they fail to adequately review and verify the specifics and reliability of the AI systems they employ [42]. In 2017, the European Parliament released a resolution titled Civil Law Rules on Robotics. This document proposed the "human in command approach," a principle ensuring that decisions pertaining to healthcare planning and treatment remain with humans and do not rely solely on AI tools. This principle was further developed in the European Commission's White Paper on AI, released in 2020. This paper stated that the autonomous behavior of AI during their life cycle could cause significant changes to the software or algorithm, which could have a direct impact on safety and necessitate continuous risk assessment. Additionally, the Civil Law Rules on Robotics suggested that in cases of AI medical errors, responsibility should be proportional to the instructions given to the machine and its level of autonomy — the greater the learning capacity or decisional autonomy of a machine, the greater the responsibility of its human trainers should be [43].

Lack of AI democratization and widening inequality in healthcare

In the context of healthcare, democratization refers to providing universal access to AI-based solutions and ensuring that everyone can benefit equally from this technology. The potential for disparate impact of AI-based solutions on underprivileged groups is a major concern. Research indicates that AI-based solutions can be biased in favor of the privileged and wealthy populations. This means that the AI-based solutions may not be equally accessible to those from other backgrounds. Furthermore, AI systems are only as good as the data used to train them. If the data used to train the AI system does not adequately represent the diversity of the population, then the AI system will produce biased outcomes [1, 44, 45]. For example, if the data used to train the AI system comes predominantly from Caucasian patients, then the AI system may be less accurate in detecting eye diseases in other ethnicities.

Discussion

The use of AI in ophthalmology has been gaining increased attention in recent years due to its potential to improve patient care and outcomes. AI-based technologies have been found to provide accurate and timely diagnosis and management of eye diseases, as well as to reduce medical errors. However, the use of AI in ophthalmology is not without controversy. AI-based systems may not be able to account for the complexity of human decision-making, which can lead to errors or bias, leading to harmful consequences for vision and expose ophthalmologists to liability issues. This narrative review identified and evaluated the controversial aspects and implications of AI use in ophthalmology through the analysis of six key issues. Potential solutions to address them are:

1. Bias and distributional shift can be limited by ensuring that the population of the training data matches the population of the target population, by collecting data of high quality, and by using standardized data collection methods. Researchers should use methods such as stratified sampling and cross-validation with the aim to ensure that the results of ML models are accurate and reliable, upholding alignment with current clinical evidence. Real-life validation of a model's performance is also needed to guarantee that the AI system is up to the task of making accurate and reliable diagnoses in a real-life setting. Additionally, the AI system must be tested in a variety of different settings to allow it to perform consistently in different environments [2, 44].
2. Healthcare organizations should adopt stronger security measures for their data infrastructure, such as encryption and authentication protocols, monitoring and logging systems, and AI-powered solutions certified by third-party organizations. Misuse of health data can lead to a breach of confidentiality and violation of patient privacy, so a secure approach to training AI systems is necessary. Federated learning is the ideal approach for ophthalmology and healthcare due to its capacity to protect patient data while still enabling the AI system to learn [4, 46]. Decentralizing the learning process ensures that data remains local and is not exposed to any third party. As such, model updates can be distributed to each device to improve performance while protecting patient data.
3. To ensure the ethical use of AI-based ophthalmology tools, there is a need for greater transparency in how patient data is collected and used. This should include clear guidelines on data handling and storage, as well as consent forms which explicitly outline how the data will be used. In addition, independent reviews of the algorithms used to power such applications should be conducted on a regular basis to ensure accuracy and fairness.
4. Strategies to address the black-box issue of AI in ophthalmology should be supported, in order to increase the confidence of the medical society in the clinical decisions suggested by the algorithms. One such strategy is to use explainable AI algorithms, which are designed to provide insight into the decision-making process of AI-based algorithms [5, 47]. Explainable AI algorithms are based on the idea that AI-based decisions should be explainable and interpretable, allowing medical professionals to better understand the decisions being made by the AI-based algorithms. Additionally, explainable AI algorithms can be used to provide better insight into the behavior of the algorithm and to identify potential bias or errors in the decision-making process. Another strategy is the use of ensemble methods. Ensemble methods involve combining multiple AI-based algorithms to create a more accurate and reliable decision-making process [6, 48].
5. We believe that AI-based healthcare solutions should be evaluated and approved by appropriate regulatory bodies, country-specific medical boards, and scientific societies as part of the "standard of care." This would

not only allow doctors to use the full potential of AI systems, but also provide a legal safety framework in the event of AI-related medical malpractice. However, this raises the additional question of whether manufacturers and sellers of AI-related products should be held liable under product liability law. Whatever the outcome, it is clear that the use of AI in ophthalmology requires careful consideration of legal and ethical implications.

6. Ensure that AI-based solutions are made accessible to everyone and that any potential for disparate impact on underprivileged groups is minimized. AI systems must be trained on large datasets that accurately represent the diversity of the population, which may be particularly challenging for rare diseases with limited data.

In conclusion, the use of AI in ophthalmology has the potential to improve patient care and outcomes, but it is not without controversy. It is important to consider the ethical and safety implications of AI-based decision-making, and to ensure that any decision made is in the best interest of the patient.

Code availability Not applicable.

Funding Open access funding provided by Università degli Studi di Udine within the CRUI-CARE Agreement.

Data availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest Daniele Veritti, Leopoldo Rubinato, Valentina Sarao, Axel De Nardin, and Gianluca Foresti declare they have no financial interests. Paolo Lanzetta is a consultant to Aerie, Apellis, Bayer, Biogen, Centervue, Novartis, and Roche.

References

1. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25:44–56. https://doi.org/10.1038/s41591-018-0300-7
2. Challen R, Denny J, Pitt M et al (2019) Artificial intelligence, bias and clinical safety. BMJ Qual Saf 28:231–237. https://doi.org/10.1136/bmjqs-2018-008370
3. Godin K, Stapleton J, Kirkpatrick SI et al (2015) Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. Syst Rev 4:138. https://doi.org/10.1186/s13643-015-0125-0
4. Yu K-H, Kohane IS (2019) Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 28:238–241. https://doi.org/10.1136/bmjqs-2018-008551
5. Rajpurkar P, Irvin J, Zhu K et al (2017) CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. https://doi.org/10.48550/ARXIV.1711.05225
6. Wong A, Otles E, Donnelly JP et al (2021) External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med 181:1065. https://doi.org/10.1001/jamainternmed.2021.2626
7. Sarao V, Veritti D, Borrelli E et al (2019) A comparison between a white LED confocal imaging system and a conventional flash fundus camera using chromaticity analysis. BMC Ophthalmol 19:231. https://doi.org/10.1186/s12886-019-1241-8
8. Sarao V, Veritti D, Lanzetta P (2020) Automated diabetic retinopathy detection with two different retinal imaging devices using artificial intelligence: a comparison study. Graefes Arch Clin Exp Ophthalmol 258:2647–2654. https://doi.org/10.1007/s00417-020-04853-y
9. Lee AY, Yanagihara RT, Lee CS et al (2021) Multicenter, head-to-head, real-world validation study of seven automated artificial intelligence diabetic retinopathy screening systems. Diabetes Care 44:1168–1175. https://doi.org/10.2337/dc20-1877
10. Nakayama LF, Kras A, Ribeiro LZ et al (2022) Global disparity bias in ophthalmology artificial intelligence applications. BMJ Health Care Inform 29:e100470. https://doi.org/10.1136/bmjhci-2021-100470
11. Khan SM, Liu X, Nath S et al (2021) A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. Lancet Digit Health 3:e51–e66. https://doi.org/10.1016/S2589-7500(20)30240-5
12. Esteva A, Kuprel B, Novoa RA et al (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542:115–118. https://doi.org/10.1038/nature21056
13. Coventry L, Branley D (2018) Cybersecurity in healthcare: a narrative review of trends, threats and ways forward. Maturitas 113:48–52. https://doi.org/10.1016/j.maturitas.2018.04.008
14. Sulleyman A (2017) NHS cyber attack: why stolen medical information is so much more valuable than financial data. The Independent http://www.independent.co.uk/life-style/gadgets-and-tech/news/nhs-cyber-attack-medical-data-records-stolen-why-so-valuable-to-sell-financial-a7733171.html. Accessed 18 January 2023
15. Landi H (2022) Healthcare data breaches hit all-time high in 2021, impacting 45M people. Fierce Healthcare https://www.fiercehealthcare.com/health-tech/healthcare-data-breaches-hit-all-time-high-2021-impacting-45m-people. Accessed 18 January 2023
16. Zhou Q, Zuley M, Guo Y et al (2021) A machine and human reader study on AI diagnosis model safety under attacks of adversarial images. Nat Commun 12:7281. https://doi.org/10.1038/s41467-021-27577-x
17. Harris B (2019) FDA issues new alert on Medtronic insulin pump security. HealthcareITNews. https://www.healthcareitnews.com/

news/fda-issues-new-alert-medtronic-insulin-pump-security. Accessed 18 January 2023

18. Eddy N (2019) Infusion pump-linked workstations contain critical security flaw. HealthcareITNews. https://www.healthcareitnews.com/news/infusion-pump-linked-workstations-contain-critical-security-flaw. Accessed 18 January 2023
19. Ilyasova NY, Demin NS (2022) Application of artificial intelligence in ophthalmology for the diagnosis and treatment of eye diseases. Pattern Recognit Image Anal 32:477–482. https://doi.org/10.1134/S1054661822030166
20. Perez MV, Mahaffey KW, Hedlin H et al (2019) Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 381:1909–1917. https://doi.org/10.1056/NEJMoa1901183
21. Yim J, Chopra R, Spitz T et al (2020) Predicting conversion to wet age-related macular degeneration using deep learning. Nat Med 26:892–899. https://doi.org/10.1038/s41591-020-0867-7
22. Powles J (2017) Why are we giving away our most sensitive health data to Google? The Guardian https://www.theguardian.com/commentisfree/2017/jul/05/sensitive-health-information-deepmind-google. Accessed 18 January 2023
23. Abdullah YI, Schuman JS, Shabsigh R et al (2021) Ethics of artificial intelligence in medicine and ophthalmology. Asia Pac J Ophthalmol (Phila) 10:289–298. https://doi.org/10.1097/APO.0000000000000397
24. Abramoff MD, Tobey D, Char DS (2020) Lessons learned about autonomous AI: finding a safe, efficacious, and ethical path through the development process. Am J Ophthalmol 214:134–142. https://doi.org/10.1016/j.ajo.2020.02.022
25. Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nat Mach Intell 1:389–399. https://doi.org/10.1038/s42256-019-0088-2
26. Nguyen Q, Woof W, Kabiri N (2023) Eye2Gene Patient Advisory Group, et al Can artificial intelligence accelerate the diagnosis of inherited retinal diseases? Protocol for a data-only retrospective cohort study (Eye2Gene). BMJ Open 13:e071043. https://doi.org/10.1136/bmjopen-2022-071043
27. Ballantyne A (2020) How should we think about clinical data ownership? J Med Ethics 46:289–294. https://doi.org/10.1136/medethics-2018-105340
28. London AJ (2019) Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hast Cent Rep 49:15–21. https://doi.org/10.1002/hast.973
29. Zednik C (2021) Solving the black box problem: a normative framework for explainable artificial intelligence. Philos Technol 34:265–288. https://doi.org/10.1007/s13347-019-00382-7
30. Evans NG, Wenner DM, Cohen IG et al (2022) Emerging ethical considerations for the use of artificial intelligence in ophthalmology. Ophthalmol Sci 2:100141. https://doi.org/10.1016/j.xops.2022.100141
31. Loyola-Gonzalez O (2019) Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view. IEEE Access 7:154096–154113. https://doi.org/10.1109/ACCESS.2019.2949286
32. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215. https://doi.org/10.1038/s42256-019-0048-x
33. Araújo T, Aresta G, Mendonça L et al (2020) DR|GRADUATE: uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images. Med Image Anal 63:101715. https://doi.org/10.1016/j.media.2020.101715
34. Ayhan MS, Kühlewein L, Aliyeva G et al (2020) Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection. Med Image Anal 64:101724. https://doi.org/10.1016/j.media.2020.101724
35. Abràmoff MD, Lou Y, Erginay A et al (2016) Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Invest Ophthalmol Vis Sci 57:5200. https://doi.org/10.1167/iovs.16-19964
36. Huff DT, Weisman AJ, Jeraj R (2021) Interpretation and visualization techniques for deep learning models in medical imaging. Phys Med Biol 66:04TR01. https://doi.org/10.1088/1361-6560/abcd17
37. Jiang H, Xu J, Shi R et al (2020) 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, Montreal, QC, Canada, pp 1560–1563
38. Goodman B, Flaxman S (2017) European Union regulations on algorithmic decision-making and a "right to explanation." AIMag 38:50–57. https://doi.org/10.1609/aimag.v38i3.2741
39. Gunasekeran DV, Zheng F, Lim GYS et al (2022) Acceptance and Perception of Artificial Intelligence Usability in Eye Care (APPRAISE) for ophthalmologists: a multinational perspective. Front Med 9:875242. https://doi.org/10.3389/fmed.2022.875242
40. Price WN II (2017) Artificial intelligence in health care: applications and legal implications. SciTech Lawyer 14:10–13
41. Price WN, Gerke S, Cohen IG (2019) Potential liability for physicians using artificial intelligence. JAMA 322:1765. https://doi.org/10.1001/jama.2019.15064
42. Gerke S, Minssen T, Cohen G (2020) Ethical and legal challenges of artificial intelligence-driven healthcare. In: Artificial Intelligence in Healthcare. Elsevier, pp 295–336
43. European Commission (2020) White Paper on artificial intelligence: a European approach to excellence and trust. European Commission [online]. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en. Accessed 20 Dec 2022
44. Vokinger KN, Feuerriegel S, Kesselheim AS (2021) Mitigating bias in machine learning for medicine. Commun Med 1:25. https://doi.org/10.1038/s43856-021-00028-w
45. Daich Varela M, Sen S, De Guimaraes TAC et al (2023) Artificial intelligence in retinal disease: clinical application, challenges, and future directions. Graefes Arch Clin Exp Ophthalmol 9:1–15. https://doi.org/10.1007/s00417-023-06052-x
46. Xu J, Glicksberg BS, Su C et al (2021) Federated learning for healthcare informatics. J Healthc Inform Res 5:1–19. https://doi.org/10.1007/s41666-020-00082-4
47. Singh A, Sengupta S, Lakshminarayanan V (2020) Explainable deep learning models in medical image analysis. J Imaging 6:52. https://doi.org/10.3390/jimaging6060052
48. Beltrami AP, De Martino M, Dalla E et al (2022) Combining deep phenotyping of serum proteomics and clinical data via machine learning for COVID-19 biomarker discovery. IJMS 23:9161. https://doi.org/10.3390/ijms23169161

Authors and Affiliations

Daniele Veritti[1] · Leopoldo Rubinato[1] · Valentina Sarao[1,2] · Axel De Nardin[3] · Gian Luca Foresti[3] · Paolo Lanzetta[1,2]

✉ Daniele Veritti
daniele.veritti@uniud.it

[1] Department of Medicine – Ophthalmology, University of Udine, Udine, Italy

[2] Istituto Europeo di Microchirurgia Oculare – IEMO, Udine, Italy

[3] Department of Mathematics, Informatics and Physics, University of Udine, Udine, Italy