Title
Artificial intelligence extension of the OSCAR-IB criteria

Permalink
https://escholarship.org/uc/item/5ts6f6cc

Journal
Annals of Clinical and Translational Neurology, 8(7)

ISSN
2328-9503

Authors
Petzold, Axel
Albrecht, Philipp
Balcer, Laura
et al.

Publication Date
2021-07-01

DOI
10.1002/acn3.51320

Peer reviewed

REVIEW

Artificial intelligence extension of the OSCAR-IB criteria

Axel Petzold[1,2] (iD), Philipp Albrecht[3] (iD), Laura Balcer[4], Erik Bekkers[5], Alexander U. Brandt[6] (iD), Peter A. Calabresi[7] (iD), Orla Galvin Deborah[8], Jennifer S. Graves[9] (iD), Ari Green[10], Pearse A Keane[1], Jenny A. Nij Bijvank[2], Josemir W. Sander[11,12,13], Friedemann Paul[14], Shiv Saidha[7], Pablo Villoslada[15] (iD), Siegfried K Wagner[1] (iD), E. Ann Yeh[16], the IMSVISUAL, ERN-EYE Consortium,[a]

[1]Moorfields Eye Hospital, City Road, The National Hospital for Neurology and Neurosurgery, Queen Square, UCL Queen Square Institute of Neurology, London, UK
[2]Neuro-ophthalmology Expert Center, Amsterdam UMC, The Netherlands
[3]Department of Neurology, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany
[4]Departments of Neurology, Population Health and Ophthalmology, NYU Grossman School of Medicine, New York
[5]AMLAB, Amsterdam, The Netherlands
[6]Department of Neurology, University of California, Irvine, California
[7]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland
[8]Retina International, Dublin, Ireland
[9]Department of Neurosciences, UC, San Diego, California
[10]Department of Neurology, University of California San Francisco, San Francisco, California
[11]NIHR UCL Hospitals Biomedical Research Centre, UCL Queen Square Institute of Neurology, London, UK
[12]Chalfont Centre for Epilepsy, Chalfont St Peter, UK
[13]Stichting Epilepsie Instellingen Nederland (SEIN),Heemstede, The Netherlands
[14]Experimental and Clinical Research Center, Max Delbrück Center for Molecular Medicine and Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Germany
[15]Institut d'Investigacion Biomediques August Pi Sunyer (DIBAPS) and Hospital Clinic, University of Barcelona, Barcelona, Spain
[16]Division of Neurology, Department of Pediatrics, Hospital for Sick Children, Division of Neurosciences and Mental Health SickKids Research Institute, University of Toronto, Canada

Abstract

Artificial intelligence (AI)-based diagnostic algorithms have achieved ambitious aims through automated image pattern recognition. For neurological disorders, this includes neurodegeneration and inflammation. Scalable imaging technology for big data in neurology is optical coherence tomography (OCT). We highlight that OCT changes observed in the retina, as a window to the brain, are small, requiring rigorous quality control pipelines. There are existing tools for this purpose. Firstly, there are human-led validated consensus quality control criteria (OSCAR-IB) for OCT. Secondly, these criteria are embedded into OCT reporting guidelines (APOSTEL). The use of the described annotation of failed OCT scans advances machine learning. This is illustrated through the present review of the advantages and disadvantages of AI-based applications to OCT data. The neurological conditions reviewed here for the use of big data include Alzheimer disease, stroke, multiple sclerosis (MS), Parkinson disease, and epilepsy. It is noted that while big data is relevant for AI, ownership is complex. For this reason, we also reached out to involve representatives from patient organizations and the public domain in addition to clinical and research centers. The evidence reviewed can be grouped in a five-point expansion of the OSCAR-IB criteria to embrace AI (OSCAR-AI). The review concludes by specific recommendations on how this can be achieved practically and in compliance with existing guidelines.

Introduction

Sophisticated artificial intelligence (AI)-based algorithms not only enable discrete layer segmentation of retinal optical coherence tomography (OCT) scans, but can identify novel retinal features. A prominent recent example is automatic clinic referral via deep learning based on retinal layer analysis.[1] The current OSCAR-IB quality control

(QC) criteria take algorithmic quality components into account (the A-criterion) but the criteria are not adapted to AI-based algorithms.[2] Given the recent success and rapid developments in this field, it is timely to build on the OSCAR-IB QC criteria to address the challenges of AI and big data specifically.

To this purpose, it is critical to acknowledge that accuracy is paramount to the interpretation of retinal OCT in neurological disease. Judgments are highly dependent on quantitative data of individual retinal layers. Key components are thickness, degree of change, and alteration of the topography. The retinal layer thickness changes seen in neurological disorders are much more subtle[3-7] than the pathologies seen in the ophthalmologic diseases, now successfully detected by AI-based methods.[1,8,9] For neurodegenerative diseases, relevant annual retinal layer atrophy rates are just above the axial image resolution of contemporary spectral-domain and swept-source OCT techniques.[5] For this reason, image QC is paramount. Over the past decade, big OCT data have accumulated in neurodegenerative and neuroinflammatory diseases. These data are attractive for the development of AI strategies. The expectation is to improve the accuracy of OCT-based quantification, diagnostic sensitivity and specificity, discover novel surrogates for monitoring disease progression as well as outcome metrics for clinical trials. Fully automated AI-based strategies are transferable from highly specialized services to primary care. The test throughput is also scalable to include, for example, high street opticians. Both will aid with the logistics of patient care through local centers.

In 2012, we proposed the first consensus OCT QC criteria, OSCAR-IB.[2] The name served as a mnemonic for seven distinct QC criteria to be remembered; (i) **O**bvious errors, (ii) **S**ignal strength, (iii) **C**entration of scan, (iv) **A**lgorithm failure, (v) **R**etinal pathology, (vi) **I**llumination, and (vii) **B**eam placement. This was followed by international validation[10] and endorsement to reporting guidelines.[11] The OSCAR-IB QC criteria were developed in a multiple sclerosis (MS) network, and have since been broadly accepted. This success is at least in part due to the demand for clarity and transparency on being practical to "get simple things right". A similar approach is warranted for the role of AI in relation to OCT data in neurodegenerative diseases.

Review of evidence for need of QC

AI-based strategies are at risk for propagation of systematic errors, imbalance, and bias due to subtle differences in data acquisition or postprocessing. There is justified concern about the lack of QC standards for big data.[12,13] The rise of big data is in part driven by the hope of improving P4 medicine: predictive, preventive, personalized, and participatory care.[14] For example, integrative prediction models encompassing as many as 63 variables have been proposed to enable personalized predictions of individuals' outcomes and guide treatment decisions in myeloproliferative neoplasms.[15] The implications that AI-driven approaches will have for individuals can easily be influenced by bespoke sources of bias fed into the model as discussed below.

Importantly, minimal variation in image acquisition can cause substantial errors in the quantitative data.[16,17] Strategies based on AI are excellent in recognizing changes between images, but do not necessarily know how the human OCT operators have acquired an image. This can mislead the AI-based strategy with a downstream effect of possible misdiagnosis, mismanagement, and harm. The risk for such a situation to occur increases with rapidly rising numbers of OCT scans to be evaluated. It may introduce systematic errors if imbalances exist between populations and centers, for example, due to service capacity issues or automation of our health-care systems. The possible medico-legal ramifications are also evident.

Review of failures and successes of retinal OCT in neurological diseases

To date, OCT data in MS and related disorders are most consistent as most reports adhered to the OSCAR-IB QC criteria and followed the APOSTEL reporting guidelines.[2,11] Results are more heterogeneous for other neurological diseases because of lack of standardization. There is evidence for an early publication bias in Alzheimer disease (AD) reports.[6] Subsequent data were not supportive of the earlier enthusiasm. Few of the reports on AD followed a rigorous QC approach. This similarly applies to reports of OCT in Parkinson disease (PD),[18,19] amyotrophic lateral sclerosis (ALS),[20-23] stroke,[24] epilepsy,[4,25] and schizophrenia.[26]

Review of successes of AI in neurological diseases

There are critical successes for the use of AI in neurological disease. For example, urgent triaging of individuals from brain imaging to neurosurgery[27]; earlier diagnosis of AD[28]; identifying suitable candidates for epilepsy surgery[29]; and regulation of adaptive deep brain stimulation in movement disorders.[30] Imaging-based trial outcome measures in neurology include almost all neurodegenerative, neurovascular, and neuroinflammatory conditions alongside tumors.[31] Imaging data have become multimodal. This adds to complexity and time needed by

human readers and reporting. Likewise, histological data can now be used for machine and deep learning.[32]

The review committee

The ophthalmological community has driven advances in AI-based analysis of retinal OCT.[1] The committee for this review has been expanded (Table 1). We included representatives from neurological sub-specialties who have used retinal OCT data for diagnostic and prognostic purposes, as well as treatment trial outcome measures. We have also engaged with experts in the fields of AI and bio-engineering and bio-statistics and two not-for-profit organizations (www.ern-eye.eu and www.imsvisual.org).

The patient voice

The importance of patient involvement as a key stakeholder has been recognized[33] and contributed to the development of conceptual models.[34] On a day-to-day practical level, the experience has demonstrated that individuals tolerate retinal OCT well. It is noninvasive, noncontact, quick, and provides instant feedback. The possibility to display images directly to the individuals and discuss changes has given them more confidence and insight in their care.[35] This good partnership has helped in working together to build trust, supporting treatment decisions and making OCT scans available for research. There is a need to maintain this mutual trust at a time where the immense amount of data accumulated now permit AI-inspired projects on big data. None of this is possible without patient participation, their consent, and feedback. A key concern of patients and their advocates is that their data will be misused. Individuals have a higher level of confidence in not-for-profit stakeholders than in government or private companies.[36]

Data protection and privacy

Due to the requirement of very large training datasets for optimal performance, most current clinical AI systems have been developed using routinely collected data which have been anonymized. Anonymization of medical images presents specific challenges, however, particularly images of individually unique structures such as the neurosensory retina.[37] Even when carefully anonymized, there is at least a theoretical risk of re-identification for such images, either now or with some future technology.[38] Therefore, we recommend a multistep approach to addressing data protection and privacy. Firstly, retinal OCT scans should be anonymized according to current national and international standards.[39] This includes removal of any imaging meta-data such as patient names, dates of birth, or medical record numbers, obscuration of hospital visit dates, plus careful consideration of any associated clinical meta-data (e.g., merging of categories/classes if they contain only a limited number of examples).[40] Secondly, a range of additional safeguards should be put in place. Technical safeguards include the requirement to store data in trusted research environments with access controls and audit logs; contractual safeguards include prohibitions against linkage or attempted re-identification of data. Importantly, every attempt should be made to minimize the data shared to that required for the clinical or research purpose – this is a fundamental principle of much data protection regulations, including the European General Data Protection Regulation (GDPR). Finally – and perhaps most importantly – it is vital to engage in patient and public engagement and involvement at the earliest possible stage. This includes making patients aware that their data are being used for research, publishing study protocols, and giving patients the opportunity to opt-out. By adopting a cautious and engaged approach such as this, we believe it is possible to reduce any data protection risks while maximizing the potential for future patient benefit. In the future, a range of technical solutions, including federated learning and homomorphic encryption, should help further mitigate these risks.[41]

Search strategy and selection criteria

We reviewed three databases, PubMed, Web of Science, and Google Scholar, between 01 January 1963 and 23 April 2020 without language restriction. We chose the English version of a manuscripts if the same group had published similar data in Dutch, French, German, Italian, or Spanish. The search terms used were "optical coherence tomography" or "OCT" combined with "artificial intelligence", "machine learning", "deep learning", "multiple sclerosis", "optic neuritis", "dementia", "Alzheimer", "Parkinson", "motor neuron disease", "amyotrophic lateral sclerosis", "stroke", "cerebrovascular accident", "schizophrenia", "patient voice". We also reviewed articles included in three systematic reviews previously conducted.[3,5,8]

Methods

Firstly, we reviewed the original OSCAR-IB criteria to clarify which of the QC failures require an individual to be re-assessed or to be excluded if, for example, post hoc homogenization approaches fail. Having to recall a patient for a failed test is not desirable, is problematic, and is expensive. Secondly, we reviewed approaches to rectifying QC failures by image postprocessing. Thirdly, we examined the outcome of our AI-based methods for

Table 1. Expertise of the literature review committees.

Expertise	Members
Patient voice	Nils Wiegerink (patient), Russel Wheeler (patient advocate), Christiaan Waters (President of patient organization), Avril Daily (Retina International, ERN-EYE), Christina Fasser (Retina International, ERN-EYE), Orla Galvin Deborah (Retina International, ERN-EYE), and Oshakuade (Retina International, ERN-EYE)
AI	Erik Bekkers, Siegfried Wagner, and Pearse Keane
Public relation & Media	Avril Daily
ALS	Philip Albrecht and Orhan Atkas
Alzheimer disease	Thomas Wisnewski
Epilepsy	Josemir W. Sander
Parkinson disease	Alexander Brandt, Philipp Albrecht, and Orhan Atkas
Stroke	Shadi Yaghi and Arvind CHANDRATHEVA
Multiple Sclerosis	Alexander Brandt, Peter Calabresi, Laura Balcer, Elliot & Tree Frohman, Friedeman Paul, Ari Green, Pablo Villoslada, Axel Petzold, Philipp Albrecht, Orhan Aktas, E. Ann Yeh, Bernardo Sanchez-Dalmau, Jen Graves, Shiv Saidha, Robert Bermel, IMSVISUAL, and ERN-EYE
Rare Diseases	Alexander Brandt, Philipp Albrecht, Orhan Atkas, Axel Petzold, Friedeman Paul, Frederike Oertel, Alexander Brandt, E. Ann Yeh, Avril Daily (Retina International, ERN-EYE), Christina Fasser (Retina International, ERN-EYE), Orla Galvin Deborah (Retina International, ERN-EYE) Oshakuade (Retina International, ERN-EYE), Bernardo Sanchez-Dalmau, and ERN-EYE
Ophthalmology	Bernardo Sanchez-Dalmau, Pearse Keane, Siegfried Wagner and ERN-EYE
Neuro-ophthalmology	Fiona Costello, Ari Green, Axel Petzold, Laura Balcer, Bernardo Sanchez-Dalmau, Jen Graves, and ERN-EYE
OCT	Alexander Brandt, Frederike Oertel, Hannah Zimmerman, Philipp Albrecht, Orhan Atkas, Peter Calabresi, Axel Petzold, Jen Graves, Rachel Nolan-Kennedy, Laura Balcer, Shiv Saidha, Bernardo Sanchez-Dalmau, Pablo Villoslada, and Robert Bermel
OCTA	Benjamin Knier, Shiv Saidha, Axel Petzold and IMSVISUAL
Clinical trials OCT QC	Alexander Brandt, Friedeman Paul, Sven Schippling, Axel Petzold, Robert Bermel, Laura Balcer and IMSVISUAL
Statistics and epidemiology	David Crabb, Gary Cutter, Laura Balcer, Jen Graves, Rachel Nolan-Kennedy, Kathryn Fitzgerald, and Zhaoxia Yu

irregularities, identical to the approach taken in the original OSCAR-IB report.[2] The terminology of terms explicitly related to AI is summarized in Table 2.

Defining QC for AI: RASCO

Firstly, there is a clear and justified fear of the misuse of big data.[37] Secondly, the patient–physician relationship must be supported to provide an optimal experience. Thirdly, demonstration of the capability of the AI strategy enhances the ability to produce high-quality and relevant effectiveness research. Fourthly, it promotes accountability. Fifthly, it provides grounding for the production of reproducible studies. Together, the definition of QC for AI can be summarized by five pillars which were named individually or in combination in the literature reviewed (Figure 1). The mnemonic, **RASCO,** stands for **R**eproducibility (**R**), **A**ccountability for decisions made (A), to be **S**upportive of the patient–physician relationship (S), **C**apability ranging from machine learning (ML)-supported OCT quality control assessment to time and resource-efficient decision-making (C), and **O**penness with and trust in public opinion (O) is pertinent, given the personal data protection issues discussed above.[42]

Big data

The utility of AI in medical applications is more dependent on data quality than quantity. The new research field of big data has contributed considerably to the advancement of medical science by analysis of large datasets. Until recently, it had not been easy to accumulate enough data to create a large data repository and analyses were too complicated or lacked statistical and computer power. A critical area of weakness of big data can be the granularity and quality of the source data entered. In essence, the quality of outputs or results of AI-based assessments should not be expected to exceed the underlying quality of the data being analyzed (input data). This underpins the importance of maintaining the highest standards of quality, even in the AI space. As we are at the dawn of AI for OCT research, one of our aims is to facilitate the generation of the high-quality data needed for future research in the field.

Prospective OCT image QC

Each OCT scan should prospectively be labeled as QC fail or not. There are several reasons why a scan may fail QC. Each failed scan should be annotated with a complete list of reasons. An efficient way is to use the capital letters of the OSCAR-IB criteria.[10] To avoid a potential bias by eliminating scans from the sickest individual who may have difficulties with the test, this needs to be explicitly noted. Retinal and systemic co-morbidities require careful clinical evaluation with more in-depth ophthalmic phenotyping than hitherto done in most neurological studies.[43]

Table 2. Terminology and basic concepts.

Artificial Intelligence (AI)	Computer or machine-based intelligence which enables "learning" and "problem solving"
Machine learning (ML)	One subset of AI. Typically algorithms improve automatically through experience after training on a dataset. ML can be supervised or unsupervised
Deep learning	One subset of ML essentially based on artificial neuronal networks. Very efficient and the basis of most contemporary AI-based studies on image recognition
Supervised	Supervised ML works on a labeled training dataset (e.g., OSCAR-IB OCT scans) and reproduces the desired outcome
Unsupervised	Unsupervised ML tries to discover previously undetected patterns in a dataset
Over-fitting	Over-fitting can be a problem with ML, a source of over-enthusiastic reporting and reason for lack of reproducibility

QC failure may result in two situations: (1) where an error can be corrected at postprocessing; and (2) where the error requires recalling and repeating the test.

Human-led OCT image QC is a time-consuming task, so it is desirable for this to be performed using AI strategies. We suggest making use of the above-described annotation of failed scans for ML of OSCAR-IB criteria (Figure 2). This will enable the training of future AI algorithms to separate good from insufficient quality OCT scans. The next application step within the pool of scans designated as being of inadequate quality will be to identify those scans which may be subjected to post-acquisition correction approaches, thereby making them high quality, and enabling their safe and accurate utilization. This is a crucial step as it allows for AI training in auto-correction. Scans which failed OSCAR-IB and are not correctable must be excluded from any further AI steps.

Taken together this leaves a staged approach to QC in AI: (1) Automated AI OCT QC rating using validated OCT QC criteria[2,10]; (2) where possible AI QC correction during image postprocessing, or if not possible patient recall for repeat acquisition; and (3) final step of AI-based image analysis. This will typically make use of pattern recognition and be the key step forward for the primary research questions. A limitation to keep in mind is that

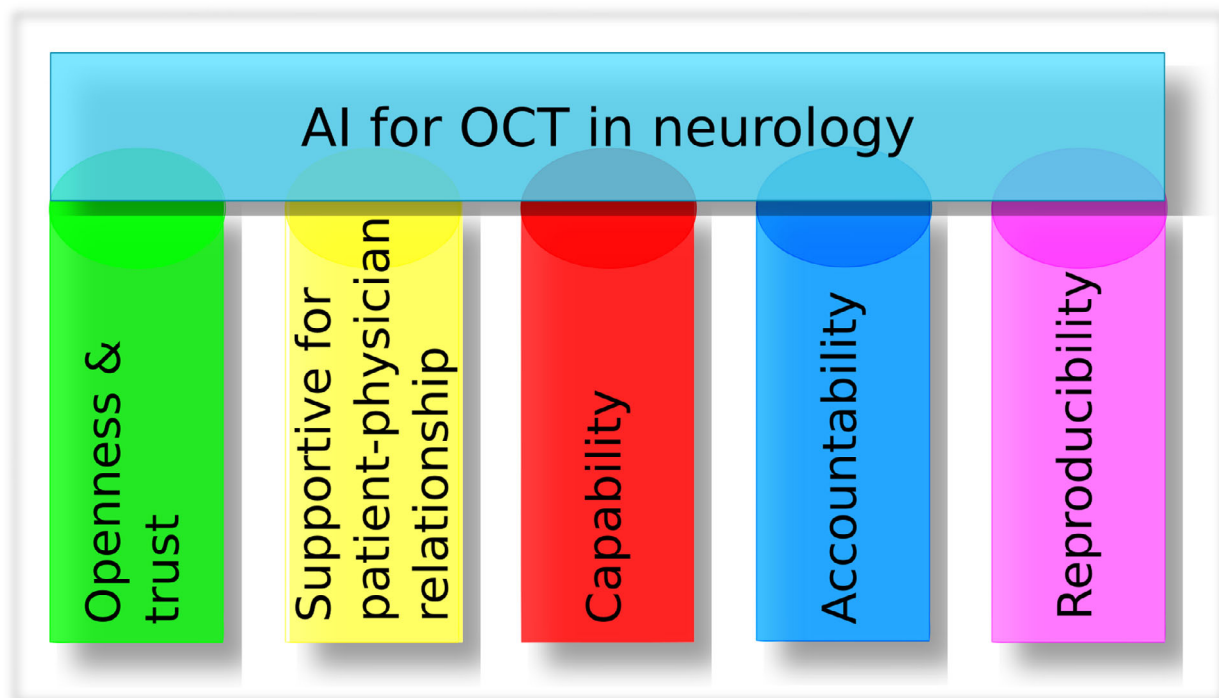**Figure 1.** The goal of quality control in Artificial Intelligence (AI) rests on five pillars: RASCO. (1) Openness with and trust in the public opinion, (2) to be Supportive for the patient–physician relationship, (3) Capability ranging from machine learning (ML)-supported OCT quality control assessment to time and resource-efficient decision-making, (4) Accountability for decisions made, and (5) Reproducibility (RASCO).

presently it is not possible to exclude ophthalmological co-morbidities without a clinical assessment.

AI artifact vulnerability

We have not identified reports on the vulnerability of algorithms to misclassification due to the use of different OCT devices or software versions. Even seemingly small updates have the potential to cause significant differences which if left unnoticed can bias results.[44]

Ground truth

The definition of ground truth is disease-specific. It should be stated explicitly how the ground truth was defined. At the minimum for AD and other neurodegenerative dementias, epilepsy, MS, optic neuritis (ON), neuromyelitis optica spectrum disorder (NMOSD), PD, adherence to consensus investigation protocols, and diagnostic criteria will be required. As diagnostic criteria in most neurological diseases are regularly updated, this needs to be taken into account.

Statistics

The descriptive statistics reviewed were mostly based on binary classifiers such as a disease is present "yes/no". These models should include a comment on proportional bias.[45] This is needed to interrogate how much the AI-based prediction agrees with the ground truth. The definition for an acceptable ground truth needs to include the level of evidence on which it was based. For binary and multiclassifier models, the degree of inter-rater agreement should be stated to permit judging on how stable the ground is.

Graphs can be presented in a way that allows judgment of the degree of over-fitting and underestimation relevant in comparing differences between AI and ground truth. Many studies used Bland-Altman plots[46] or analyzed the performance of AI and ground truth based on a receiver operator curve (ROC)-based area under the curve (AUC). This gives comparative estimates of sensitivity, specificity, and the positive predictive value (PPV) as a measure of overall accuracy. This is particularly relevant for relatively rare diseases. It was recommended to complementing area
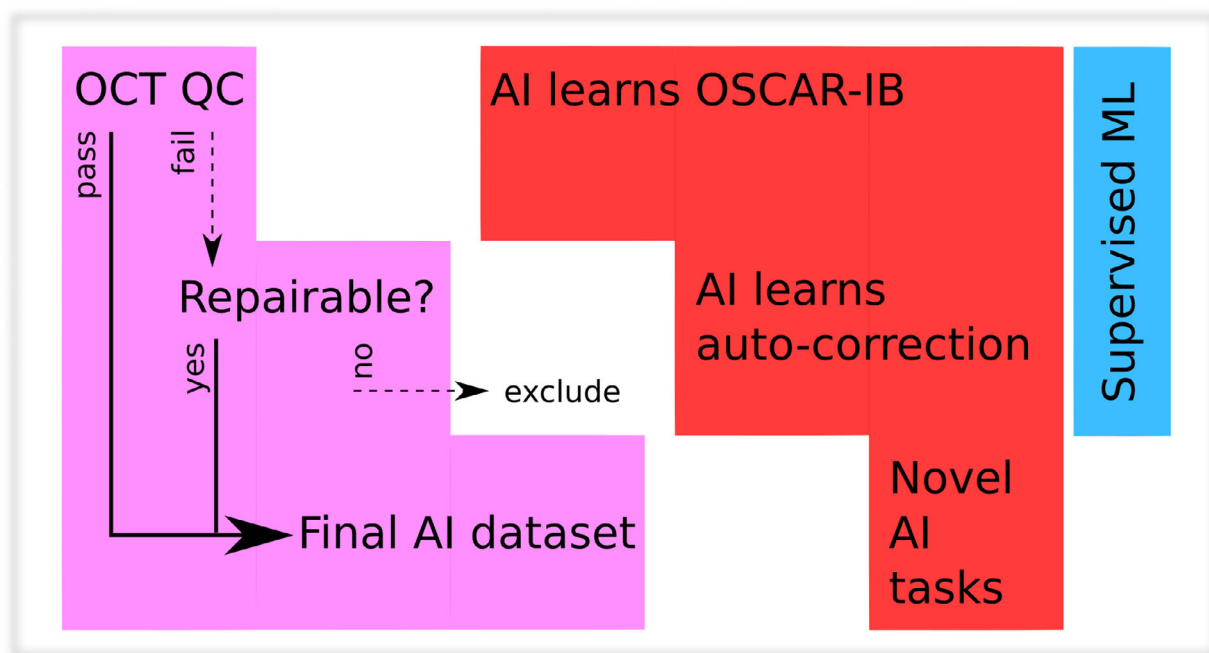

Figure 2. The capability of AI to contribute interpreting OCT images depends on the optimization of each step contributing to the decision tree. The first step relates to the quality of the raw data. Validated QC criteria for OCT image have been summarized as OSCAR-IB.[2] The ground truth of whether or not an OCT passes QC is based on human assessment. The seven OSCAR-IB criteria for QC rejection by a human assessor can directly be used to train AI. Annotation of corrupted OCT scans permits for two outcomes: (1) image postprocessing and repair of artifacts or (2) complete rejection and (if feasible) recall of patient and OCT rescan. Only a dataset that passed OCT image QC should be used for further AI interpretation.

under the curve ROC (AUROC) values by precision-recall (precision is the PPV and recall is the sensitivity in the AI literature) curves (AUPRC).[8] This was found to be of relevance for unbalanced datasets (substantially more subjects in one of the groups compared).

Cut-off level calculation

Reporting of calculation of cut-off values included the use of independent cohorts, a graphical ROC-based approach, the Youden Index, k-fold cross-validation, or hold-out validation approaches to obtain accurate estimations of AI-based cut-off performance.

Power calculations

We did not yet find consistent reports of the inclusion of power calculations to studies, which are relevant for randomized controlled trials using AI-based outcome measures.[47] It is recommended that sample size estimates be performed before developing an algorithm and repeated after study completion. The gain in power, meaning a more robust statistical result, is just as informative for future research as the potential cost savings by optimizing numbers. Lastly, the standardized effect size, likely to come from AI, was recommended to be aligned with distribution, and anchor health economics to inform clinical trials on what will be a realistic difference.[47,48]

Cohort description

On review, cohort descriptions were mostly conform to contemporary standards on demographic characteristics. Cohort descriptions are relevant for AI, and will also greatly limit/determine the usability of the system. This reinforces the need to build on successful initiatives such as the established Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online TeleHealth (CONSORT-EHEALTH)[49] and the CONSORT-AI guidelines.[50] Documentation of developmental changes is also relevant throughout pediatric care and at the transition to adult care. A novel source of potential biases related to the disease diagnostic criteria used. For many of the conditions of interest to retinal OCT, subsequent diagnostic criteria were published over the past decades. While generally aimed at improving practicability, sensitivity, and specificity, this bears the risk that cohorts which are supposed to have the same disease can be quite different in their composition. For example, subsequent diagnostic criteria for MS, AD, and PD have profoundly reshaped the patient base for clinical research over time. Contemporary cohorts tend to be milder than historical cohorts.[51] Clinical trial populations are different from observational studies. The co-morbidity burden is relevant. Relevant items for the pooling of big data are: reporting of the exact diagnostic criteria, a detailed listing of all inclusion and exclusion criteria, recruitment, referrals, and capability of individuals to comply with the examinations. Minimization of the risk of systematic bias will ensure that validation of AI in other cohorts will be comparable.

Validation

For all AI algorithm development efforts, data used for this purpose should be clearly described for discovery and replication analyses. To avoid obtaining a distorted/biased view on performance, data that are used for validation (e.g., to assess performance) should not have been used during algorithm development. There is a real risk for over-fitting the AI models. It was recommended to perform a validation of the algorithm with the aid of a comparable out-of-sample population. Each AI classification scheme should be rated to whether or not an **external validation** was performed. This can be supported by publishing details on the building blocks of the AI. Relevant are precise and meaningful definitions on a functional and performance level. This entails a detailed description of the AI architecture, hyper-parameters, as well as details on how the available data were used to train such systems, preferably via open access code repositories. One of the challenges with AI at regulatory level but also at the clinical level found was the fact that neural networks can learn with data and improve their performance. For this reason, it was suggested to define in advance which type of learning is allowed without requiring validation, approval, or lack clinical risks.

Human versus machine and human with machine

It was reported that AI might improve over human performance in terms of accuracy and speed.[52] For this "machine versus human" approach reporting included data on sensitivity, specificity, positive, and negative predictive values including the 95% confidence intervals (CI) and the numbers on which the calculations were based. These data permit to answer the question if AI can outperform humans not only as seen with Chess and "Go" games,[53,54] but also for classification of retinal appearances.[55]

There is a second, equally important question to be answered. How can AI be used to enhance human performance?[56] Therefore, it was recommended to test if there is a synergistic effect if the AI and human approach are combined. This is typically referred to as human–AI symbiont/symbiotic.[57]

Clinical practice

The relevance of potential clinical downstream effects has been recognized.[58-60] The big chances are to reduce the burdens on physicians and help with service capacity issues. It was recommended to indicate if an algorithm is useful for clinical practice. This requires to test the algorithm in clinical routine. There were different levels at which algorithms added information: on an individual level, on a cohort level, or for screening purposes. There can be important consequences for daily clinical care and health systems. Concerns reported related to misdiagnosis and practicability. This has implications for disease classification.[61]

Guidelines

At the time of the review, the following guidelines were relevant: APOSTEL, TRIPOD-AI, CONSORT-AI,[50] SPIRIT-AI, and STROBE[62] guidelines. They are regularly updated and latest information can be found on the equator network website at www.equator-network.org/.

Open access

Open access and data sharing were found to be essential for accountability and reproducibility. The classified sample dataset is just as valuable as the developed algorithm. Datasets can potentially be used by other groups, to facilitate even greater improvements. Accordingly, data availability may accelerate development in the field. Algorithms can also be transferable and codes can be shared.

Black box

On review, there is a need to understand on what basis an algorithm came to a particular conclusion.

Review of the black box of OCT in neurology

There are a few careful predictions one can make regarding the "black box" for neurodegeneration based on anatomy and progression pattern.

Firstly, anatomically each area in the retina is connected by axons with a corresponding area in the brain because of the hard-wired retino-cortical projections.[3,5] The location of damage to the brain will determine the location of expected OCT changes in a determined area of the retina, a "Region of Interest" (ROI). It has been shown that an ROI-based approach to quantification of inner retinal layer atrophy is superior to occasionally performed sector analysis[5-7] or the generally adopted global averaged approach because it can mask small areas of atrophy.[63,64]

Secondly, the progression pattern is determined by location and size of a lesion damaging the retino-cortical projections.[64-66] The speed of progression is highest and the area of inner retinal atrophy most extensive with direct retrograde axonal degeneration as seen with optic nerve damage. More distal brain damage will still cause localized atrophy in the retina by a mechanism called retrograde trans-synaptic axonal degeneration.[65,67] On sequential OCT imaging, the time course of atrophy is shorter with small brain lesions compared to larger brain lesions.[64] It can be anticipated that a smoldering, slowly enlarging brain lesion will continue to drive the expansion of OCT detectable retinal atrophy.[68]

Thirdly, inflammatory activity in demyelinating disease has been related to transient increase of the inner nuclear layer (INL) volume.[69-73] Part of this INL thickening is related to the development of microcystic macular edema (MME).[69,70,74] Vitreous traction had been implicated, but is not required for the development of MMO.[75] In most (>80%) cases, MMO is a transient phenomenon.[74] In the remainder, it remains static over the years[74,76] and is considered by some to represent a retrograde maculopathy[77] due to axonotmesis in the anterior visual pathways as known from experimental models.[78]

Fourthly, there are qualitative observations on the OCT images, which have not yet been translated into automated forms of quantification. One example is the presence of hyper-reflective spots.[74] There are two types of these hyper-reflective spots on OCT, and one is static and particularly visible at the upper and lower border of the INL. With the advent of OCT-Angiography (OCTA) and adaptive optics, it has become clear that they represent reflectivity changes from the inner retinal vasculature.[79] There is at least another type of hyper-reflective spot noticed on serial OCT images, which migrates vertically through the retina.

Fifthly, the vitreous has specific OCT signal characteristics which can be reliably quantified from the raw image data.[80,81] The technique is useful in neurological disease affecting younger adults where the vitreous body still adheres to the retina such as the majority of people with MS.[82] The evaluation of the raw OCT data, rather than analysis of an already postprocessed screen image, is required due to signal changes.

Sixth, advanced image shape analyses now permit for quantitative data on qualitative characteristics of the optic disc. The technique has proved valuable in idiopathic intracranial hypertension[83,84] and possibly also idiopathic moyamoya angiopathy.[85] Similarly, the presence of peripapillary hyper-reflective ovoid mass-like structures (PHOMS) is a novel OCT finding,[86] which akin to MMO

remained undetectable on conventional funduscopic examination. Likewise, shape analysis of the fovea has become possible.[87,88]

Seventh, functional assessment of individual retinal layers by OCT is possible using, for example, a dark adaptation.[79] One can anticipate that with the availability of OCTA, the retinal equivalent of a blood-oxygen-level-dependent (BOLD) signal for the brain will emerge.[89] Increased, localized retinal metabolic activity will demand increased oxygen supply and cause elevated perfusion of the microvasculature.[79] Pioneering data on OCTA in MS imply that there is a need for AI-supported QC to exclude artifacts.[90-92] This will be relevant for reliable quantitative OCTA data on the retinal microvasculature

which may help to differentiate between disease entities such as MS and NMOSD.[93]

Eighth, inter-eye differences of individual retinal layers are an attractive and highly sensitive method to screen for optic neuritis and MS.[43,94-101] Expanding on these findings, there is a field for AI-based analyses of patterns of retinal asymmetry in MS.[43]

Lastly, reflectivity changes of individual layers can be interrogated to estimate tissue properties indirectly.[102,103]

Based on the above combination of numerous quantitative and qualitative changes in retinal (neural and non-neural tissue) architecture in neurological disease, there are promising avenues for a supervised ML approach to the analysis and interpretation of OCT data. Equally, for researchers who prefer to follow a nonsupervised ML approach, the committee recommends checking if findings may be explainable, at least in part, by the above summary of anatomically, biologically, and pathologically plausible observations.

Summary

In summary, we reviewed several levels of AI-based OCT research in neurology. The main points arising from this review are summarized in Table 3 and based on five pillars (RASCO). The practical conclusions from the multiple levels of evidence reviewed and the summary table may be found helpful on a practical level for future research in the field.

Table 3. Summary of key points from the literature review on OCT and AI research in neurology. The categories are based on the mnemonic "RASCO". This table may be found helpful in guiding future use of the reported data for AI-based studies.

Question	Answer
REPRODUCIBLITY	
OSCAR-IB OCT quality control compliant?	Yes / No
APOSTEL OCT reporting guideline compliant?	Yes / No
TRIPOD-AI compliant?	Yes / No
CONSORT-AI compliant?	Yes / No
SPIRIT-AI compliant?	Yes / No
STROBE compliant?	Yes / No
ACCOUNTABILITY	
Training, test & validation sets explained?	Yes / No
Potential for bias[1] in big data addressed?	Yes / No
Ground truth explicitly stated?	Yes / No
Statement on proportional bias given?	Yes / No
Precision-recall curves provided?	Yes / No
Power calculations included?	Yes / No
SUPPORTS PATIENT DOCTOR RELATIONSHIP	
Patient voice included?	Yes / No
Conflicts of interest, including political, explained?	Yes / No
Shows how AI is used to enhance human performance?	Yes / No
Tested in clinical practice?	Yes / No
CAPABILITY OF ALGORITHM	
Unsupervised AI?[2]	Yes / No
Has QC capabilities?[3]	Yes / No
Provides a glimpse into the black box?[4]	Yes / No
Vulnerabilities of AI explained?[5]	Yes / No
External Validation?	Yes / No
OPENNESS	
Data availability statement?	Yes / No
Data deposited in repository?	Yes / No
AI deposited in open access code repository?	Yes / No

[1]Sources of bias can be analytical, clinical, statistical, imbalance in populations, or centres where the original research was conducted.
[2]SeeTable 2
[3]SeeFigure 2
[4]See Figure 3
[5]Vulnerabilities to artifacts, use of different devices, hard- or software updates of the OCT device.

Isabelle Audo, Jagannadha Avasarala, Daly Avril, Francesca R. Bagnato, Brenda Banwell, Amit Bar-Or, Raed Behbehani, Arnaldo Belzunce Manterola, Jeffrey Bennett, Leslie Benson, Jacqueline Bernard, Dominique Bremond-Gignac, Josefine Britze, Jodie Burton, Jonathan Calkwood, William Carroll, Arvind Chandratheva, Jeffrey Cohen, Giancarlo Comi, Christian Cordano, Silvana Costa, Fiona Costello, Ardith Courtney, Anes Cruz-Herranz, Gary Cutter, David Crabb, Lindsey Delott, Jerome De Seze, Ricarda Diem, Helene Dollfuss, Nabil K. El Ayoubi, Christina Fasser, Carsten Finke, Dominik Fischer, Kathryn Fitzgerald, Pedro Fonseca, Jette L. Frederiksen, Elliot Frohman, Teresa Frohman, Kazuo Fujihara, Iñigo Gabilondo Cuellar, Steven Galetta, Elena Garcia-Martin, Gavin Giovannoni, Brigita Glebauskiene, Inés González Suárez, Gorm Pihl Jensen, Steffen Hamann, Hans-Peter Hartung, Joachim Havla, Bernhard Hemmer, Su-Chun Huang, Jaime Imitola, Vytautas Jasinskas, Hong Jiang, Rahele Kafieh, Ludwig Kappos, Randy Kardon, David Keegan, Eric Kildebeck, Ungsoo Samuel Kim, Sasha Klistorner, Benjamin Knier, Scott Kolbe, Thomas Korn, Lauren Krupp, Wolf Lagrèze, Letizia Leocani, Netta Levin, Petra Liskova, Jana Lizrova Preiningerova, Birgit Lorenz, Eugene May, David Miller, Janine Mikolajczak, Saddek Mohand Saïd, Xavier Montalban, Mark Morrow, Ellen Mowry, Joaquim Murta, Carlos Navas, Rachel Nolan, Katarzyna Nowomiejska, Frederike Cosima Oertel, Jiwon Oh, Celia Oreja-Guevara, Christophe Orssaud, Benjamin Osborne, Olivier Outteryck, Catarina Paiva, Jacky Palace, Athina Papadopoulou, Nikos Patsopoulos, Jana Lizrova Preiningerova, Nikolas Pontikos, Markus Preising, Jerry Prince, Daniel Reich, Robert Rejdak, Marius Ringelstein, Luis Rodriguez de Antonio, Jose-Alain Sahel, Bernardo Sanchez-Dalmau, Jaume Sastre-Garriga, Sven Schippling, Joel Schuman, Ken Shindler, Robert Shin, Neil Shuey, Kerstin Soelberg, Svenja Specovius, Agnese Suppiej, Alan Thompson, Ahmed Toosy, Rubén Torres, Valérie Touitou, Susanne Trauzettel-Klosinski, Anneke van der Walt, Patrick Vermersch, Angela Vidal-Jordana, Amy T. Waldman, Christian Waters, Russell Wheeler, Owen White, Helmut Wilhelm, Kimberly M. Winges, Nils Wiegerinck, Lenja Wiehe, Thomas Wisnewski, Sui Wong, Jens Würfel, Shadi Yaghi, Yuyi You, Zhaoxia Yu, Patrick Yu-Wai-Man, Reda Žemaitienė, and Hanna Zimmermann. More details about these collaborators are provided in Text S1.

Author Contributions

Axel Petzold: Study concept and design, project supervision, monthly project meetings, review of literature, drafting of manuscript, and revisions of manuscript. Phillip Albrecht: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript. Laura Balcer: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript. Erik Bekkers: Review of literature, intellectual content, and revisions of manuscript. Alexander U. Brandt: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript. Peter A. Calabresi: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript. Orla Galvin Deborah: Patient perspective, revision of content, and final version of the manuscript. Jennifer S. Graves: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript. Ari Green: Critical revision of the manuscript for important intellectual content. Pearse A. Keane: Critical revision of the manuscript for important intellectual content. Jenny A. Nij Bijvank: Critical revision of the manuscript for important intellectual content. Friedemann Paul: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript. Shiv Saidha: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript. Josemir W. Sander: Critical revision of the manuscript for important intellectual content. Pablo Villoslada: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript. Siegfried K. Wagner: Critical revision of the manuscript for important intellectual content. E. Ann Yeh: Monthly project meetings, review of literature, intellectual content, and revisions of manuscript.

Conflict of Interest

A. Petzold is part of the steering committee of the ANGI network which is sponsored by ZEISS, steering committee of the OCTiMS study which is sponsored by Novartis, and reports speaker fees from Heidelberg Engineering. P. Albrecht reports consulting fees, research grants, and nonfinancial support from Allergan, Biogen, Celgene, Ipsen, Merck Serono, Merz Pharmaceuticals, Novartis, and Roche, consulting fees, and nonfinancial support from Bayer Healthcare, and Sanofi-Aventis/Genzyme, outside the submitted work. L. Balcer reports personal fees from Biogen; she is editor in chief of the Journal of Neuro-Ophthalmology. E. Bekkers has nothing to disclose. A. Brandt is cofounder and shareholder of startups Motognosis and Nocturne. He is named as inventor on several patent applications description MS serum biomarkers, perceptive visual computing, and retinal image analysis. R. Bremel has served as a consultant for Biogen, EMD Serono, Genzyme/Sanofi, Genentech/Roche, Novartis, and Viela Bio. He receives ongoing research support directed to his institution from Biogen, Genentech, and Novartis. P.A. Calabresi has received consulting fees for serving on scientific advisory boards for Biogen and Disarm Therapeutics, and

is PI on grants to Johns Hopkins from Biogen, Gentech, and Annexon. O. Galvin has nothing to disclose. J.S. Graves has grant/contract research support from the National MS Society, Biogen, and Octave Biosciences. She serves on a steering committee for a trial supported by Novartis. She has received honoraria for a nonpromotional, educational activity for Sanofi-Genzyme. She has received speaker fees from Alexion and BMS and served on an advisory board for Genentech. A. Green reports grants and other support from Inception Biosciences; grants from the National Multiple Sclerosis Society and from the US. National Institutes of Health; additional support from MedImmune, Mylan, Sandoz, Dr Reddy, Amneal, Momenta, Synthon, and JAMA Neurology, outside the submitted work; and that the Multiple Sclerosis Center, Department of Neurology, University of California San Francisco has received grant support from Novartis for participating in the OCTIMS study. P.A. Keane is supported by a Clinician Scientist award (CS-2014-14-023) from the National Institute for Health Research. J. Nij Bijvank has nothing to disclose. J.W. Sander has been consulted by and received research grants and fees for lectures from Eisai, UCB, Zogenix, and GW Pharmaceuticals, outside the submitted work. F. Paul receives funding from Deutsche Forschungsgemeinschaft, Bundesministerium für Bildung und Forschung, and Guthy Jackson Charitable Foundation. FC has received consulting fees from Clene, EMD Serono, and PRIME, and is participating as a site investigator in the Novartis-funded OCTIMS study. S. Saidha has received consulting fees from Medical Logix for the development of CME programs in neurology and has served on scientific advisory boards for Biogen-Idec, Genzyme, Genentech Corporation, EMD Serono, and Celgene. He was the site investigator of a trial sponsored by Med-Day Pharmaceuticals, and is the PI of investigator-initiated studies funded by Genentech Corporation and Biogen Idec, and received support from the Race to Erase MS foundation. He has received equity compensation for consulting from JuneBrain LLC, a retinal imaging device developer. P. Villoslada has received an honorarium from Heidelberg Engineering in 2014, has received unrestricted research grants from Novartis (including for the OCTIMS study), Biogen, Genzyme, and Roche, and has participated in advisory boards for Novartis, Roche, Genzyme, and Biogen. PVi holds stocks in the following spin-off companies: Bionure Inc, Spire Bioventures, Mintlabs, and Health Engineering. S. Wagner has nothing to disclose. E. Ann Yeh has received research funds from NMSS, CIHI, CIHR, NIH, OIRM, MS Society of Canada, Mario Battaglia Foundation, SickKids Foundation, CBMH Innovation Fund, CMSC, Stem Cell Network, Department of Defense, Rare Diseases Foundation, and Biogen. Unrestricted educational funds from Teva and Guthy-Jackson Foundation. She has served on a scientific advisory panel for Hoffmann-La Roche and Biogen and has received speaker's honoraria from Novartis, CMSC, MS at the Limits, and Canadian Rheumatological Association.

Funding information

References

1. Fauw JD, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat Med 2018;24:1342–1350.

2. Tewarie P, Balk L, Costello F, et al. The OSCAR-IB consensus criteria for retinal OCT quality assessment. PLoS One 2012;7:e34823.

3. Petzold A, de Boer JF, Schippling S, et al. Optical coherence tomography in multiple sclerosis: a systematic review and meta-analysis. Lancet Neurol 2010;9:921–932.

4. Peng Y, Zhao Y, Hu W, et al. Reduction of retinal nerve fiber layer thickness in vigabatrin-exposed patients: a meta-analysis. Clin Neurol Neurosurg 2017;157:70–75.

5. Petzold A, Balcer LJ, Calabresi PA, et al. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. Lancet Neurol 2017;16:797–812.

6. Chan VT, Sun Z, Tang S, et al. Spectral-domain OCT measurements in Alzheimer's disease. Ophthalmology 2019;126:497–510.

7. Chrysou A, Jansonius NM, van Laar T. Retinal layers in Parkinson's disease: a meta-analysis of spectral-domain optical coherence tomography studies. Parkinsonism Relat Disord 2019;64:40–49.

8. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. The Lancet Digital Health 2019;1:e271–e297.

9. Milea D, Najjar RP, Jiang Z, et al. Artificial intelligence to detect papilledema from ocular fundus photographs. N Engl J Med 2020;382(18):1687–1695.

10. Schippling S, Balk L, Costello F, et al. Quality control for retinal OCT in multiple sclerosis: validation of the OSCAR-IB criteria. Mult Scler 2015;21:163–170.

11. Cruz-Herranz A, Balk LJ, Oberwahrenbrock T, et al. The APOSTEL recommendations for reporting quantitative optical coherence tomography studies. Neurology 2016;86:2303–2309.

12. Editors TL. Expression of concern: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. Lancet (London, England): 2020.

13. Mehra MR, Desai SS, Kuy SR, et al. Cardiovascular disease, drug therapy, and mortality in Covid-19. N Engl J Med 2020;382(25):e102.

14. Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. New Biotechnol 2012;29:613–624.

15. Grinfeld J, Nangalia J, Baxter EJ, et al. Classification and personalized prognosis in myeloproliferative neoplasms. N Engl J Med 2018;379:1416–1430.

16. Balk LJ, de Vries-Knoppert WAEJ, Petzold A. A simple sign for recognizing Off-Axis OCT measurement beam placement in the context of multicentre studies. PLoS One 2012;7:e48222.

17. Oberwahrenbrock T, Weinhold M, Mikolajczak J, et al. Reliability of intra-retinal layer thickness estimates. PLoS One 2015;10:e0137316.

18. Albrecht P, Müller A-K, Südmeyer M, et al. Optical coherence tomography in Parkinsonian syndromes. PLoS One 2012;7:e34891.

19. Borm CD, Smilowska K, de Vries NM, et al. The neuro-ophthalmological assessment in Parkinson's disease. J Parkinson's Dis 2019;9:427–435.

20. Roth NM, Saidha S, Zimmermann H, et al. Optical coherence tomography does not support optic nerve involvement in amyotrophic lateral sclerosis. Eur J Neurol 2013;20:1170–1176.

21. Ringelstein M, Albrecht P, Südmeyer M, et al. Subtle retinal pathology in amyotrophic lateral sclerosis. Ann Clin Transl Neurol 2014;1(4):290–297.

22. Hübers A, Müller HP, Dreyhaupt J, et al. Retinal involvement in amyotrophic lateral sclerosis: a study with optical coherence tomography and diffusion tensor imaging. J Neural Transm 1996;2016(123):281–287.

23. Simonett JM, Huang R, Siddique N, et al. Macular sub-layer thinning and association with pulmonary function tests in Amyotrophic Lateral Sclerosis. Sci Rep 2016;6:29187.

24. Pula JH, Yuen CA. Eyes and stroke: the visual aspects of cerebrovascular disease. Stroke Vasc Neurol 2017;2:210–220.

25. Clayton LM, Dévilé M, Punte T, et al. Retinal nerve fiber layer thickness in vigabatrin-exposed patients. Ann Neurol 2011;69:845–854.

26. Silverstein SM, Rosen R. Schizophrenia and the eye. Schizophrenia Res Cognition 2015;2:46–55.

27. Chilamkurthy S, Ghosh R, Tanamala S, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. The Lancet 2018;392(10162):2388–2396.

28. Marinescu RV, Oxtoby NP, Young AL, et al. TADPOLE Challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data, Predictive intelligence in medicine : second International Workshop. PRIME 2019, held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings. PRIME (Workshop) (2nd : 2019 : Shenzhen Shi, China) 2019;11843:1–10.

29. Rasheed K, Qayyum A, Qadir J, et al. Machine learning for predicting epileptic seizures using EEG signals: a review. IEEE Rev Biomed Eng 2021;14:139–155.

30. Neumann W-J, Turner RS, Blankertz B, et al. Toward electrophysiology-based intelligent adaptive deep brain stimulation for movement disorders. Neurotherapeutics 2019;16:105–118.

31. Bashyam VM, Erus G, Doshi J, et al. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. Brain. 2020;143:2312–2324.

32. Cornblath EJ, Robinson JL, Irwin DJ, et al. Defining and predicting transdiagnostic categories of neurodegenerative disease. Nat Biomed Eng 2020;4:787–800.

33. Dean S, Mathers JM, Calvert M, et al. "The patient is speaking": discovering the patient voice in ophthalmology. Br J Ophthalmol 2017;101:700–708.

34. Deverka PA, Lavallee DC, Desai PJ, et al. Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. J Comp Eff Res 2012;1:181–194.

35. Braithwaite T, Wiegerinck N, Petzold A, Denniston A. vision loss from atypical optic neuritis: patient and physician perspectives. Ophthalmology and Therapy 2020;9(2):215–220.

36. Courbier S, Dimond R, Bros-Facer V. Share and protect our health data: an evidence based approach to rare disease patients' perspectives on data sharing and data protection-quantitative survey and recommendations. Orphanet J Rare Dis 2019;14:175.

37. Tom E, Keane PA, Blazes M, et al. Protecting data privacy in the age of AI-enabled ophthalmology. Transl Vis Sci Technol 2020;9:36.

38. Larson DB, Magnus DC, Lungren MP, et al. Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework. Radiology 2020;295:675–682.

39. Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. https://ico.org.uk/media/1061/anonymisation-code.pdf. Accessed December 31, 2020.

40. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. Trials 2010;11.

41. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence 2020;2:305–311.

42. Lewis P, Conn D, Pegg D. (2020). UK government using confidential patient data in coronavirus "Response.

43. Petzold A, Chua SYL, Khawaja AP, et al. Retinal asymmetry in multiple sclerosis. Brain 2020.

44. Coric D, Petzold A, Uitdehaag BMJ, Balk LJ. Software updates of OCT segmentation algorithms influence longitudinal assessment of retinal atrophy. J Neurol Sci 2018;387:16–20.

45. Yu M, Tham Y-C, Rim TH, et al. Reporting on deep learning algorithms in health care. The Lancet Digital Health 2019;1:e328–e329.

46. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol 2003;22:85–93.

47. Cook JA, Julious SA, Sones W, et al. DELTA, javax.xml.bind.JAXBElement@5af99c39, guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. BMJ (Clinical research ed.) 2018;363:k3750.

48. Cook JA, Hislop J, Adewuyi TE, et al. Assessing methods to specify the target difference for a randomised controlled trial: DELTA (Difference ELicitation in TriAls) review. Health Technol Assess 2014;18(28):1–175.

49. Eysenbach G. CONSORT-EHEALTH: Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions. J Med Internet Res 2011;13:e126.

50. CONSORT-AI and Group, S.-A. S. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. Nat Med 2019;25:1467–1468.

51. Beiki O, Frumento P, Bottai M, et al. Changes in the risk of reaching multiple sclerosis disability milestones in recent decades: a nationwide population-based cohort study in Sweden. JAMA Neurol 2019;76:665–671.

52. Cook TS. Human versus machine in medicine: can scientific literature answer the question? The Lancet Digital Health 2019;1:e246–e247.

53. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016;529:484–489.

54. Gibney E. What Google's winning Go algorithm will do next. Nature 2016;531:284–285.

55. Biousse V, Newman NJ, Najjar RP, et al. Optic disc classification by deep learning versus expert neuro-ophthalmologists. Ann Neurol 2020.

56. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020;22:e15154.

57. Nagao K. Symbiosis between Humans and Artificial Intelligence. In ed. Artificial Intelligence Accelerates Human Learning. Springer, 2019.

58. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Quality Safety 2014;23:727–731.

59. Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med 2019;25:433–438.

60. Gruber K. Is the future of medical diagnosis in computer algorithms? The Lancet Digital Health 2019;1:e15–e16.

61. Sanchez-Martinez S, Camara O, Piella G, et al. Machine Learning for Clinical Decision-Making: Challenges and Opportunities, 2019.

62. Little J, Higgins JPT, Ioannidis JPA, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. PLoS Med 2009;6:e22.

63. Wu Z, Weng DSD, Thenappan A, et al. Evaluation of a region-of-interest approach for detecting progressive glaucomatous macular damage on optical coherence tomography. Transl Vis Sci Technol 2018;7:14.

64. de Vries-Knoppert WA, Baaijen JC, Petzold A. Patterns of retrograde axonal degeneration in the visual system. Brain 2019;142:2775–2786.

65. Jindahra P, Petrie A, Plant GT. The time course of retrograde trans-synaptic degeneration following occipital lobe damage in humans. Brain 2012;135:534–541.

66. Panneman E, Coric D, Tran L, et al. Progression of anterograde trans-synaptic degeneration in the human retina is modulated by axonal convergence and divergence. Neuro-Ophthalmol 2019;43:382–390.

67. Jindahra P, Petrie A, Plant GT. Retrograde trans-synaptic retinal ganglion cell loss identified by optical coherence tomography. Brain 2009;132:628–634.

68. Sinnecker T, Oberwahrenbrock T, Metz I, et al. Optic radiation damage in multiple sclerosis is associated with visual dysfunction and retinal thinning - an ultrahigh-field MR pilot study. Eur Radiol 2014;25:122–131.

69. Gelfand JM, Nolan R, Schwartz DM, et al. Microcystic macular oedema in multiple sclerosis is associated with disease severity. Brain 2012;135:1786–1793.

70. Saidha S, Sotirchos ES, Ibrahim MA, et al. Microcystic macular oedema, thickness of the inner nuclear layer of the retina, and disease characteristics in multiple sclerosis: a retrospective study. Lancet Neurol 2012;11:963–972.

71. Gelfand JM, Cree BA, Nolan R, et al. Microcystic inner nuclear layer abnormalities and neuromyelitis optica. JAMA Neurol 2013;1–5.

72. Knier B, Schmidt P, Aly L, et al. Retinal inner nuclear layer volume reflects response to immunotherapy in multiple sclerosis. Brain 2016;139:2855–2863.

73. Balk LJ, Coric D, Knier B, et al. Retinal inner nuclear layer volume reflects inflammatory disease activity in multiple sclerosis; a longitudinal OCT study. Mult Scler J Exp Transl Clin 2019;5(3):205521731987158.

74. Burggraaff MC, Trieu J, de Vries-Knoppert WAEJ, et al. The clinical spectrum of microcystic macular edema. Invest Ophthalmol Vis Sci 2014;55:952–961.

75. Brandt AU, Oberwahrenbrock T, Kadas EM, et al. Dynamic formation of macular microcysts independent of vitreous traction changes. Neurology 2014;83:73–77.

76. Balk LJ, Killestein J, Polman CH, et al. Microcystic macular oedema confirmed, but not specific for multiple sclerosis. Brain 2012;135:e226.

77. Abegg M, Dysli M, Wolf S, et al. Microcystic macular edema: retrograde maculopathy caused by optic neuropathy. Ophthalmology 2014;121:142–149.

78. van Buren J. Trans-synaptic retrograde degeneration in the visual system of primates. J Neurol Neurosurg Neuropsych 1963;26:402–409.

79. Kleerekooper I, Petzold A, Trip SA. Anterior visual system imaging to investigate energy failure in multiple sclerosis. Brain 2020;143(7):1999–2008.

80. Keane PA, Balaskas K, Sim DA, et al. Automated analysis of vitreous inflammation using spectral-domain optical coherence tomography. Transl Vis Sci Technol. 2015;4 (5):4–10.

81. Montesano G, Way C, Ometto G, et al. Optimizing OCT acquisition parameters for assessments of vitreous haze for application in uveitis. Sci Rep 2018;8:1648.

82. Coric D, Ometto G, Montesano G, et al. Objective quantification of vitreous haze on optical coherence tomography scans: no evidence for relationship between uveitis and inflammation in multiple sclerosis. Eur J Neurol 2019.

83. Kaufhold F, Kadas EM, Schmidt C, et al. Optic nerve head quantification in idiopathic intracranial hypertension by spectral domain OCT. PLoS One 2012;7: e36965.

84. Albrecht P, Blasberg C, Ringelstein M, et al. Optical coherence tomography for the diagnosis and monitoring of idiopathic intracranial hypertension. J Neurol 2017;264:1370–1380.

85. Albrecht P, Blasberg C, Lukas S, et al. Retinal pathology in idiopathic moyamoya angiopathy detected by optical coherence tomography. Neurology 2015;85:521–527.

86. Malmqvist L, Bursztyn L, Costello F, et al. The optic disc drusen studies consortium recommendations for diagnosis of optic disc drusen using optical coherence tomography. J Neuroophthalmol 2018;38:299–307.

87. Yadav SK, Motamedi S, Oberwahrenbrock T, et al. CuBe: parametric modeling of 3D foveal shape using cubic Bézier. Biomedical Optics Express 2017;8:4181–4199.

88. Motamedi S, Oertel FC, Yadav SK, et al. Altered fovea in AQP4-IgG–seropositive neuromyelitis optica spectrum disorders. Neurol Neuroimmunol Neuroinflamm 2020;7: e805.

89. Bennett MR, Farnell L, Gibson WG. Quantitative relations between transient BOLD responses, cortical energetics, and impulse firing in different cortical regions. J Neurophysiol 2019;122:1226–1237.

90. Feucht N, Maier M, Lepennetier G, et al. Optical coherence tomography angiography indicates associations of the retinal vascular network and disease activity in multiple sclerosis. Multiple Sclerosis J 2018;25:224–234.

91. Murphy OC, Kwakyi O, Iftikhar M, et al. Alterations in the retinal vasculature occur in multiple sclerosis and exhibit novel correlations with disability and visual function measures. Multiple Sclerosis J 2019;26:815–828.

92. Iftikhar M, Zafar S, Gonzalez N, et al. Image artifacts in optical coherence tomography angiography among patients with multiple sclerosis. Curr Eye Res 2019;44:558–563.

93. Kleerekooper I, Houston S, Dubis AM, et al. Optical coherence tomography angiography (OCTA) in multiple sclerosis and neuromyelitis optica spectrum disorder. Front Neurol 2020;11.

94. Petzold A, Wattjes MP, Costello F, et al. The investigation of acute optic neuritis: a review and proposed protocol. Nat Rev Neurol 2014;10(8): 447–458.

95. Coric D, Balk LJ, Uitdehaag BMJ, Petzold A. Diagnostic accuracy of optical coherence tomography Inter-Eye Percentage Difference (IEPD) for optic neuritis in multiple sclerosis. Eur J Neurol 2017;24:1479–1484.

96. Nolan-Kenney RC, Liu M, Akhand O, et al. Optimal intereye difference thresholds by optical coherence tomography in multiple sclerosis: an international study. Ann Neurol 2019;85:618–629.

97. Xu SC, Kardon RH, Leavitt JA, et al. Optical coherence tomography is highly sensitive in detecting prior optic neuritis. Neurology 2019;92:e527–e535.

98. Behbehani R, Ali A, Al-Omairah H, Rousseff RT. Optimization of spectral domain optical coherence tomography and visual evoked potentials to identify unilateral optic neuritis. Mult Scler Relat Disord 2020;41:101988.

99. Davion J-B, Lopes R, Drumez É, et al. Asymptomatic optic nerve lesions. Neurology 2020;94(23):e2468–e2478.

100. Outteryck O, Lopes R, Drumez É, et al. Optical coherence tomography for detection of asymptomatic optic nerve lesions in clinically isolated syndrome. Neurology 2020;95:e733–e744.

101. Villoslada P, Sanchez-Dalmau B, Galetta S. Optical coherence tomography: A useful tool for identifying subclinical optic neuropathy in diagnosing multiple sclerosis. Neurology 2020;95:239–240.

102. Burns SA, Elsner AE, Sapoznik KA, et al. Adaptive optics imaging of the human retina. Prog Retin Eye Res 2019;68:1–30.

103. Romano F, Arrigo A, Leone PP, et al. Altered ellipsoid zone reflectivity and deep capillary plexus rarefaction correlate with progression in Best disease. Br J Ophthalmol 2020;104:461–465.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Text S1. List of committee members: IMSVISUAL and ERN-EYE.